Research Article

Mathematical Modeling and Simulation of Table Tennis Trajectory Based on Digital Video Image Processing

Chuan He,[1,2] **Xiaoquan Zhang** [1] **Yulong Gui,**[3] **Yang Liu,**[1] **and Wei Zhang**[1]

[1]*School of Kinesiology and Health Promotion, Dalian University of Technology, Panjin, 124221 Liaoning Province, China*
[2]*Moscow State Academy of Physical Education, Moscow 140043, Russia*
[3]*College of Physical Education, South-Central University for Nationalities, Wuhan, 430000 Hubei Province, China*

Correspondence should be addressed to Xiaoquan Zhang; xiaoquanzhang@dlut.edu.cn

Digital sports training based on digital video image processing promises to reduce the reliance on the experience of coaches in the table tennis training process and to achieve a more general physical education base. Based on this approach, this paper describes the specific forms of exercise content, movement characteristics, and skill levels in the table tennis framework and specifies the calculation methods of motion capture and movement characteristics suitable for table tennis. Meanwhile, to further improve the accuracy of the inertial motion capture system in restoring the position posture of the trainees, this paper improves the original inertial motion capture system from two aspects: contact judgment of both feet and correction of the position posture based on the contact position constraint. The simulation results show that the corrected human posture has good action smoothness. This paper first proposes a knowledge-based generic sports-assisted training framework based on generalizing the traditional sports training model. The framework contains four main modules: domain knowledge, trainees, sport evaluation, and controller. The domain knowledge module is a digital representation of the knowledge of the exercise content, improvement instructions, and skill indicators of the sport; the trainee module is the active response of the trainee to the exercise content and improvement instructions; the motion evaluation module uses motion capture technology to obtain the raw motion data of the trainee and further calculates the motion characteristics; the controller module proposes improvement instructions to the trainee or makes him/her practice new content based on the results of the motion evaluation. Based on the results of the motion evaluation, the controller module proposes improvement instructions or makes the trainee practice new content until the trainee achieves the desired goal.

1. Introduction

In sports training, the function of the trainer, who is the core, can be boiled down to two parts: using the eyes, ears, and other senses to observe the training content and the performance of the trainee and using the brain to process the information obtained from the observation and then give feedback to the trainee in terms of training content and training instructions. In a digital automated training system, motion capture technology can be used to record the training process digitally, while accurate modeling and computation can analyze the training effects and give appropriate guidance and feedback [1]. The raw motion information obtained by relying on motion capture devices is more accu-

rate and transparent than the naked eye observation of coaches and is also easier to record, preserve, and analyze later, while automated evaluation and training based on this information can reduce the reliance on coaches' experience. Therefore, research on digital recording and automated training in sports is of profound value.

In the training process, table tennis players are often able to hit the ball speed is super fast, the landing point is variable rubbers, out-of-bounds balls, controversial balls, etc., but the drop point of table tennis is instantaneous and fleeting, and even professional coaches are often unable to accurately identify the real situation of the drop point; moreover, athletes have to conduct a lot of training every day, using human resources to record and analyze these drop point

information is often laborious and also unable to achieve good results; therefore, using a drop point recognition system can help athletes record their training data during the training process, including ball trajectory, ball drop point, and drop point analysis [2]. These data can help athletes understand their training situation and make targeted technical improvements to objectively help them improve their competitive level. The traditional fusion algorithm: the dual-mode image fusion algorithm based on multiscale transformation is analyzed, the existing drawbacks are presented, and a dual-mode image fusion algorithm based on deep learning is proposed to address these drawbacks, which improves the accuracy of foreground and background extraction and solves the problem that the fusion rules in the traditional fusion algorithm rely on manual design. Experimental results on public datasets show that the algorithm proposed in this paper outperforms the traditional fusion algorithm under several evaluation metrics. In traditional drop point recognition systems, it is often necessary to resort to a large number of sensors, such as vibration sensors, sound sensors, and pressure sensors. This approach is not only hardware intensive but also cumbersome and complex to implement; so, it is often difficult to apply on a large scale. In recent years, machine vision and image processing technology flourishes, especially that the target detection, tracking, and other technologies become more and more mature, so that the development of the falling point recognition technology gets more support, not only the recognition accuracy of the system is more and more improved but also the cost of consuming hardware equipment decreases continuously. In the future, the combination of image technology and falling point recognition technology will be more closely. In the development of modern competitive sports, gradually present high, difficult, and precise as well as sharp characteristics, in this environment, the application of modern technology means in sports training is also more important. To achieve the maximum excavation of human potential, it is also necessary to increase the application of science and technology [3]. Based on this, it is also necessary to realize the comprehensive application of knowledge of sports science-related disciplines and to implement systematic and comprehensive research on the characteristics of the inner laws of sports.

2. Related Work

Traditional motion target detection algorithms contain background subtraction, interframe differencing, and optical flow methods. All of them have achieved good results in their respective suitable scenes, but all of them also have shortcomings. With the development of computer vision theory and the improvement of computer computing performance, many scholars have made a lot of improvements to the above algorithms in combination with actual scenes, such as the three-inter-frame difference algorithm based on edge features, which solves the problem of incomplete target region voids. The introduction of optical flow information in the interframe difference method improves the accuracy of detection but increases the complexity of computation. In

complex environments, such as dynamic background, light changes, and similar interference, all affect the detection effect. The background subtraction method is favored by researchers for its accurate detection, complete extracted target region, and good anti-interference effect and has been studied and innovated more intensively and widely used in engineering. The performance of this method depends on the selection and construction of the background, and the algorithms of the background modeling class are usually classified into two categories: with and without reference. Literature [4] proposed a method based on mixed Gaussian background modeling GMM (Gaussian mixture model) based on summarizing the law of statistical distribution of pixels, which has excellent detection effect in the case of relatively small background changes, but the algorithm is computationally intensive, high complexity and is a practical scene cannot guarantee real-time; so, it is difficult to promote the use.

The literature [5] uses an adaptive approach to select the number of Gaussian models to further improve the efficiency of the algorithm, which is essentially an improvement of the hybrid Gaussian model approach. The hybrid Gaussian model requires manual adjustment of many parameters, and the debugging of parameters directly affects the experimental results. To solve this problem, literature [6] proposed a modeling method KDE (kernel density estimation) for nonparametric kernel density estimation, which works well in complex environments with high target detection accuracy, but the algorithm requires a large amount of storage space and high computational complexity, making it impossible for engineering applications. A novel vibe background modeling algorithm is proposed in the literature [7], wherein in the initialization phase, an independent set of background models is first created for each pixel point, and then the model set is filled with randomly selected pixel values from the first frame in the video. In the execution target detection phase, the difference between each pixel point of the current frame and its corresponding pixel point in the model set is calculated to determine whether the pixel point belongs to the stationary background or the moving foreground, and the foreground target region is extracted. The Vibe algorithm is simple and fast to process, and this algorithm has good detection results even in the environment of dynamic background interference or slight jitter. This simple and fast background modeling method and fast and effective foreground determination method are more suitable for motion target detection in real-time scenes. Currently, the Vibe algorithm has become the mainstream algorithm for research in this field because of its many advantages and good performance. In recent years, the pose estimation efficiency problem caused by the high-dimensional properties of human motion data has started to gain attention. The literature [8] proposed the use of locally linear embedding to reduce the dimensionality of the data. The stochastic Gaussian process hidden variable model proposed in the literature [9] has achieved good results in reducing the dimensionality of data under complex motion. In [10], after treating the stored data of high-level athletes as "full scores," the trainee's sports performance score is given explicitly by comparing

the data of ordinary trainees with the data of high-level athletes. Further, the literature [11] also selects the videos of the best athletes with the most similar scores to the trainees for each sport characteristic and recommends them to the trainees as learning targets. In addition, the literature [12] provides a conceptualization on how to better monitor and analyze the data, respectively, including how to choose the type of data to be monitored in sports and how to process the data after monitoring, which provides a guideline for digitization of general sports training. The literature [13] uses an inverse reinforcement learning model to learn basic stroke strategies in table tennis, and a reward model (REM) obtained by training from sports data can extract key information (e.g., speed, and landing point) in high-level strokes. The deep feature flow algorithm published in the literature [14] uses flow estimation and flow propagation to extract CNN features of keyframes, then passes the depth features from keyframes to other frames, and uses optical flow network to pass data between keyframes, which reduces the computation and improves as the algorithm only processes keyframes detection efficiency. The literature [15] focuses its research on the software level by assuming the human skeleton as an articulated steel structure, using offline optimization methods to determine the skeleton length and using inverse motion chains to generate motion trajectories.

3. Mathematical Modeling and Simulation of Table Tennis Trajectory Based on Digital Video Image Processing

3.1. Digital Video Image Processing Algorithms. The so-called target detection technique is to distinguish the region of interest from the region of disinterest in a video file or image, qualitatively describe whether there is a target in the image, and locate the target location. Target detection is currently an important research direction in the field of computer vision. With the continuous updating of Internet technology, video processing technology, and hardware computing power, human life is filled with a large amount of video and image data, and the target detection scenarios are becoming richer and richer, which has led to the continuous development of target detection technology and the improvement of detection accuracy and efficiency [16]. The current leading target detection algorithms have been able to achieve real-time detection of targets. Contour features involve relatively few feature points; so, they also result in speed gains. Experiments have shown that such features tend to enhance the tracker's effectiveness in tracking targets in the presence of partial occlusion. Target detection technology is not only a cornerstone in the field of intelligent video surveillance but also has a wide range of applications in other fields, such as image retrieval, web data mining, remote sensing image release analysis, medical image analysis, and other industries.

The traditional target detection algorithm is based on feature engineering, and its core idea is to extract the features of the target in the candidate frame using manually constructed features and complete the classification with a classification algorithm after extracting the features. The basic detection process of the traditional target detection algorithm is shown in Figure 1 below.

The typical pattern of traditional target detection algorithms is to use manual feature construction algorithms plus classification algorithms to accomplish the target detection task. Among them, feature construction algorithms include HOG features, SIFT features, HAAR features, and LBP features, and classification algorithms include SVM algorithm and AdaBoost algorithm. Next, the AdaBoost algorithm is introduced in detail. Next, the AdaBoost algorithm is introduced in detail. The AdaBoost algorithm adopts the idea of iteration, where one weak classifier is trained in each iteration, and the weak classifier that completes the training will participate in the next iteration of training. After each round of training, the probability of the distribution of correctly classified data is reduced, the probability of the distribution of incorrectly classified data is increased in the dataset, and the next round of training is for incorrectly classified data. Eventually, N weak classifiers are combined into a strong classifier for feature classification [17]. Figure 2 shows the structure of the AdaBoost algorithm.

The 3D reconstruction of an object is the process of computing and restoring the 3D stereo geometric information of an object based on a multiview 2D picture of the same object. The derivation of the 3D reconstruction method is first based on the imaging principle of the camera, i.e., the small-aperture imaging model, which corresponds to the process by which the camera captures an image in real space: first, the camera maps the object points in the object space onto the film plane, and then, to generate the final seen planar image, the points recorded on the film plane will be projected again onto the final projection plane, which eventually constitutes the points on the imaging screen. To simplify this process further, the principle of camera imaging can be seen as a mapping of points in real space to points on image space, by directly relating the projected image to real space objects.

The vector from N to O can be expressed as vector A. By adding the axis W perpendicular to the plane to the image plane reference system, the plane reference system becomes a three-dimensional reference system, and then the W coordinates of the points on the plane are always 0. The point P is the projection of the point N on the plane along the W axis, and the straight line distance between them constitutes the principal distance. The rotational motion of a ping-pong ball in the air is directly affected by the Magnus force. From the perspective of the fluid dynamics, the analysis of the flight of a rotating object in the air, the angular velocity vector of rotation, and the flight velocity vector that does not coincide in the formation of the plane will appear a transverse force and this side of the vertical relationship, resulting in the deflection of the trajectory of the object in flight, and the transverse force is the Magnus force. Therefore, point P is also called the principal point. From point N to point I, vector B can be formed, and since O, I, and N are colinear, vector A and vector B are related as shown in Equation (1).
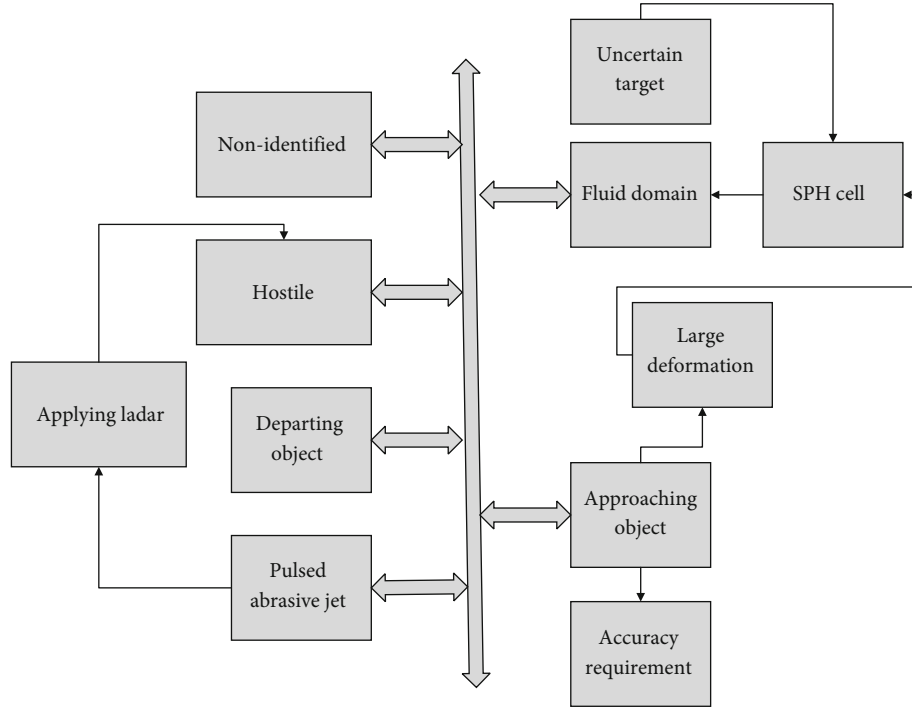
$$B = 1.34\text{c}A^T. \tag{1}$$

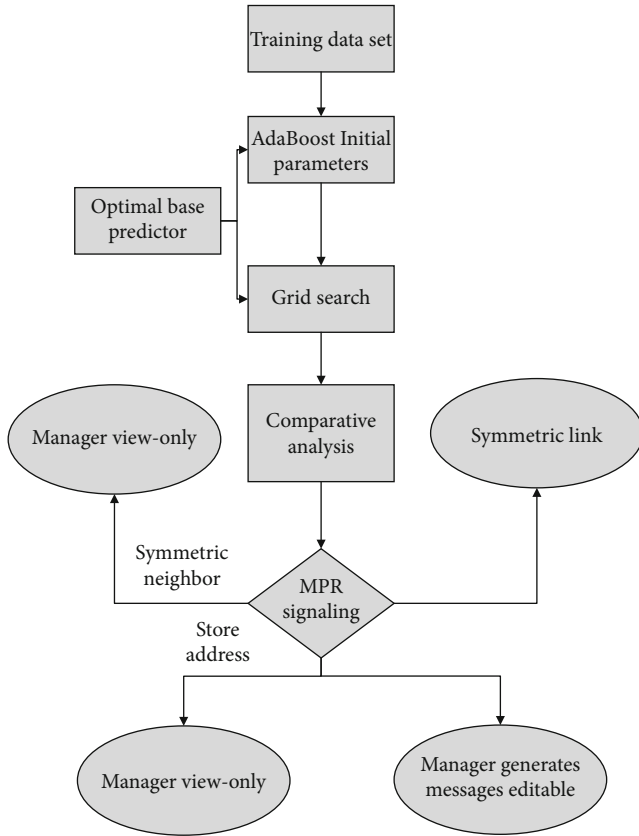FIGURE 1: Traditional target detection process.

FIGURE 2: AdaBoost algorithm flow.

The vectors A and B are under two separate reference systems and need to be transformed to the same reference system to describe them. Here, the vector A is transformed to the plane reference system, where $A(I)$ is the A vector under the image plane reference system, $A(O)$ is the A vector under the real space reference system, the matrix T is the transformation matrix, and the vectors are related under different reference systems as shown in Equations (2) and 3.

$$T_{I/O} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \tag{2}$$

$$A^T = B \cdot T_{I/O} \cdot B^{-1}. \tag{3}$$

The A vector, after conversion to the image plane reference system, is brought into Equation (1), and the resulting object space coordinates are related to the image plane coordinates as shown in Equation (4).

$$c = \frac{-b}{2a\sqrt{r_{12}(x - x_0)^2 + r_{23}(y - y_0)^2 + r_{13}(z - z_0)^2}}. \tag{4}$$

Note that the image plane reference system units for u, v, u_o, and v_o are realistic units of measure, such as centimeters, and therefore need to be converted from realistic length units to pixel units for image imaging by further bringing in the coefficients λ

$$u = \lambda k(x - y - z) + v. \tag{5}$$

By extracting the parameters from Equation (5), the relationship between the coordinates (u, v) and (x, y, z) can be

expressed in the form of Equation (6), where Δu and Δv are errors such as lens distortion of the camera.

$$(u, v) = \lambda(x, y, z) \frac{(x_0, y_0, z_0)}{\Delta u \cdot \Delta v}. \tag{6}$$

If the parametric error of the lens is introduced, Equation (6) can be transformed into the form of Equation (7), where R is the term shown in Equation (8).

$$R_k = \sup \left\{ \lim_{k \to \infty} \sum_{i=1}^{k} f_k(w_k) \right\}, \tag{7}$$

$$Q_k = \frac{\sum g_i y_{ik}^2}{Cx}. \tag{8}$$

From Equation (7), L1 to L16 are the correspondence between the real space and the image space points, where L1 to L11 are called DLT parameters, and L12 to L16 are called additional parameters. To solve the DLT parameters by the least-squares method, it is necessary to construct a system of overdetermined equations to solve the 11 unknowns, i.e., the system of equations needs to satisfy that the number of equations is greater than the number of unknowns [18]. Therefore, by finding at least 6 groups of corresponding points including noncoplanar feature points to form the corresponding control volume in real space, and bringing the corresponding points in the image plane and real space into Eq. (7) and solving them in conjunction, the corresponding DLT parameters can be obtained.

After obtaining the DLT parameters for each shot, the coordinates of the point in real space can be reconstructed if the coordinates of the point in the 2D plane are also known for the same real space recorded in each image. Equation (7) can be further transformed into the form shown in Equation (9), where x and p are as shown in Equation (10). By bringing the DLT parameters of the cameras and the 2D plane coordinates into Equation (9), the coordinates in the real space can be obtained.

$$(V_{ij})_{mn} = \begin{pmatrix} X_1 p_{11} & X_1 p_{21} & \cdots & X_1 p_{m1} \\ X_2 p_{12} & X_2 p_{22} & \cdots & X_2 p_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ X_m p_{1n} & X_m p_{2n} & \cdots & X_m p_{mn} \end{pmatrix}, \tag{9}$$

$$V = (X_i P_{ij})^{mn}. \tag{10}$$

Expansion can merge boundary points, expand the target region boundary, and is generally used to fill voids in the target region and eliminate noise from small particles. The formula is as follows.

$$H(x) = \lim_{n \to \infty} \sum_{i=1}^{n} p_i \lg \frac{1}{p_i}. \tag{11}$$

Based on the virtual environment, each joint needs to be treated differently. Each step of the simulation for the joint has a binding force that must act on the rigid body, in the application of the joint constraint equation can achieve the calculation of the dynamics engine binding force, for the rigid body in motion and all joint correlation play a role in maintaining. The usual representation of the joint constraint equation is as follows.

$$e_j = -k \lim_{n \to \infty} \sum_{i=1}^{n} p_i \lg \frac{1}{p_i}. \tag{12}$$

The collision of a ball and a racket is different from the simulation of a collision between steel, and there can be a penetration phenomenon. The collision of a ball and a racket is bound to have a deformation, and the correct simulation of the collision process will also have a penetration phenomenon. For the problem of inaccurate contact judgment in inertial motion capture, a set of plantar contact platforms based on pressure measurement is built. A set of pressure acquisition circuits is designed according to the characteristics of pressure sensors and the basic features of human motion, and a stable and efficient plantar pressure acquisition is realized by the carrier of homemade insoles. For the problem that there are differences in contact judgment thresholds for different weight groups and different sports movements, this paper proposes a set of adaptive methods for generating contact judgment thresholds, and the resulting contact judgment results are consistent with those observed in high frame rate videos. Therefore, for the implementation of the simulation process of not completely rigid objects, there must be the existence of noncompulsory constraints, or by the existence of flexible constraints (CFM), the collision of two objects in the process of allowing a certain penetration phenomenon, for example, in the simulation of a small elastic ball and the ground collision process, set the corresponding penetration more can enhance the simulation of realism. By setting the racket elasticity coefficient to K_p and the damping coefficient to K_d, the ERP and CFM are calculated as follows.

$$F = B^2 V^T K_p K_d. \tag{13}$$

Particle filtering is based on Kalman filtering and uses prior probabilities to assign weights to each target sample. The larger the weight assigned to the particle, the higher the importance of the sample, which also means closer to the true target. After that, the weights of the particles are updated according to the state transfer equation of the Kalman filter and finally, the particles are resampled based on their importance, and the particles are diffused according to their distribution and the result of the diffusion, i.e., the state of the tracking target is normalized to this result [19]. The particle filtering method compared to most tracking algorithms to a certain extent to alleviate the impact of target occlusion, in the speed, also has a certain advantage, so in the actual engineering has more and more applications.
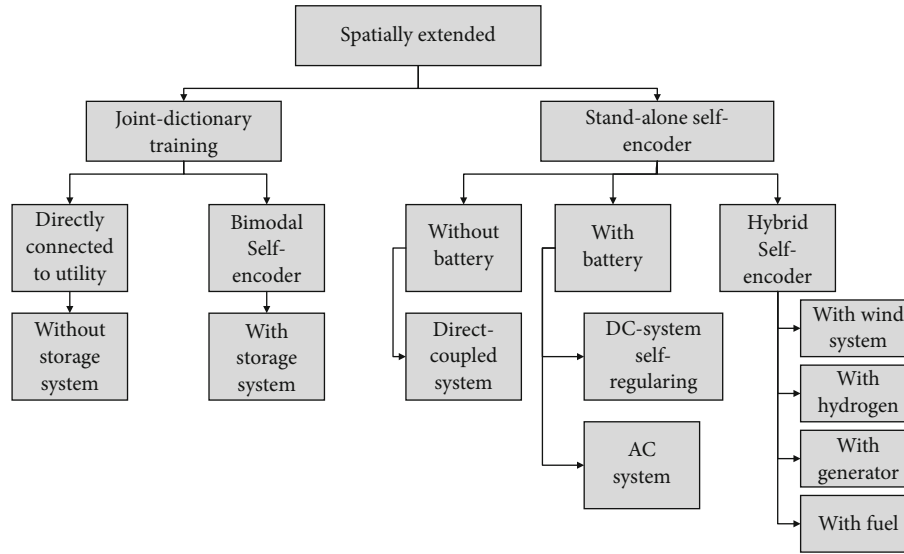
FIGURE 3: Flow chart of fusion prediction.

Several feature extraction methods are commonly used in classical tracking algorithms.

(1) Color histogram features based on target regions: color features have the properties of being unaffected by changes in the shape and contour size of the target object, the property of rotational invariance, and the property of being roughly equally distributed in color space, but perform less well for cases where the target and background colors are similar

(2) Contour features of the target: contour features involve relatively few feature points; so, they also result in speed gains. Experiments have shown that such features tend to enhance the tracker's effectiveness in tracking targets where partial occlusion exists

(3) Texture features of the target: compared to the contour features of the target, the features of the target texture are more representative and can be less disturbed by the deformation of the target; so, the tracking effect is improved compared to the algorithms using contour features. Some subsequent algorithms also do various combinations of the above basic features appropriately to optimize the tracking effect in different scenarios. With the increasing popularity of deep learning algorithms, compared to previous methods, deep learning-based target detection is faster and more representative for feature extraction; the better the feature extraction, the more accurate the target will be matched in the tracking phase

Since the training data is completely different from the test data, it is impossible to predict the size of the targets in the test data when the model is put into application, which will result in the model having different detection performances for targets of different sizes. The virtual table tennis system consists of real and virtual scenes: the real scenes include real rackets and tables, and the virtual scenes are dis-

played in three dimensions by projection and include virtual opponents, virtual balls, and virtual tables. Participants wear stereoscopic glasses and hold a special paddle to play an immersive ping pong match with a lifelike virtual opponent on the screen. Generally speaking, most networks are better at detecting large targets, but have a large gap in detection for small targets, or even fail to detect smaller targets. For this difficulty, relevant training data can be added, and also multiscale fusion prediction can be performed by fusing the features extracted from the images. Figure 3 shows the flow chart of fusion prediction.

Sometimes overlap between targets may occur due to shooting angles or crowding, etc. This phenomenon occurs when detecting scenes in which multiple targets with a large degree of overlap are often detected as the same target, resulting in a lower check-all rate, or recall, of the targets, which indirectly affects the accuracy. For solving this problem, an appropriate loss function can be considered to increase the penalty for overlapping target detection errors in the backpropagation process. The training of target detection models varies from scenario to scenario and from need to need. Although speed and accuracy are equally important. However, some applications that have been landed so far prioritize speed slightly higher while guaranteeing accuracy to meet expectations, so in terms of network selection, a lightweight, end-to-end detection network can be used. The size of the model is as small as possible to facilitate porting to embedded devices such as chips and boards. Although the size of the model trained by the lightweight network has been optimized, it is still possible to consider compressing the model by quantization, cropping, and other operations to make the model run in real-time on the computer.

3.2. Mathematical Modeling and Simulation of Table Tennis Trajectory Based on Digital Video Image Processing. The virtual table tennis system contains both real and virtual scenes: the real scenes include real rackets and tables, and the virtual scenes are displayed in stereo by projection and include

virtual opponents, virtual balls, and virtual tables. Participants put on stereoscopic glasses and hold a special paddle to play an immersive table tennis match against a virtual opponent that comes to life on the screen. The paddle is equipped with a 6-degree-of-freedom magnetic sensor and a vibrator that senses the paddle's three-dimensional position (X, Y, Z) and orientation (Yaw, Roll, Pitch) [20]. The magnetic sensor transmits the data collected in real-time to the computer, and after analysis and processing, calculates the position, direction, and speed of the paddle, to determine whether the paddle hits the virtual table tennis and the direction and strength of the hit, then controls the trajectory and landing point of the virtual ball, and finally determines whether the virtual opponent drops the ball according to the difficulty and the landing point of the ball selected by the user and decides the expression and action of the virtual opponent according to the score. After the system judges that the racket has hit the ball, it will make the human hand feel the force of touching the ball through the vibrator inside the racket. The virtual scene is created by simulating the visual image of two eyes and projected onto the same screen by two projectors, which are finally separated by the polarized glasses worn by the participants to obtain a three-dimensional feeling.

The forces acting on a game of table tennis during its run include air resistance, gravity, and Magnus force [21]. In the analysis of the force situation, the force of gravity in the running of the table tennis ball is vertical downwards in mg, $G = 9.81\,\text{m/s}^2$. The direction of the air resistance force is opposite to the direction of the running of the ball, the air density ρ, the drag coefficient C_d, the windward area of the ball A, and the linear velocity v. All the indicators have a direct impact on the air resistance, and the calculation formula obtained by combining aerodynamics is expressed as follows.

$$F = C_d \cdot \int G^\rho \cdot v(t) dv. \tag{14}$$

In the air drag coefficient detection needs to be obtained through wind tunnel experiments, in the process of this study, it is assumed that the drag coefficient is influenced by only two factors, which are translational speed and rotational speed, and the calculation formula obtained based on aerodynamic analysis is as follows.

$$C_d = \frac{25.1016}{0.1979 + (v/\omega)^{7/3}} + s. \tag{15}$$

During the rotational motion of a ping-pong ball in the air, it is directly subjected to the Magnus force. From the hydrodynamic viewpoint, the flight of a rotating object in the air, the angular velocity vector of rotation, and the velocity vector of flight do not coincide in the plane formed by a transverse force, and this side is vertical, thus causing the deflection of the trajectory of the object in flight, the transverse force is the Magnus force. The presence of this force can be realized to explain the generation of curved balls in

ball sports such as table tennis, football, and tennis. In a game of falling table tennis, the surrounding air has a direct effect on the surface of the ball, while the surface of the ball also acts on the surrounding air, showing different flow lines, as can be seen from the diagram above the ball flow lines are denser compared to the bottom; so, it can be obtained that the flow velocity above is higher, and the two forces interact in the process of pressure direction is upward. During the upward rotation of the ball, the direction of pressure is exactly opposite to the downward rotation, i.e., the downward force.

According to the study of the forces during the operation of the table tennis ball, it is found that the translation speed is v, the Magnus force is Fm, the air resistance is Fd, the gravity is G, the angle between v and the plane XOY is α, the angle between the projection of the translation speed vector on the XOY plane and the X-axis is ϕ, the angle between the Magnus force and the XOY plane is β, the angle between the projection of the Magnus force on the XOY plane and the X-axis is θ, and with the application of Newton's second law $F = ma$, the following equation can be obtained.

$$F(\rho) = \int \phi^{-i2\pi\theta\omega x} f(x) dx. \tag{16}$$

In the model construction, the mathematical modeling of the motion trajectory of the table tennis ball and the simulation is completed based on the above model construction.

4. Experimental Verification and Conclusions

Each trainee was asked to receive three different difficulty serves (slight topspin, strong topspin, and backspin) to differentiate between trainees' levels. For a specific difficulty, the tee was set to serve at a fixed point and to the trainee's forehand position to ensure that the lowest level trainee could effectively receive the ball at the lowest difficulty. The trainees were required to make 7-10 catches to the diagonal, center edge, and ipsilateral edge corners of the opposite table, respectively, while prioritizing that the catch would land on the table. That is, each trainee needs to catch approximately: 3 difficulty levels × 3 target drop points × $(13 - 17) = (117 - 153)$ balls. In this case, the three target drop points were set to further differentiate the levels between trainees. In general, it was easier to return to the middle of the opposite table than to the edge and to the opposite corner than to the edge corner on the same side. The 13-17 serves per drop point were set to reduce the impact caused by trainee errors. Considering the cost of the experiment, in addition to the design of the above experimental components, the experiment was implemented with the requirement of synchronizing the optical and inertial systems and recording as much information as possible. In this paper, a migration learning approach is used to fine-tune the parameters on a dual-mode image dataset using the weights of a target detection model pretrained on a large dataset. Since the distribution of target image data for target detection on bimodal fused images is the concatenation of infrared image

dataset and visible image dataset, the training dataset needs to encompass both infrared and visible images. In this experiment, we specify that the trainee strikes the table with the paddle at the beginning and end moments of the experiment to synchronize the timing of the two systems by making the paddle rigid body in the optical system and the wrist sensor in the inertial system have significant acceleration at the same time; we specify the initial position of the trainee during the recording in the inertial system to simplify the positional transformation of the two systems; we specify the initial position of the ping pong paddle during the recording in the optical system to simplify the paddle The posture calculation of the.

Only raw motion data can be obtained from the motion capture system, i.e., the position, velocity, acceleration, angular velocity, and stance of each human skeleton at all times, the 3D trajectory of the ball, and the position and stance of the racket. What is more important for motion evaluation is the motion data for each round or even each phase of the round, from which advanced characteristics such as skill metrics can be assessed. The complete process of processing from raw optical and inertial motion capture data to a single round even performance can be divided into optical data preprocessing, data segmentation and complementation, serve and catch-return judgment, and feature calculation after a successful catch-return.

The motion data from the 21 bones output by the inertial motion capture system is already of high quality and can be used directly for data segmentation. In contrast, because of spatial clutter, the coexistence of multiple balls, object occlusion, etc., there is a loss of racket data obtained directly from the optical system (especially in the few frames near the touching ball), and a complete table tennis track is often recognized by the system as multiple segments. If the manual correction method is used, the processor needs to check the video against the optical lens one by one to identify the different segments belonging to the same table tennis, stitch them together, then remove stray points and remark the valid reflective points for identification and reading. The whole process is laborious and error-prone.

This paper uses a self-developed computer program that allows automatic and efficient data stitching, stray point removal, and tagging for optical motion capture. The program requires that the initial segments of all valid reflective balls be marked first, and then the initial frames of unmarked segments and the end frames of marked segments are compared using the spatiotemporal characteristics of the flight of table tennis. If the difference between the two frames of data is small in time and the velocity calculated from the position after stitching has good continuity, the two segments are considered to correspond to the same table tennis flight trajectory and are stitched together. In addition, the program can determine and crop uninformative data segments of the table tennis, for example, by detecting the ball speed and height from the ground in all segments to remove segments with essentially zero ball speed as well those with significantly lower height than the table. In this way, the optical motion capture data processed using the program can be used for the next step of data segmentation.

The effect of optical data processing is shown in Figure 4, indicating that untagged data segments are cut to the corresponding tagged segments according to spatiotemporal continuity over a time range of 600 frames (96 frames/second).

After the above data processing, the typical serve and return trajectories for the three serve patterns of slight topspin, strong topspin, and bottom spin can be obtained. The single-round table tennis trajectories corresponding to the slight topspin, strong topspin, and bottom spin serves are shown in Figure 5, where the X-axis of the coordinate system is along the long side of the table, pointing from the trainee to the direction of the server machine, the Y-axis is vertically upward, and the Z-axis is pointing to the direction of the short side of the table according to the right-hand rule. Since the system requires low real-time performance, the fusion algorithm with the best performance is selected: a deep learning-based IR and visible image fusion algorithm, which fuses the read source IR and visible video frame by frame, and each fused image is input to the target detection network for subsequent processing. The four key points in a single round, such as serve, serve down the table, strike, and return bounce, are indicated at the corresponding locations in the figure. The position of the inflection point not marked at the end of the curve indicates that the return ball returns to the pocket after impacting the recovery net, as can be seen by the gradual return to stability after the X-axis coordinates of the ball reach an extreme value well above the initial point. Intuitively, the three trajectory curves are not very different, except that in the strong topspin serve pattern the trainee needs to rub the ball more with the racket in the return, and accordingly, the trajectory changes more slowly near the point of impact.

The above are just the ideal case of serve and return trajectories. The more specific determination of the serve and receive-return situation is crucial for both the ensuing action recognition and feature calculation, as well as for more detailed guidance feedback in the future. Based on the relationship between the table tennis and the table (where it lands on the table), the net (number of times it crosses the net), and the racket (whether it hits or not). Among them, the contact or not of the table tennis with the table is judged based on the distance between the ball's position and the table and whether the velocity direction changes after touching the table; the passing of the table tennis over the net or not is based on the distance between the ball and the net plane and the over height; as for judging whether the paddle hits, the two conditions that the ball is close to the paddle and the velocity of the ball changes significantly after hitting need to be satisfied at the same time.

The two main causes of common serve errors in the experiment are when the service is stopped by the net and when the initial landing point is on the side of the tee. The former can be seen in the fact that the trajectory on the X-axis (along the long side of the table) only reaches half the length of the table, while the latter is reflected in a premature bounce on the Y-axis (vertically upwards), where the trajectory is "depressed" on the X-axis because the ball slowly rolls back to the tee side after bouncing off the net. The common types of errors in the catch and return are more frequent,
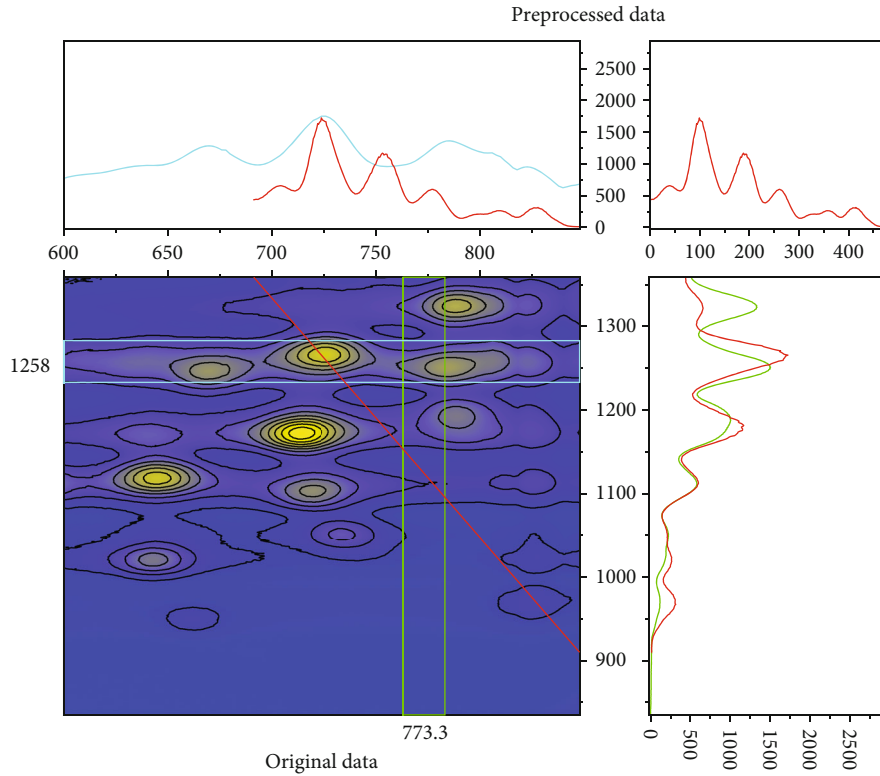
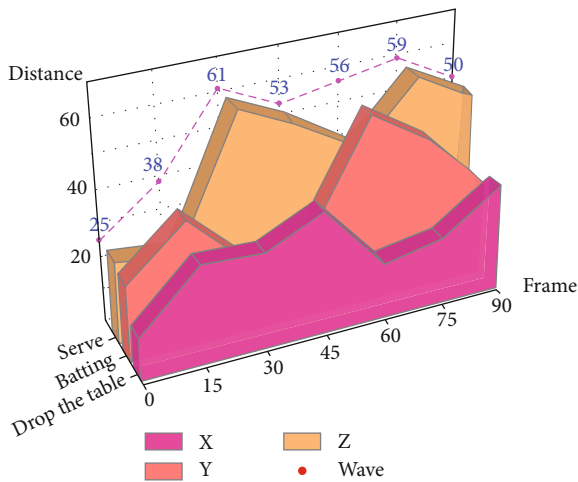FIGURE 4: Effect of processing optical motion capture data.

FIGURE 5: Single-round table tennis trajectory with three difficult serves.
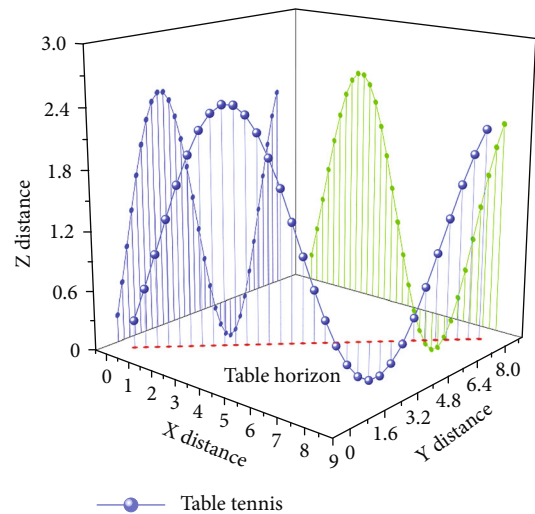

FIGURE 6: trajectory of the missed motion of a table tennis.

with the trainee failing to catch the ball and the corresponding X-axis coordinate dropping from its initial value (on the server side) to zero (on the trainee's side), while the Y-axis also maintains a stable approximate parabolic curve after the first touch; the curves of the three axes still show a slight change after the first touch and bounce, reflecting the trainee's interference with the trajectory when the racket touches the ball but fails to hit it back and the catch and return, interference with the trajectory when the racket touched the ball but failed to hit it back, and the case where

the catch-return ball did not touch the opponent's table, as the trajectory Y-axis did not bounce after the ball was hit, and the case representing the catch-return ball touching the net. Figure 6 shows the commonly missed motion trajectories of table tennis balls.

The image tracking effect when the table tennis is doing uniform acceleration is similar to the image tracking effect when it is moving at a uniform speed, the Kalman-based predictive controller successfully predicts the position information of the target when it is occluded, the image tracking
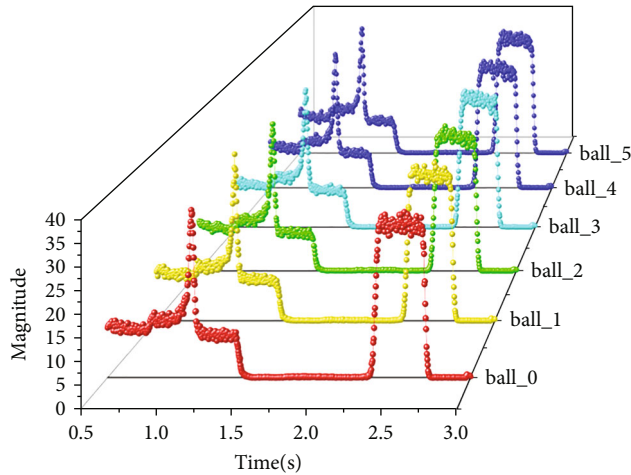
FIGURE 7: The plot of tracking error of the system using predicted data for tracking during uniformly accelerated motion of a table tennis.

algorithm directly resumes normal tracking when the target exits the occluder, and the target is still located in the center of the image, indicating that the controller closed loop is effective. The tracking error plot of the system is shown in Figure 7. In the case of occlusion, the tracking error of the X-axis reaches 20 dB, while the tracking error of the Y-axis is around 10 dB, and the data in the plot indicates that the X-axis tracking error is larger than the Y-axis tracking error. According to the analysis, the target does uniform acceleration on the X-axis and uniform velocity on the Y-axis. The reason is that the prediction effect of Kalman for uniform acceleration is worse than that for uniform velocity, which leads to a larger tracking error in uniform acceleration than that in uniform velocity in the occlusion, but the tracking error is still within the acceptable range trajectory during accelerated motion.

5. Conclusions

In response to the current situation of an insufficient number of coaches and subjective evaluation of referees in China's sports, this paper forms a set of simulation models of table tennis trajectories based on digital video image processing methods, especially motion evaluation based on motion capture technology, concerning the ideas of interactive training and personalized feedback in traditional sports. The main findings of the paper are as follows.

(1) Application of digital video image processing methods to table tennis. Using table tennis as an example, this paper describes the implementation difficulties and error handling of a generic sports-assisted training framework when applied to a specific sport. The specific forms of practice content, immediate performance, and skill metrics in table tennis are described, and the use of a digital video image processing method to capture the system to capture the trajectory of the table tennis ball and racket as well to restore the body position is clarified

(2) To address the problem of inaccurate contact judgment in inertial motion capture, a set of plantar contact platforms based on pressure measurement is built. A set of pressure acquisition circuits is designed according to the characteristics of pressure sensors and the basic features of human motion, and a stable and efficient plantar pressure acquisition is realized by the carrier of homemade insoles. Data transmission level to the collected video image information, through the data transmission equipment in the application environment, to build high-speed wireless data transmission network, the collected video image compression coding at a certain frame rate through the wireless LAN to the background data processing equipment to ensure real-time transmission of the front-end acquisition data. For the problem that there are differences in contact judgment thresholds for different weight groups and different sports movements, this paper proposes a set of adaptive methods for generating contact judgment thresholds, and the resulting contact judgment results are consistent with those observed in high frame rate videos

(3) To address the problem of errors in skeletal position measurement in inertial systems, this paper proposes a human posture correction method based on an optimization framework. The contact position of both feet is used as a constraint, and the weighted value of the measured pose in the current frame and the optimized pose in the previous frame is used as the initial value of optimization, which can make the corrected human pose have good smoothness under the condition of satisfying the contact position constraint and further improve the human inertial motion capture accuracy

Data Availability

All data, models, and code generated or used during the study appear in the submitted.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. G. Melo, M. F. Pinto, A. L. M. Marcato, I. Z. Biundini, and N. M. S. Rocha, "Low-cost trajectory-based ball detection for impact indication and recording," *Journal of Control, Automation and Electrical Systems*, vol. 32, no. 2, pp. 367–377, 2021.

[2] C. Lu, "Kalman tracking algorithm of ping-pong robot based on fuzzy real-time image," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 4, pp. 3585–3594, 2020.

[3] J. Wang, K. Zhao, D. Deng et al., "Tac-simur: tactic-based simulative visual analytics of table tennis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 407–417, 2020.

[4] L. Chaoming, "Prediction and analysis of sphere motion trajectory based on deep learning algorithm optimization," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 5, pp. 6275–6285, 2019.

[5] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi, "Ball tracking in sports: a survey," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1655–1705, 2019.

[6] O. Koç, G. Maeda, and J. Peters, "Optimizing the execution of dynamic robot movements with learning control," *IEEE Transactions on Robotics*, vol. 35, no. 4, pp. 909–924, 2019.

[7] R. Prakash, L. Behera, S. Mohan, and S. Jagannathan, "Dynamic trajectory generation and a robust controller to intercept a moving ball in a game setting," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 4, pp. 1418–1432, 2020.

[8] C. Cuevas, D. Quilon, and N. García, "Techniques and applications for soccer video analysis: a survey," *Multimedia Tools and Applications*, vol. 79, no. 39-40, pp. 29685–29721, 2020.

[9] M. Lahby, A. Essouiri, and A. Sekkaki, "A novel modeling approach for vertical handover based on dynamic k-partite graph in heterogeneous networks," *Digital Communications and Networks*, vol. 5, no. 4, pp. 297–307, 2019.

[10] M. Knauer and C. Büskens, "The augmented reality sandbox makes optimization visible," *IFAC-Papers OnLine*, vol. 53, no. 2, pp. 17572–17577, 2020.

[11] C. Wateosot and N. Suvonvorn, "Group activity recognition with an interaction force based on low-level features," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 14, no. 7, pp. 1061–1073, 2019.

[12] A. E. Abulwafa, A. I. Saleh, H. A. Ali, and M. S. Saraya, "A fog based ball tracking (FB2T) system using intelligent ball bees," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 5735–5754, 2020.

[13] D. Parambanchary and V. Malleswara Rao, "WOA-NN: a decision algorithm for vertical handover in heterogeneous networks," *Wireless Networks*, vol. 26, no. 1, pp. 165–180, 2020.

[14] M. D. Foote, P. E. Dennison, A. K. Thorpe et al., "Fast and accurate retrieval of methane concentration from imaging spectrometer data using sparsity prior," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 6480–6492, 2020.

[15] Z. Zhuang, "Optimization of building model based on 5G virtual reality technology in computer vision software," *Mathematical Biosciences and Engineering*, vol. 18, no. 6, pp. 7936–7954, 2021.

[16] K. Szelag, P. Kurowski, P. Bolewicki, and R. Sitnik, "Real-time camera pose estimation based on volleyball court view," *Opto-Electronics Review*, vol. 27, no. 2, pp. 202–212, 2019.

[17] Y. Zhang, "Detection and tracking of human motion targets in video images based on camshift algorithms," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11887–11893, 2020.

[18] J. Montalban, G. M. Muntean, and P. Angueira, "A utility-based framework for performance and energy-aware convergence in 5G heterogeneous network environments," *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 589–599, 2020.

[19] G. Maeda, O. Koc, and J. Morimoto, "Phase portraits as movement primitives for fast humanoid robot control," *Neural Networks*, vol. 129, pp. 109–122, 2020.

[20] Y. Cao and H. Mao, "High-dimensional multi-objective optimization strategy based on directional search in decision space and sports training data simulation," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 159–173, 2022.

[21] Z. Zhao, M. Karimzadeh, L. Pacheco et al., "Mobility management with transferable reinforcement learning trajectory prediction," *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 2102–2116, 2020.