Hindawi
Complexity
Volume 2024, Article ID 9823476, 1 page
https://doi.org/10.1155/2024/9823476

Retraction

Retracted: Simulation of Tennis Match Scene Classification Algorithm Based on Adaptive Gaussian Mixture Model Parameter Estimation

Complexity

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wang and M. Wen, "Simulation of Tennis Match Scene Classification Algorithm Based on Adaptive Gaussian Mixture Model Parameter Estimation," *Complexity*, vol. 2021, Article ID 3563077, 12 pages, 2021.

Hindawi
Complexity
Volume 2021, Article ID 3563077, 12 pages
https://doi.org/10.1155/2021/3563077

Research Article

Simulation of Tennis Match Scene Classification Algorithm Based on Adaptive Gaussian Mixture Model Parameter Estimation

Yuwei Wang[1] **and Mofei Wen**[2]

[1]*Chengdu Sport University, Sichuan, Chengdu 61000, China*
[2]*Physical Education College, Chengdu University, Sichuan, Chengdu 61000, China*

Correspondence should be addressed to Mofei Wen; wenmofei@cdu.edu.cn

Received 17 April 2021; Revised 5 May 2021; Accepted 8 May 2021; Published 19 May 2021

Academic Editor: Zhihan Lv

This paper presents an in-depth analysis of tennis match scene classification using an adaptive Gaussian mixture model parameter estimation simulation algorithm. We divided the main components of semantic analysis into type of motion, distance of motion, speed of motion, and landing area of the tennis ball. Firstly, for the problem that both people and tennis balls in the video frames of tennis matches from the surveillance viewpoint are very small, we propose an adaptive Gaussian mixture model parameter estimation algorithm, which has good accuracy and speed on small targets. Secondly, in this paper, we design a sports player tracking algorithm based on role division and continuously lock the target player to be tracked and output the player region. At the same time, based on the displacement information of the key points of the player's body and the system running time, the distance and speed of the player's movement are obtained. Then, for the problem that tennis balls are small and difficult to capture in high-speed motion, this paper designs a prior knowledge-based algorithm for predicting tennis ball motion and landing area to derive the landing area of tennis balls. Finally, this paper implements a prototype system for semantic analysis of real-time video of tennis matches and tests and analyzes the performance indexes of the system, and the results show that the system has good performance in real-time, accuracy, and stability.

1. Introduction

With the Olympic Games, World Cup, and other large sports events, people are becoming increasingly obsessed with sports. With the improvement of national living standards, outbound tourism has shown a high growth trend in recent years, the residents' consumption upgrade has promoted the transformation and upgrading of the tourism industry, and compared with other industries, tourism consumption is more inclined to experience-based consumption, tourism, and other industries to integrate the development of new business models to meet diversified consumer demand; "tourism" will become the new trend of the next phase of the development of the outbound tourism industry. As a popular ball sport, every tennis fan will pay attention to the four major open tennis tournaments. However, the traditional form of watching tennis matches has been unable to meet the needs of people's sports entertainment, and with the advent of the era of artificial intelligence, the traditional form of tennis matches broadcast intelligent upgrade has given the time to wait for a wide audience of badminton naturally also affected by the development of technology and the birth of some intelligent devices based on data analysis [1]. However, the current intelligent application in the field of badminton is very limited, and the judgment part of the game is still subject to the naked eye judgment of the referee. In competitive badminton, there are explicit rules for all the movements of the players, and the referee is required to make a penalty according to the rules of the game. However, the referee's behaviour may be influenced by many factors such as the player's desire to win or lose the game and personal habits, which may lead to some referees not complying with the rules of the game to make penalties, such as misjudgements

and even black whistles [2]. With the use of various new technologies in the broadcasting and adjudication of different sports events, the traditional way of adjudication by the naked eye alone is bound to suffer a huge impact. In recent years, researchers have gradually started to study the badminton game scenario, and the research topics range from the penalty system for the whole badminton field to the tee system only for badminton training and slowly emerge from the study of the characteristics of the player's action statistics and swing action judgment [3]. In this paper, we design and implement a machine vision-based badminton serve violation detection system, which is a third-party mechanism not influenced by subjective factors, to improve the impartiality of the referee's decision by realizing an auxiliary penalty when serving the ball.

To solve the shortcomings of the traditional model of the video surveillance system, researchers began to try to combine computer vision-related technology with a video surveillance system, by using the camera instead of the human eye to monitor, through the computer instead of the human brain to observe and analyse, forming a complete set of the intelligent video surveillance system [4]. The so-called intelligent video surveillance system refers to the intelligent processing of the video information obtained by the camera through computer vision-related technologies, then the acquired image information for motion target detection to extract useful information, and then the target information for event behaviour analysis, and finally according to the preset rules to make the corresponding response. In the whole intelligent surveillance system, the surveillance system itself alone can lock the specific target of interest and achieve automatic analysis and judgment of the relevant situation of the moving target and make timely response and processing of some potential abnormal behaviour [5].

These systems have played an important role in modern competitive sports, however, there is no research on the semantic parsing of the entire tennis match video. This paper is the first attempt to perform real-time semantic analysis of tennis match video and develop a novel prototype system for real-time video semantic analysis of tennis matches. Specifically, this paper selects the video of a tennis match from a monitoring perspective as the object of analysis. The semantic analysis of a tennis match in real-time includes two major aspects: on the one hand, the semantic analysis of the players' motion information, including motion data and motion categories and on the other hand, the analysis of the trajectory and landing point of the tennis ball motion. Through the research and implementation of this semantic analysis technology, the movement on the tennis court is tracked, recorded, and analysed, and the data feedback and guidance are provided in real-time. This combination of online semantic analysis and offline matches provides teaching aids for umpires and a fresh viewing experience for spectators, which is of great research significance.

1.1. Status of Research. The problem of target detection has been of great interest in the field of computer vision. Target detection has been a challenging problem due to the influence of natural factors such as illumination, occlusion, and various reasons such as different appearance, shape, and pose of objects [6]. Finding the location, size, and shape of the target that may appear in the image is the fundamental task of target detection. Since the targets have different aspect ratios, it is too expensive to solve the generic target detection problem using the classical sliding window +image scaling scheme [7]. Tennis related research is relatively very rare, and for badminton there are many studies; we study related algorithms, and the similarity between the two is higher, so we choose badminton-related content. With the advancement of computer hardware, deep learning has been able to develop rapidly [8]. This also brought a breakthrough in target detection. Region-Convolution Neural Network (r-CNN) first used the CNN approach for target detection by determining whether multiple borders extracted from an image corresponding to a certain class of objects [9]. Then, the new network FastR-CNN2, also proposed by Ross Kirchick, solved the problem of slow speed left by R-CNN. The "Faster R-CNN" network structure was then proposed by a team from Microsoft for real-time target detection via a region proposal network that reuses the same CNN results from multiple region proposals [10]. The difference between the SPP-Net and the R-CNN is that the input does not need to be scaled down to a specified size and a spatial pyramid pooling layer is added, so that features are extracted only once per image [11].

The literature [12] introduced a variety of methods for badminton boundary detection: digital detection systems and line trial aids based on piezoelectric sensing technology, digital detection systems mainly using optical real-time tracking and capturing equipment, data information processing systems, and digital display devices to achieve the capture of badminton and calculate the corresponding coordinates to determine whether the boundary; the disadvantage is the requirement of the game ball after infrared spraying and being susceptible to Infrared interference outside the field. The calculation of the coordinates of the ball also has a large error. Line trial assisted device is by laying a strip with the insulating substrate with positive and negative wires on both sides of the court boundary and attaching conductive material to the ball, using the ball to touch the ground with the borderline to generate electrical induction to achieve the determination of the ball's landing point; the disadvantage of such devices is that the accuracy is easily affected by wet weather and player sweat, but also there is a need to change the standard construction of badminton; literature [13] further improves the sports action. To further improve the accuracy of classification, an optimized Back Propagation (BP) neural network was proposed to train motion features with better results; nowadays, using keyframes in videos as the basis for action determination has also started to gradually become a popular method, which extracts representative key gestures as keyframes in an action sequence [14]. Because of this, Kinect, for example, can read the human skeleton information consisting of the key points of the human body in real-time through a depth camera to estimate the human pose actions [15]. There are two difficult aspects of skeleton-based

behaviour recognition techniques: first, how to identify generic and strongly distinguishing features, and second, how to model the dynamic changes in behavioural actions in terms of time-domain correlation. Behaviour recognition is performed by analysing a presegmented temporal sequence frame by frame; each frame, in turn, contains the location information of k human key points, then classifying their behavioural actions. LSTM makes good use of the time-domain correlation for dynamic modelling [16].

In the competitive badminton scenario, most of the studies focus on recognizing a certain hitting action during the game or the badminton training process, and there are relatively few studies for the badminton game serve scenario. The process of serving is an important part of badminton matches, which is still mainly judged by the naked eye. The introduction of machine vision-related technologies to analyse the serving process and establish a new auxiliary penalty model will bring more convenience and fairness to future badminton matches. The main research of this paper is to build a real-time semantic intelligent analysis system for tennis match video based on the tennis match video from the surveillance viewpoint, targeting two major sports objectives: players and tennis. The semantics defined in the system mainly includes two major blocks: one is the real-time analysis of tennis players' movements and motion statistics, and the other is the prediction and analysis of tennis ball trajectory and landing area.

2. Simulation Analysis of Adaptive Gaussian Mixture Model Parameter Estimation Algorithm for Tennis Match Scene Classification

2.1. Adaptive Gaussian Mixture Model Parameter Estimation Algorithm Design. Motion target detection is divided into two categories, one is the motion target detection in static scenes, which refers to the video based on the camera's stationary conditions, and the other is the motion target detection in dynamic scenes, which refers to the video based on the camera's motion conditions. In most cases in daily life, the camera is at rest, so this paper mainly discusses the video foreground target detection in static scenes. Not all the information in the video sequence captured by the camera is of concern to us, and all we care about are people, vehicles, and other moving objects with practical significance [17]. Motion target detection from another point of view is the segmentation of the image, i.e., the segmentation of the foreground target from the background of a video sequence, extracting the motion target while representing it explicitly with mathematical parameter information.

The background subtraction method differentiates the current frame from the background image to achieve the detection of motion regions, so the algorithm must have a background image, and the background image can be updated at any time with the change of the actual environment so that the most important process of the background subtraction method is the establishment of its background model and its update. Firstly, the corresponding background model $BG(x, y)$ is established by some algorithm, which will be used as the background image with each frame to do the difference. When the algorithm is executed, the greyed-out image $P_k(x, y)$ of each frame is differenced from the background image $D_k(x, y)$ according to

$$D_k(x, y) = P_k(x, y) + BG(x, y). \tag{1}$$

The image obtained from the above equation is called the difference image $D_k(x, y)$ and the difference image principle. "On" only contains foreground target information, and the background model established by the algorithm used in practice is different from the current actual complex dynamic background. The image obtained from the above equation is called the difference image and the difference. Therefore, the obtained differential image still contains background pixels with interference factors, so it is necessary to binarize the differential image as in (2) after the set threshold, and the resulting binarized image is $B_k(x, y)$.

$$B_k(x, y) = \begin{cases} \text{prospect}, & D_k(x, y) \le \Delta T, \\ \text{background}, & \text{othewise}. \end{cases} \tag{2}$$

According to the feature of the continuous video sequence, it can be obtained that if there is a moving foreground target in the video scene, there is an obvious change between consecutive video frames; otherwise, it means that there is no obvious change between consecutive video images [18]. The interframe difference method draws on this idea and is divided into the two-frame difference method and the three-frame difference method. The two-frame difference method is calculated by calculating the pixel value difference between two adjacent video frames at the corresponding position pixel point. Since it is assumed that there is a big difference between foreground and background pixels of video frames, the interframe difference method can distinguish foreground from background accurately by the pixel value difference information, and the specific difference process is shown in

$$D(x, y, t-1) = P(x, y, t) + P(x, y, t-1). \tag{3}$$

Only the foreground target in the slowly changing scene between the two frames of the difference method reflects a better effect, and when the motion of the foreground target movement is a fast scene, the foreground target in the adjacent two frames of the image difference is large, resulting in the two frames of the difference method after subtraction of the foreground target is not complete. To improve the completeness of the extraction results, the three-frame difference method is proposed based on the two-frame difference method. The algorithm uses the three-frame difference method to acquire the motion target images and then uses the color threshold to separate the interference of other motion targets. The experiments use indoor laser point motion records as samples, and the algorithm detects the fast-irregular movement of laser points to determine the laser point motion trajectory. Although the three-frame difference method also uses the pixel difference information of adjacent frames in the video to determine the region

where the foreground target is located, the three-frame difference method is used to determine the region of the foreground target by differencing three adjacent frames as in (4).

$$D_1(x, y, t) = P(x, y, t+1) + P(x, y, t-1). \tag{4}$$

The single Gaussian model is used to simulate the unimodal background model, which considers the sequence of grey values of each pixel point in the image at the corresponding position in all video frames as random and discrete, Set the pixel point (x, y) at moment t and the grey value in the image of the i-th frame as $D_1(x, y, t)$. The single Gaussian model considers the grey value of each pixel point in the image in all frames as conforming to Gaussian distribution; then, the probability of occurrence of pixel value x is

$$P(x_t) = \frac{1}{\sqrt{2\pi\sigma}} e^{\left((x_t-\mu_0)^2/2\sigma_0^2\right)}. \tag{5}$$

It is difficult to adapt to complex scene changes with a constant background model, and so is the Gaussian mixture model. The key point is the update of the background model, which mainly refers to the update of important parameters in the model, i.e., the mean and variance and the right value, as shown in Figure 1.

Also, if a pixel does not match all its corresponding Gaussian models, a new Gaussian distribution needs to be created to replace the one with the smallest current weight, and the parameters of the newly created Gaussian distribution are determined by the same process as the initialization, with the value of the pixel as the mean, a larger value as the variance, and a smaller value as the weight; all other Gaussian distributions will have to have their weights reduced sequentially processing until.

The authors of RetinaNet investigated the reasons for the lagging detection accuracy of the single-stage target detection algorithm and proposed a novel training loss function, Focal loss, by emphasizing the class imbalance problem in the training process. This loss function can reduce the contribution of easy-to-classify samples to the loss function in the training process and prevent the training process from being dominated by many simple samples. To verify the effectiveness of Focal loss, the authors design a novel network structure named RetinaNet, which uses the ResNet classification network as the feature extraction network and adopts the feature pyramid network architecture to fuse the features extracted from different convolutional layers in a top-down manner.

RetinaNet uses the ResNet classification network as the feature extraction network and adopts the feature pyramid network architecture to fuse the features extracted from different convolutional layers in a top-down manner to build a multilevel feature pyramid structure containing rich semantic information. A classification subnetwork is used on the feature layers at different scales for target class prediction, and a regression subnetwork is used for object location regression. The whole network structure is shown in Figure 2. The ResNet network is used as the basic feature extraction network, the feature pyramid is constructed from the features extracted by ResNet, and the output detection results are predicted on the constructed feature pyramid. The whole network consists of convolutional and pooling layers, which is a fully convolutional network.

Focal loss is based on the traditional cross-entropy loss and gives different weights to the cross-entropy loss of different samples, which makes the contribution of easy-to-classify samples to the overall loss function decreases and the contribution of hard-to-classify samples to the overall loss function increases. The cross-entropy loss function of two categories is introduced here as an example of dichotomous classification, and the traditional cross-entropy loss is defined as

$$CE(p, y) = \begin{cases} \log(p), & y = 1, \\ \log(1-p), & y = -1, \end{cases} \tag{6}$$

where y denotes the category label of the sample, and since there are only two classes, y takes the values 1 and −1, respectively. p is the probability that the network predicts that the target belongs to category 1, and p takes the values in the range [0, 1]. For the convenience of presentation, the definition of p_t:

$$p_t = \begin{cases} p, & y = 1, \\ 1+p, & y = -1. \end{cases} \tag{7}$$

Equation (7) shows that the traditional cross-first loss treats both hard-to-classify samples and simple samples with the same degree of contribution to the overall loss function, which tends to lead to the learning process of the network being dominated by a large number of easy-to-classify negative samples when the single-stage target detector is trained [19]. To distinguish the contribution of easily classified samples and hard-to-classify samples to the overall loss function, Focal loss adds a weighting factor to the traditional cross-entropy loss, which has different weight sizes for easily classified samples and hard-to-classify samples. The focal loss function is defined as follows:

$$FL(p_t) = -(1+p_t)^{\gamma}\log(p_t), \tag{8}$$

where $FL(p_t)$ is defined as the modulation factor and γ is defined as the focus factor. For easy-to-classify samples, the network predicts a larger p_t, then the modulation factor is smaller, and the cross-first loss of easy-to-classify samples becomes smaller with the focusing factor fixed. For hard-to-classify samples, the network predicts a smaller p_t, the modulation factor is larger, and the cross-first loss of easy-to-classify samples becomes larger when the focus factor is fixed. The focus factor v is responsible for modulating the sampling weights of the easily classified samples, and when $\gamma = 0$, the Focal loss degenerates to the traditional cross-first loss. As y increases, the weight of easy-to-classify samples in the total loss becomes smaller and the weight of hard-to-classify samples in the overall loss becomes larger. The contribution of easy-to-classify samples in the overall loss function is reduced and the contribution of hard-to-classify samples in the overall loss function is increased, thus

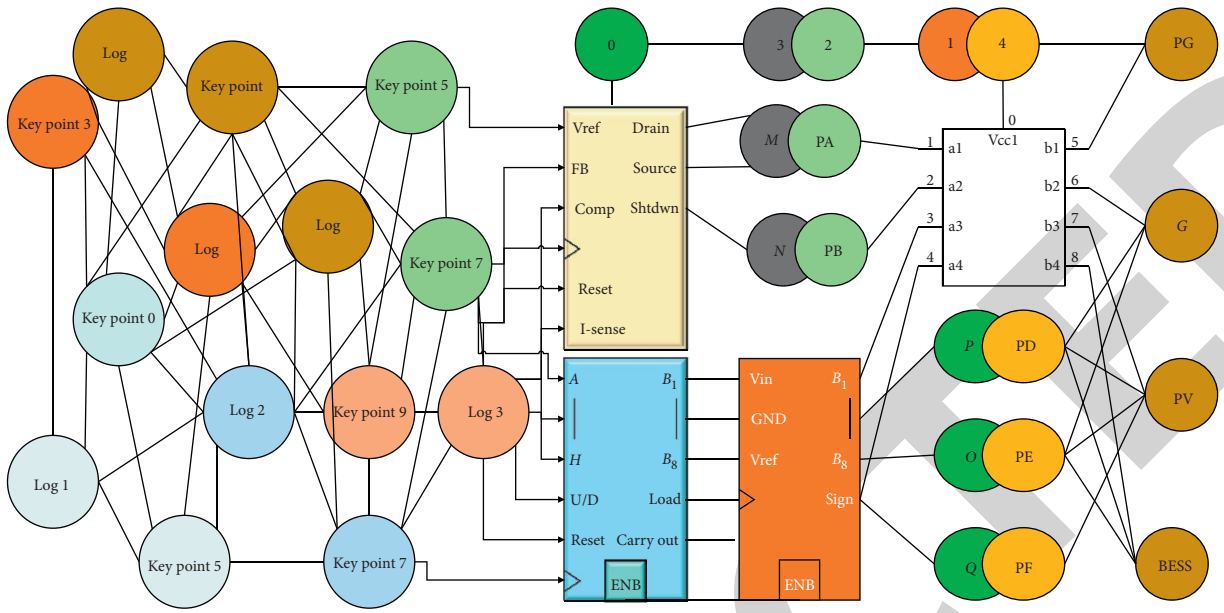

Figure 1: Adaptive Gaussian mixture model.

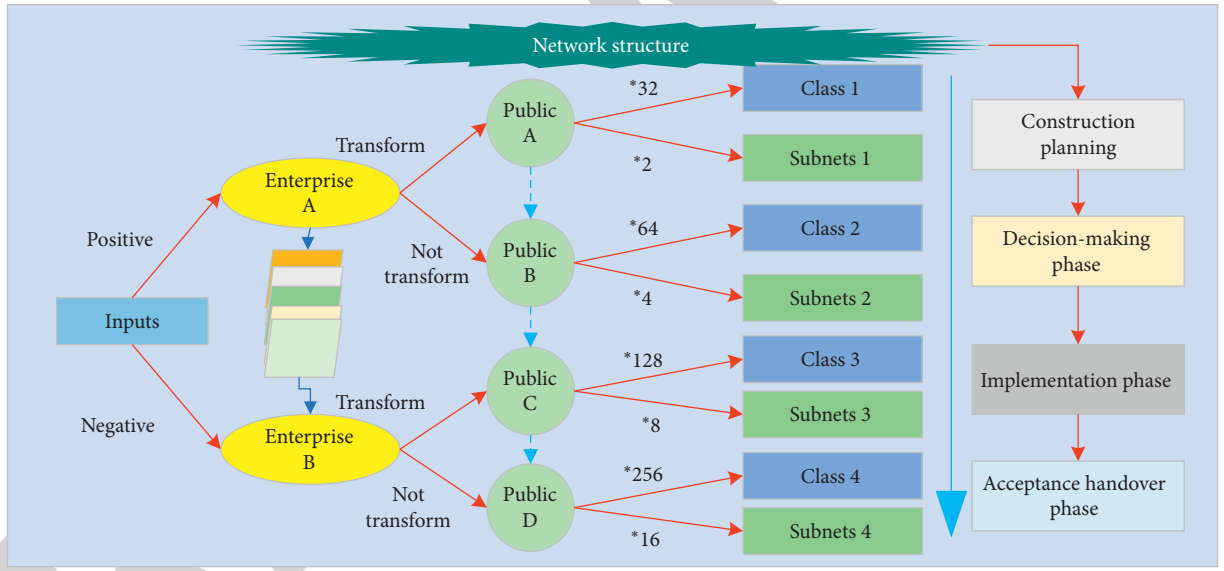

Figure 2: Network structure.

alleviating the category imbalance problem in the training phase of the single-stage target detector and improving the detection accuracy of the single-stage target detector.

2.2. Tennis Game Scene Classification System Design. To make the algorithm universal and the system practical, we build a complete tennis match dataset by collecting HD videos of all tennis matches. The dataset covers the four major tennis tournaments, which correspond to the Australian Open hard courts, French Open red clay courts, Wimbledon grass courts, and US Open hard courts. The specific data distribution is shown in Figure 3, with 20 segments of 12 s each for the Australian Open hard courts, 25 segments of 10 s each for the French Open red clay courts, 15 segments of 8 s each for the Wimbledon grass courts, and 18 segments of 8 s each for the US Open hard courts, for a total dataset size of about 23,000 video frames. The dataset was labelled using the label image tool, and the labelling categories were divided into two categories, tennis and people.

The module analyses the target sports player in the tennis match video, the input is the video frame, and the output is the semantic information in two aspects, including action type, running distance, running speed, movement time, and system running speed fps. The semantic information in

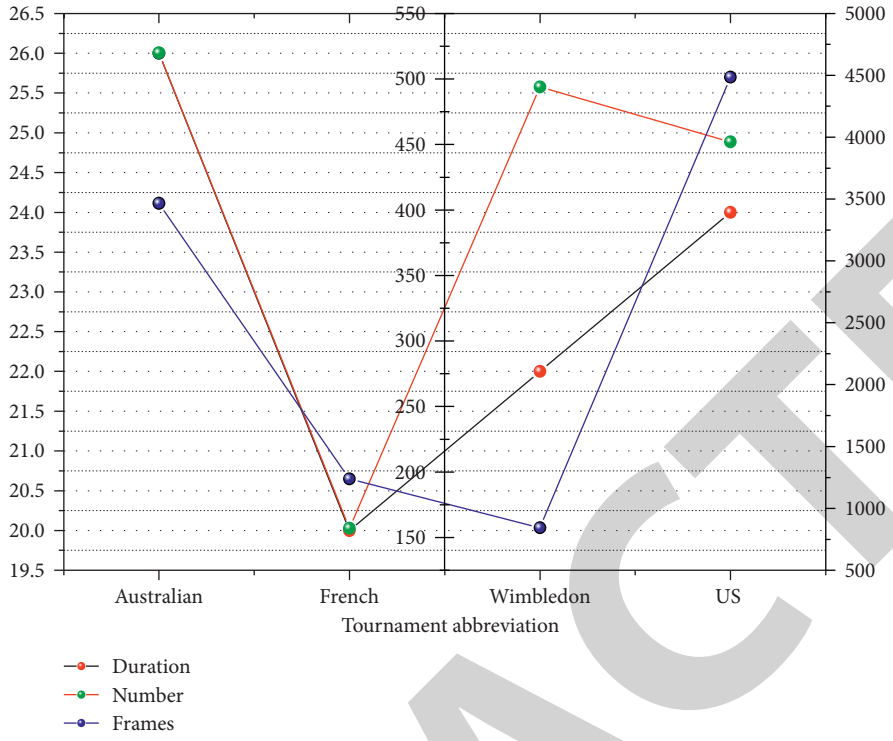

Figure 3: HD tennis match video statistics.

string format is sent to the client via a socket on the server-side. Starting from the input video frame of the tennis match, the series of steps in the dashed box is a coarse to fine process, initially finding out all the people on the court from the video frame with $1280 * 720$ resolution, then distinguishing the different roles according to the a priori knowledge of the tennis match occupancy, to find out the position of the target player in this frame, and then tracking him in each frame using Deep Sort algorithm [20]. After that, the player's body skeleton is analysed and the 18 key points of the centre of the body and the key points with the highest correlation with the swinging action are identified, and the action type is output according to the action discrimination rules formulated in the algorithm; similarly, the displacement of the centre of the body is calculated frame by frame to calculate the movement distance and running speed.

This module analyses the tennis ball in the video, the input is the video frame, and the output is the prediction result of the tennis ball's landing area. In this module, we take the a priori knowledge that the motion of tennis ball in the horizontal direction is uniform linear motion as the basis, simulate the trajectory of the tennis ball in the horizontal direction according to the recorded detection results, use the start and end signals inputted from the keyboard as the timestamp of the tennis ball's motion to finally find out its landing area, and finally encapsulate the output result and transmit it to the client using socket. The video and semantic display module should not only display the video transmitted back by socket, but also draw a series of results processed by the server-side, including the detection frame of human and tennis ball, the target player area, the target player's body skeleton, and key points, and of course, it also needs to display the semantic information output by the server-side, as shown in Figure 4.

In the previous section, a contour with a rotation angle was constructed for the target area, which was used to represent the racket state. By calculating the angle of this contour, the racket state was quantified as an angle value to determine whether the serve was overhand. However, the representation of the angle differs when the target area contour is located at the two ends of the net. The left end of the net is assumed to be the left end of the net, and the corresponding criteria are set, while the right end of the net is set to the opposite parameter. When the target profile is at the left end of the net, there are four different states of the racket in the judgment keyframe.

To make the status judgment mechanism as simple as possible, the process of judgment needs to be standardized by setting the first and fourth status as 0° and 90° respectively, while the second and third status are distinguished according to the positive or negative angle of the calculated angle, the difference being that the angle indicated by the second status is negative and the angle indicated by the third status is positive. To sum up, at the left end of the net, only the negative angle is considered a legal serve, while the positive angle, 90° or 0°, is considered an overhand violation; at the right end of the net, only the positive angle is considered a legal serve, while the negative angle, 90° or 0°, is considered an overhand violation. The target areas detected during the experiment were all parts of the racket structure, which did not affect the final judgment because the

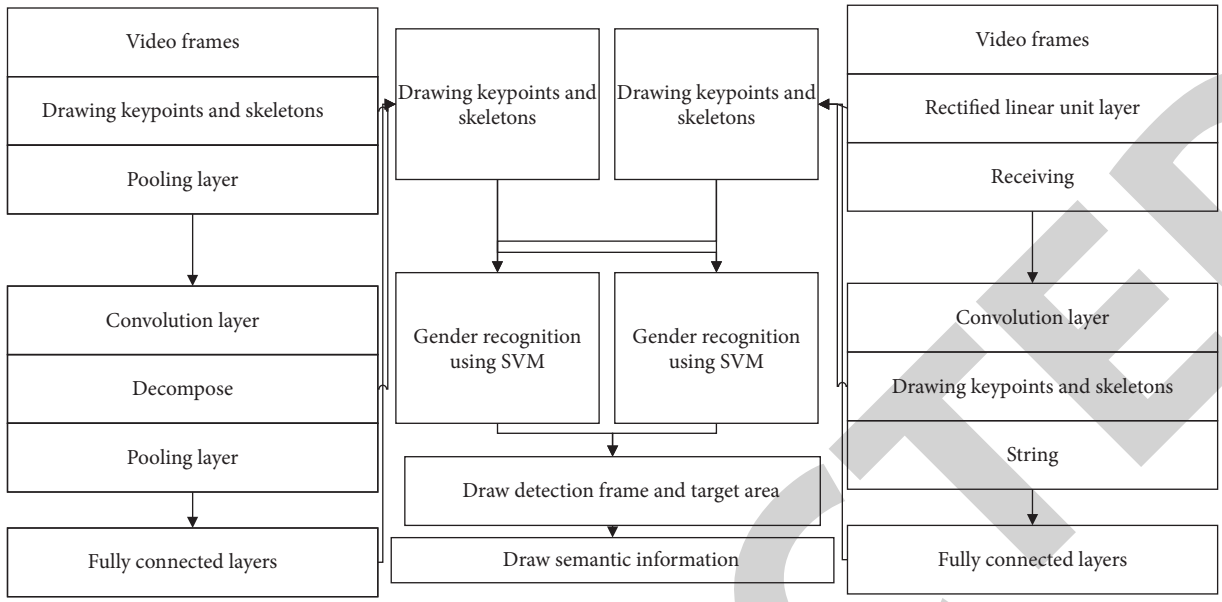

Figure 4: Flowchart of video and semantic display module.

quantification of the racket state into angle values did not require the entire racket profile, but only the angle of the main part of the profile could be calculated to represent the angle of the entire racket.

$$R(x,y) = \begin{cases} R(x,y) - \alpha \mathrm{d}p(x,y), & \mathrm{d}p(x,y) \le R(x,y), \\ R(x,y) + \beta \mathrm{d}p(x,y), & \text{otherwise.} \end{cases} \tag{9}$$

In this case, if we want to detect the highly dynamic background as the background pixel, we need to increase the matching radius adaptively to identify the highly dynamic background pixel; otherwise, when the average distance between any two pixels in the background model is larger than the initial value of the matching radius, it means that the background change is more dynamic. Otherwise, when the average distance of any pixel in the background model is less than the initial value of the matching radius, which means that the background is static, the adaptive matching radius $R(x, y)$ needs to be reduced to enable the detection of subtle background points and some noise, and α and β in the equation represent the two coefficients of the adaptive radius.

To solve the dragging problem caused by the original visual background model-based motion target detection algorithm that directly uses the first frame of the video as the initialized background model, many researchers have made improvements for this purpose. The improvements are all about how to improve the quality of the initialized model to be able to accurately measure the background condition of the current video. This has led many researchers to work on background extraction, and some researchers use the background extracted by multiframe averaging as the initialization model based on the visual background model, which is certainly more accurate than the original direct initialization value of the first frame of the video. However, this improved method has two shortcomings: first, when the number of frames is small, the background extracted by multiframe averaging is good, but as the number of frames increases, this background extraction algorithm is less effective, and second, when the number of frames is large, it greatly increases the time complexity of the program. Other researchers use the statistical histogram-based background extraction algorithm as the background initialization model for vision-based background modelling methods [21–27]. The shortcoming of this improvement is that the background extraction is good when in simple scenes, but for variable scenes, the background image extracted by this algorithm contains a lot of noise, resulting in still poor target detection results. The detection effect of using the background image extracted by the statistical histogram-based algorithm as the visual background-based motion target detection is similar to that of the multiframe averaging method, both of which are extracted with some foreground trailing shadows in the results, and this kind of extraction results with low accuracy will seriously affect the subsequent target detection process.

3. Results and Discussion

3.1. Adaptive Gaussian Mixture Model Performance Results. According to the videos selected in this paper by foreground target detection based on Gaussian mixture model modelling, the original visual background modelling method, the existing improved visual background modelling method, and foreground target detection by the improved visual background modelling algorithm proposed in this paper, the quantitative indexes accuracy rate, recall rate, and comprehensive evaluation are analysed according to the quantitative indexes proposed in the previous section, and the specific comparison data are shown in Figure 5.

From Figure 5, we know that in terms of accuracy index, the Gaussian modelling method has the lowest accuracy in

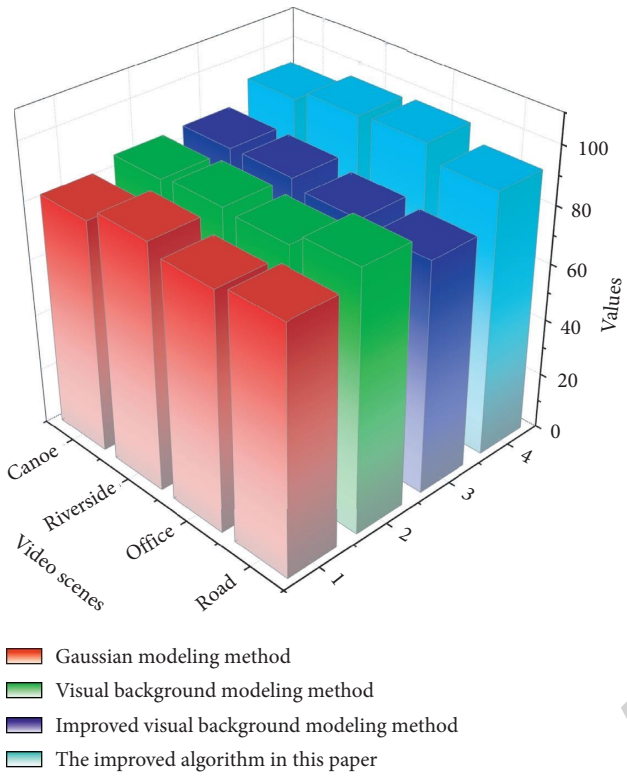

Figure 5: Comparison of accuracy of different algorithms.

Road, Office, and Riverside videos because it mistakenly detects the background pixel points (e.g., shaking leaves and water waves) as foreground pixel points in the videos, resulting in the smallest accuracy value, while Canoe video has the lowest accuracy of visual background modelling method because the method's large number of shaking. The improved visual background modelling method has slightly improved the accuracy compared with the first two methods because it adds the mechanism of flicker detection, which to some extent avoids flicker detection as foreground target pixels; and this paper extracts the pure background as the initial value of the background model and adds the improved algorithm based on the adaptive matching threshold. The method of image computation uses image local thresholds to replace global thresholds, specifically for images with excessive changes in light and shadow, or images with less significant color differences in the range. It is an adaptive means to ensure that the computer can iterate by judging and calculating the average threshold value obtained for that image region. This paper extracts the pure background as the initial value of the background model and adds the improved algorithm based on the adaptive matching threshold to effectively identify the dynamic background in the scene, which is better than the other three algorithms in terms of detection accuracy index.

The relative positional relationship between each calibration plate and the camera in space is obtained at the end of calibration, which can simulate the state of each calibration plate under the camera viewpoint. The calibration results are back-projected onto each calibration picture and statistically analysed with the original corner point detection position as shown in Figure 6, and the average error of calibration is 0.32 pixels.

The motion of the mechanical mechanism is manipulated by a control system that develops a motion control strategy for the hardware so that the robot or mechanical claw obtains the specified motion. The prerequisite for successful control is the establishment of an accurate mathematical description model for the robot arm and robot claw, i.e., the use of mathematical models to portray the relative position relationships and relative motion relationships of the robot linkages and joints.

In this paper, the experimental task is to make the robot run successfully to this location with a known target posture. In the operation space, firstly, a series of intermediate points are obtained by interpolating between the current end posture and the desired end posture with certain rules, and the velocity and acceleration of each point are obtained with time constraints, forming a time series describing the end-effector position and posture following the constraints over time. Then, we use the inverse kinematic method to obtain the angle of each joint corresponding to the pose of each intermediate point to form the time series of joints; finally, we control the rotation of each joint motor according to the above kinematic law to ensure that the robot reaches the target pose smoothly and continuously. IK solutions are applied as standard transformation animation key points. "Applied IK" uses the linked hierarchy of all objects, except for bones using the "HI solver" or "IK limb solver." It is possible to combine forward and reverse kinematics on the same object. It can be applied automatically to a range of frames, or interactively to a single frame. The above process is called robot motion planning, and the main motion planning algorithms are divided into three categories: potential field method, graph search method, and random sampling method. The potential field method models the target gravitational force and the obstacle repulsion as parameters, which is fast but easy to fall into local minima; the graph search method divides the entire robot motion space into a grid-like traversal to search for feasible paths, which is reliable but very computationally intensive. The random sampling method takes a random series of points in space and searches for feasible paths from the current to the target among the generated random points, which gives up the completeness, but the search efficiency is greatly improved, more suitable for complex tasks in high-dimensional space, but some random search algorithms are probabilistically complete, such as the fast extended random tree algorithm (RRT). The starting and target bit pose is determined given that a long enough time RRT must be planned successfully.

3.2. System and Experimental Results. The performance of the classification model is evaluated by recording a test video with a camera at 30 fps in three environments, detecting the target in each frame with the classification model, ensuring that the target is constantly moving at different positions in the video, saving the detected frames of the target object for the interception, and finally counting the number of frames

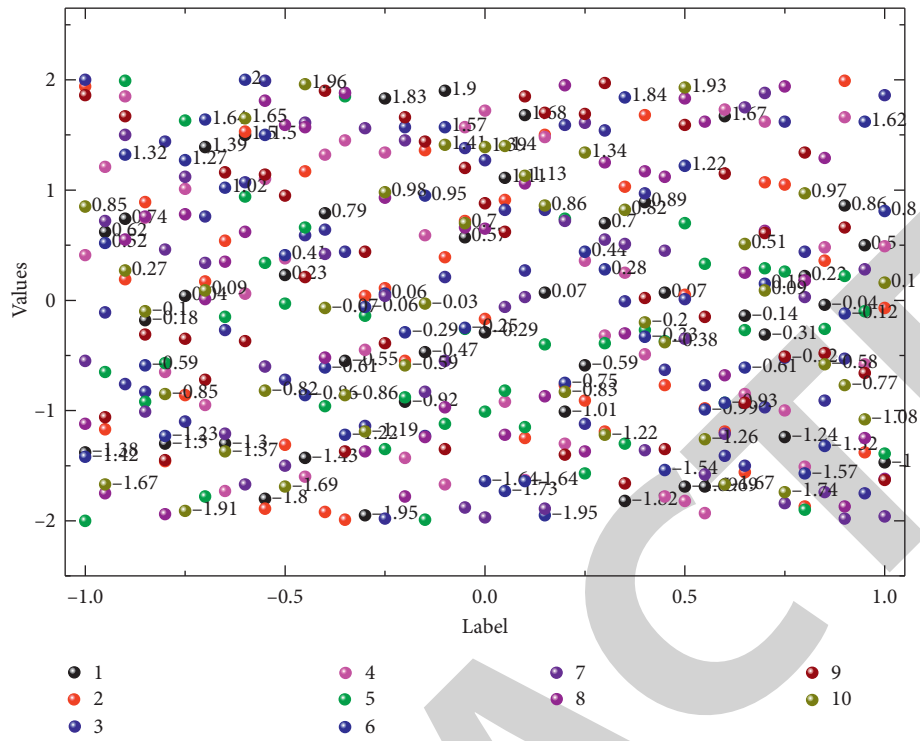

Figure 6: Calibration error analysis chart.

containing the target in the total video duration and calculating the recall at an intersection-over-Union ratio of 0.5 used to calculate the recall rate. Given that the badminton serve time does not exceed one minute, the total duration of the recorded video is set to one minute and thirty seconds, which is slightly longer than the server time. The badminton classification model was tested in three environments, and the results are shown in Figure 7.

The content of Figure 7 shows that the badminton classification model trained in this project has an average recall of 91.4% under different lighting environments and an average recall of 90.7% when the intersection ratio is greater than 0.5, which has a certain performance. The reason for the low average recall of detection is that the badminton is gradually pulled away in the video sequence, and the detection failure may occur when it is pulled away to a certain distance. The possible reason for the different performance of the classification model in different lighting environments is that no positive samples are made in multiple lighting environments. The racket classification model performs better in the artificial light environment and slightly decreases in low light scenes.

A possible reason for the difference in the quality of the training positive samples is the difference in the structure of the badminton and racket, which leads to the difference in the performance of the trained cascade classifier.

Figure 8 shows the performance of the system operating speed within 1 hour. This experiment uses MATLAB's fitting function to fit the trend graph of the operating speed, and all four curves are relatively smooth, and none of them have been declining and slipping for a long time, which proves that the system is relatively stable in terms of operating speed. In Figure 8, we are using python 3.6 for the analysis, and the final image that comes out is plotted using origin9.

However, from Figure 8 above, it can be found that between 10 min and 25 min on the French Open red clay court, there is a rapid decline in the running speed, but from 25 min to 35 min, there is a rebound, which indicates that the system has a jittering phenomenon during this period, but it does not affect the overall stability and is normal. The accuracy of our experimental results is increased by nearly 10% because our method is adaptive, which gives the most accurate results. A similar slippage occurred between 10 min and 20 min on the hardcourt of the Australian Open, after which the speed also rebounded and returned to the average state. Relatively speaking, the Wimbledon grass court was running more stable and had been in a relatively smooth state, which was a more ideal stable performance. The jitter phenomenon during system operation may be due to many factors: such as multiple processes are running on the same GPU at the same time; the computational resources are strained, resulting in speed degradation; the server runs too long and the temperature is too hot, resulting in speed degradation; the load is too large when the running program exceeds the processing limit of the server, also resulting in speed degradation; and the server GPU memory size and type are different; many reasons can affect the running speed of the system. The limitations of this study are mainly the inability to classify other sports or the ineffectiveness of the classification of other sports pairs.

The test time of 1 hour is limited, and it cannot be used as the basis for the stability of the system. Due to the time of the

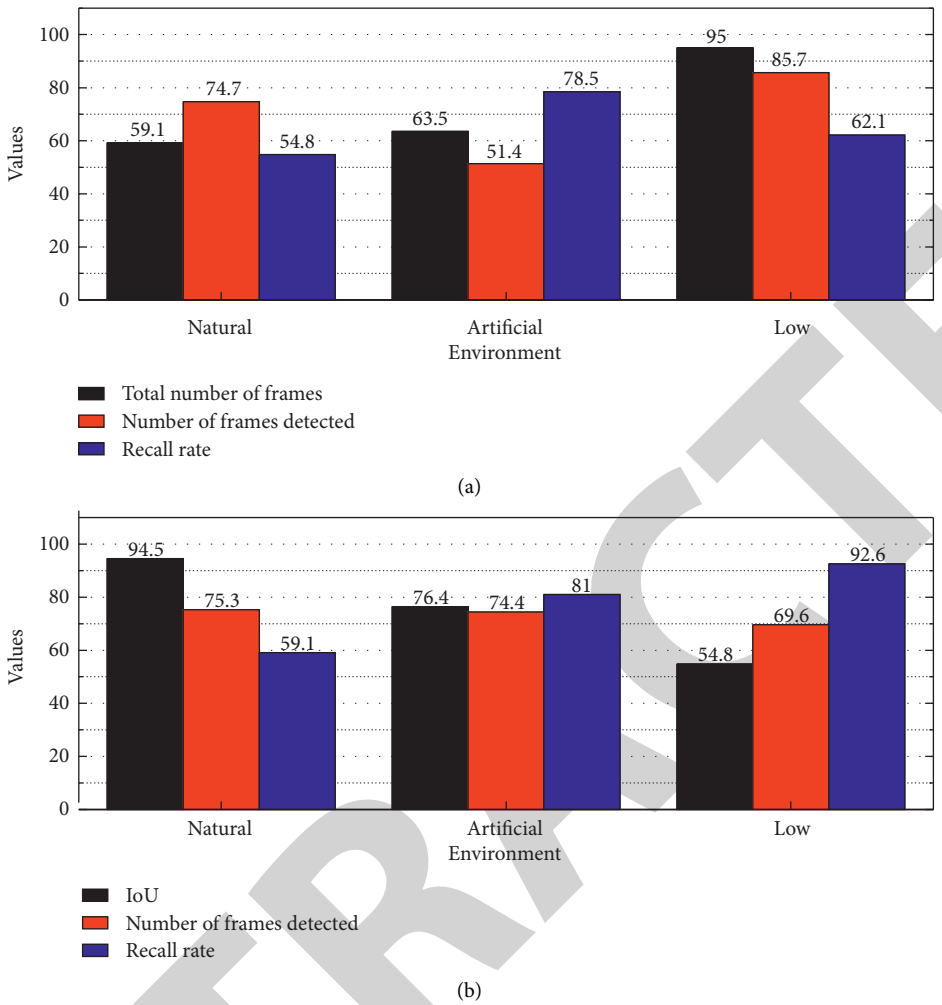

Figure 7: Badminton classification model detection statistics and recall rate.

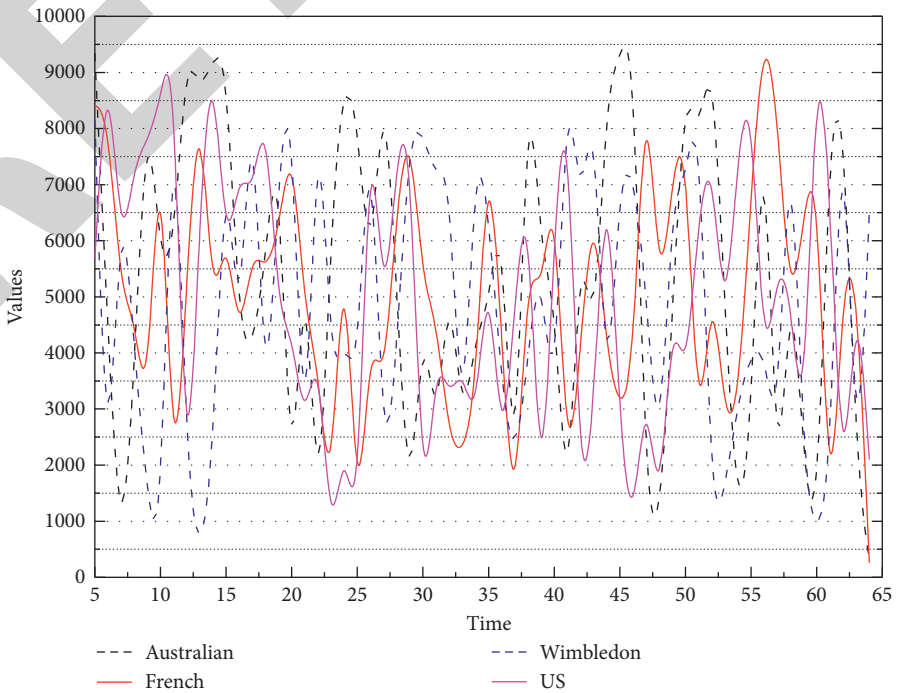

Figure 8: Stability performance graph of running speed.

experiment and the consumption of various resources, we cannot test the performance of the system stability for an infinite period, but we can infer the trend of the system running speed based on the existing experimental results, and continuing this trend, we can infer that the system running speed will be bumpy for a short period after one hour, but the overall view is stable.

4. Conclusions

This paper further describes the design and implementation of a prototype system for real-time semantic analysis of tennis match videos and provides a comprehensive analysis of each performance index of the whole system through experiments. The overall architecture design of the system is introduced first, and then the functional implementation of each submodule is presented separately. Finally, the real-time, accuracy, stability, and comprehensive impact factors on the system are demonstrated. The proposed algorithm is compared and analysed with existing and improved algorithms through quantitative analysis of target detection metrics, recall, and accuracy. The experimental results show that the algorithm proposed in this paper effectively solves the shortcomings of the above extraction, and the algorithm can have a better detection effect in many different scenarios and improve the real-time performance of the algorithm.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Consent

Informed consent was obtained from all individual participants included in the study references.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Yuwei Wang and Mofei Wen contributed equally to this work.

References

[1] M. Ponzano and M. Gollin, "Movement analysis and metabolic profile of tennis match play: comparison between hard courts and clay courts," *International Journal of Performance Analysis in Sport*, vol. 17, no. 3, pp. 220–231, 2017.

[2] G. Chen, X. Zhang, X. Tan et al., "Training small networks for scene classification of remote sensing images via knowledge distillation," *Remote Sensing*, vol. 10, no. 5, p. 719, 2018.

[3] T. Mildestvedt, O. Johannes Hovland, S. Berntsen, E. Tufte Bere, and L. Fegran, "Getting physically active by E-bike: an active commuting intervention study," *Health*, vol. 4, no. 1, pp. 120–129, 2020.

[4] G. Tian, W. Deng, Y. Gao et al., "Rich lamellar crystal baklava-structured PZT/PVDF piezoelectric sensor toward individual table tennis training," *Nano Energy*, vol. 59, pp. 574–581, 2019.

[5] S. A. Kovalchik, M. K. Bane, and M. Reid, "Getting to the top: an analysis of 25 years of career rankings trajectories for professional women's tennis," *Journal of Sports Sciences*, vol. 35, no. 19, pp. 1904–1910, 2017.

[6] Y. Iino, "Hip joint kinetics in the table tennis topspin forehand: relationship to racket velocity," *Journal of Sports Sciences*, vol. 36, no. 7, pp. 834–842, 2018.

[7] G. P. Lynch, J. D. Périard, B. M. Pluim et al., "Optimal cooling strategies for players in Australian tennis open conditions," *Journal of Science and Medicine in Sport*, vol. 21, no. 3, pp. 232–237, 2017.

[8] S. A. Kovalchik and J. Albert, "A multilevel Bayesian approach for modeling the time-to-serve in professional tennis," *Journal of Quantitative Analysis in Sports*, vol. 13, no. 2, pp. 49–62, 2017.

[9] D. Chadefaux, G. Rao, J.-L. Le Carrou, E. Berton, and L. Vigouroux, "The effects of player grip on the dynamic behaviour of a tennis racket," *Journal of Sports Sciences*, vol. 35, no. 12, pp. 1155–1164, 2017.

[10] 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.

[11] P. Le Noury, T. Buszard, M. Reid, and D. Farrow, "Examining the representativeness of a virtual reality environment for simulation of tennis performance," *Journal of Sports Sciences*, vol. 39, no. 4, pp. 412–420, 2021.

[12] B. Lane, P. Sherratt, X. Hu, and A. Harland, "Characterisation of ball degradation events in professional tennis," *Sports Engineering*, vol. 20, no. 3, pp. 185–197, 2017.

[13] J. Haut and C. Gaum, "Does elite success trigger mass participation in table tennis? an analysis of trickle-down effects in Germany, France and Austria," *Journal of Sports Sciences*, vol. 36, no. 23, pp. 2760–2767, 2018.

[14] Y. Inaba, S. Tamaki, H. Ikebukuro, K. Yamada, H. Ozaki, and K. Yoshida, "Effect of changing table tennis ball material from celluloid to plastic on the post-collision ball trajectory," *Journal of Human Kinetics*, vol. 55, no. 1, pp. 29–38, 2017.

[15] J. Zheng, T. Oh, S. Kim, G. Dickson, and V. De Bosscher, "Competitive balance trends in elite table tennis: the olympic games and world championships 1988–2016," *Journal of Sports Sciences*, vol. 36, no. 23, pp. 2675–2683, 2018.

[16] J. V. D. M. Leite, F. A. Barbieri, W. Miyagi, E. D. S. Malta, and A. M. Zagatto, "Influence of game evolution and the phase of competition on temporal game structure in high-level table tennis tournaments," *Journal of Human Kinetics*, vol. 55, no. 1, pp. 55–63, 2017.

[17] G. Elgado-García, J. Vanrenterghem, A. Muñoz-García et al., "Probabilistic structure of errors in forehand and backhand groundstrokes of advanced tennis players," *International Journal of Performance Analysis in Sport*, vol. 19, no. 5, pp. 698–710, 2019.

[18] N. Myers, W. Kibler, A. Axtell, and T. Uhl, "The sony smart tennis sensor accurately measures external workload in junior tennis players," *International Journal of Sports Science & Coaching*, vol. 14, no. 1, pp. 24–31, 2019.

[19] J. Wen, J. Yang, B. Jiang, H. Song, and H. Wang, "Big data driven marine environment information forecasting: a time series prediction network," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, 2020.

[20] D. Ndahimana, S.-H. Lee, Y.-J. Kim et al., "Accuracy of dietary reference intake predictive equation for estimated energy

requirements in female tennis athletes and non-athlete college students: comparison with the doubly labeled water method," *Nutrition Research and Practice*, vol. 11, no. 1, pp. 51–56, 2017.

[21] M. King, A. Hau, and G. Blenkinsop, "The effect of ball impact location on racket and forearm joint angle changes for one-handed tennis backhand groundstrokes," *Journal of Sports Sciences*, vol. 35, no. 13, pp. 1231–1238, 2017.

[22] J. Yang, C. Wang, H. Wang, and Q. Li, "A RGB-D based real-time multiple object detection and ranging system for autonomous driving," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11959–11966, 2020.

[23] B. Yang, X. Cheng, D. Dai, T. Olofsson, H. Li, and A. Meier, "Real-time and contactless measurements of thermal discomfort based on human poses for energy efficient control of buildings," *Building and Environment*, vol. 162, p. 106284, 2019.

[24] J. Wang, Y. Liu, S. Niu, and H. Song, "Beamforming-constrained swarm UAS networking routing," *IEEE Transactions on Network Science and Engineering*, p. 1, 2020.

[25] H. Ge, Z. Zhu, K. Lou et al., "Classification of infrared objects in manifold space using Kullback-Leibler divergence of Gaussian distributions of image points," *Symmetry*, vol. 12, no. 3, p. 434, 2020.

[26] S. Xia, D. Peng, D. Meng et al., "A fast adaptive *k*-means with no bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2020.

[27] J. Yang, T. Liu, B. Jiang et al., "Panoramic video quality assessment based on non-local spherical CNN," *IEEE Transactions on Multimedia*, vol. 23, pp. 797–809, 2020.