Artificial Intelligence-based Volleyball Target Detection and Behavior Recognition Method

Jieli Huang, Wenjun Zou*
Physical education institute, Nanchang Jiaotong University, Nanchang 330100, China

***Abstract*—Volleyball has limitations in relying on judges' subjective judgments alone to call penalties for infractions in the court. While video detail enhancement technology is extremely useful for target tracking and extraction in sports video, the current research on video detail enhancement technology does not pay much attention to the development of ball game violation tracking and recognition. Therefore, the study uses the fusion algorithm of wavelet exchange method and three-frame difference method and background subtraction method to detect and extract the motion targets, and uses the improved CamShift tracking algorithm and HMM to track and identify the tracking targets for the violation actions. Comprehensively, the study constructs a tracking recognition model for volleyball violation based on video enhancement technology to achieve accurate penalty in intense rivalry games. Through experimental analysis and comparison, the tracking F-measure value of the model constructed by the study is 0.89, which can achieve a good tracking effect, the recognition accuracy is 99.76%, and the average error is 0.003, which can effectively realize the tracking recognition of players' illegal actions during volleyball, and objectively make court penalties to guarantee the fairness and justice of the game.**

Keywords*—Volleyball; video detail enhancement; hmm; CamShift tracking; detection; recognition

I. INTRODUCTION

In competitive competitions, the subjective judgment of the referee may be influenced by different perspectives and observation results, leading to doubts about the accuracy of the judgment. Video detail enhancement technology has become an important research direction for achieving accurate punishment of illegal actions in volleyball matches. However, despite the rapid development of computer and network technology, significant progress has been made in related research fields; the application of video detail enhancement technology in sports video still faces some challenges [1]. Due to the uncertainty of lighting conditions at the competition site, limitations of collection equipment, and the presence of noise interference, the quality and clarity of volleyball game video images are often low. This poses a challenge to the effectiveness of video detail enhancement algorithms. In addition, as the competition progresses, moving targets may experience problems such as rapid movement, deformation, and occlusion, which also increases the difficulty of tracking and recognizing moving targets. In addition, due to the individual differences of different athletes and the complexity of game rules, further research is needed on the identification methods for illegal actions [2]. With the current computer and networking skills, video detail enhancement has matured. In the process of image and video acquisition, due to the lighting environment, acquisition equipment, noise interference and other factors, it will lead to the degradation of image and video visual effects, so detail enhancement algorithms need to be used to highlight some of the detailed information in the image and video for subsequent processing [3]. There are two main types of video detail enhancement algorithms, namely, detail enhancement based on multiple input images and detail enhancement based on single input images [4]. While video detail enhancement techniques are extremely useful in the tracking and extraction of targets in motion video, not much research attention has been paid to video detail enhancement techniques [5]. The study uses video detail enhancement techniques to process video images in volleyball and to achieve detection and tracking of motion targets, and to identify the offending actions accordingly. The images are first preprocessed using the wavelet variation algorithm RGB. After pre-processing, the motion targets are detected and extracted using a fusion algorithm with three-frame difference method and background subtraction method. After the motion targets are detected and extracted, the motion targets are tracked and identified using CamShift tracking algorithm and Hidden Kolff (HMM), which is found to be lost when the target tracking may be obscured. To address this, the study adjusts the position of the CamShift algorithm's finding glass by introducing a Kalman filter (Kalman) to predict the motion parameters of the target. As a result, the study constructs a volleyball motion violation tracking recognition model based on video detail enhancement, which realizes the intelligence of game penalty and effectively solves the penalty problem caused by viewing angle and other reasons during the game. The importance of research lies in improving the accuracy and fairness of competition judgments, optimizing the effectiveness of video image processing and object tracking, and providing useful references for penalty issues in other sports. The innovation lies in the refinement of the application of video detail enhancement technology. Through the combination of various algorithms and models, a volleyball movement violation tracking and recognition model based on video detail enhancement has been constructed. This study has important practical significance in improving the accuracy of competition judgments. This study is mainly divided into five sections. The second section summarizes the research on video tracking, motion recognition, and other technologies both domestically and internationally. The third section is to construct a proposed volleyball foul action tracking and recognition model, analyzing image data processing, image detection, and target tracking. The fourth section is to analyze the performance of the constructed model and verify the

superiority of the proposed model. The fifth section discusses the results and analyzes deeper conclusions. The final section summarizes the results and proposes the shortcomings of the research and future research directions.

II. Related Works

In the process of image and video acquisition, under the influence of lighting environment, acquisition equipment, noise interference and other factors, it will lead to the degradation of image and video visual effect, and need to use detail enhancement algorithm to highlight some information in the image and video for easy discrimination and processing. For this, many scholars conducted research to improve the image visual effect. Xue et al. [6] found that most video enhancement algorithms depend on optical flow to enroll frames in video sequences, but flow estimation is difficult, so they proposed a TOFlow to achieve enhancement of image data, and found excellent optimization by comparing three functional tasks, frame interpolation, video denoising/denoising, and video super-resolution, with conventional optical flow for standard benchmark tests. Guan et al. [7] proposed a MFQE method for the lower house video and designed a new MF-CNN to improve the quality of compressed video by addressing the problem that existing methods to improve the quality of compressed images/videos mainly focus on improving the quality of individual frames without considering the similarity between consecutive frames, which effectively improves the effectiveness and generalization in terms of the latest image quality of highly compliant videos. Wang et al. [8] found that the key challenge of video SR is to effectively exploit the correlated asphyxia among consecutive frames, and that available deep learning-based methods typically estimate the optical flow among LR frames to provide temporal dependence, proposing an end-to-end video SR network to super-resolve optical flow and image. Zheng et al. [9] constructed an unsaturated magnetic excitation-based MFL measurement device by converting MFL information to image representation through the Zaitong pseudo-color imaging protocol, and the maximum modulus method was used for the point location of wire breakage to extract color moments, statistical texture features, and spectral texture features from the image. Tang et al. [10] proposed a new ship detection model that can be called FLNet by combining image processing methods and deep learning target detection methods in order to solve the problem that there is a large amount of background information and noise information similar to the ship in the image due to the mechanism of imaging by SAR, which badly affects the ship detection model performance.

In intelligent video analysis systems, motion target tracking is widely used in intelligent surveillance, human-computer interaction, and autonomous driving. In order to be able to track and recognize with high accuracy despite the challenges such as environmental changes, occlusion deformation and scale changes of the tracking target, the research on motion tracking recognition has been increasing. Kim et al. [11] tracked the excavator by using a pre-trained detector to locate the excavator and correlate the detection results, tracked the excavator by the tracking learning clean toilet algorithm, and finally used a hybrid deep learning algorithm to model the visual features of the excavator and the operation cycle of the sequential patterns were modeled and trained to propose a vision-based framework for animal recognition. Jaouedi et al. [12] proposed a hybrid deep learning model-based approach for human action recognition using gated recurrent neural networks to classify sequence data and videos, and upper and lower feature data extraction using evaluation algorithm. Angeliniet al. [13] proposed a hybrid deep learning model-based approach for HAR due to the gap between the deep learning data requirements and the functionality provided by the framework that needs to provide the application in terms of data recording devices, a new 2D pose based pose level HAR approach (ActionXPose) was proposed by reducing the gap using the human pose provided by OpenPos. Zhang et al. [14] proposed a new pose level HAR approach (ActionXPose) in order to deal with the gaps with different temporal context information for long duration time series features and enhance spatio-temporal attention, the human action recognition problem was solved by using Conv-LSTM and FC-LSTM, and a STDAN was designed. Ge et al. [15] effectively represent the spatial static and temporal dynamic information of videos, using GoogleNet to extract the features in the video frames, processed by a spatial transformer network then modeled the sequential information of the features by convolutional LSTM, and proposed a new attention mechanism based convolutional LSTM action recognition algorithm.

In summary, the CNN model is used as the mainstream direction in speech recognition technology. Although this model can improve the performance of speech recognition, the subsequent structure construction is complicated and is not conducive to improving the operation efficiency of the model. Therefore, the research starts from neural network and proposes OPGRU to simplify the structure of speech recognition model and improve the recognition accuracy and operation efficiency of the model.

III. Volleyball Foul Action Tracking Recognition Model Construction

A. Image Data Pre-Processing and Target Detection

Common penalty errors and misjudgments have been the trigger for conflicts on the court. The requirement of action specification in volleyball is very high, in order to ensure the reduction of errors and wrong calls in volleyball and the accurate detection of fouls committed by players. The ball game video is processed and analyzed using video detail enhancement technology [16]. Before tracking and recognition of ball game violation actions, pre-processing of sports video image data is required to improve the accuracy and recognition precision of tracking afterwards. Firstly, the image is converted to grayscale image by RGB and the image is weighted, and the processing formula is shown in the following Eq. (1).

$$I_i(x,y) = \alpha r(x,y) + \beta g(x,y) + \gamma b(x,y) \quad (1)$$

The values α, β, and γ in Eq. (1) are the values for weighting the action images of volleyball players. When the

athlete action image is an action image captured in natural light is, the image weight is set to 1. When the action image is captured in a single light, the image weight is set to 0 to eliminate the shadow of the volleyball player's body, and the wavelet transform algorithm is used to reduce the noise of the data signal after the image is grayscale processed and weighted, and the wavelet transform formula is shown in Eq. (2) below.

$$WT(a,\tau)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\cdot\varphi\left(\frac{t-\tau}{a}\right)dt \tag{2}$$

In Eq. (2), a denotes the variable scale and τ denotes the variable translation. After multiple wavelet transforms, the obtained signal components are shown in Eq. (3) below.

$$S=A_n+D_1+D_2+\ldots+D_n \tag{3}$$

In Eq. (3), S denotes the original signal, D_n denotes the noisy signal obtained after n wavelet transforms, and A_n denotes the effective signal obtained after n wavelet transforms. After verification, it is found that the 3-layer wavelet transform has the best denoising effect. According to the above, the denoising and grayscale processing of volleyball sports video action data is realized. After processing the image need to detect the motion target, motion target detection is to take the target color, shape, position and size information in each frame of the video stream, and the video sequence is essentially three-dimensional data containing a time dimension, the study uses the three-frame difference method to extract the motion target, the three-frame difference method is illustrated in Fig. 1.

In Eq. (8), α is the number, which is fed into the section as needed, w is the degree, and h is the figure. Eq. (9) is obtained by "summing" the background subtraction method with the motion target information obtained from the three-frame difference method.

$$D(x,y)=DI(x,y)\otimes DB(x,y) \tag{9}$$

After obtaining the target information, the background adaptive update is performed, and the update expression is shown in Eq. (10).

$$B(x,y)=\begin{cases}B(x,y) & D(x,y)=0\\ uB(x,y)+(1-u)I_k(x,y) & D(x,y)\neq 0\end{cases} \tag{10}$$

The best value of u in Eq. (10) is 0.997, which is obtained after the study. The flow chart of the fusion algorithm of the three-frame difference method and the background subtraction method is shown in Fig. 2.

Firstly, the three adjacent frame degree values I_{k-2}, I_{k-1} and I_k are collected for the operation of the neighboring two difference absolute, and the difference map formula is obtained as shown in Eq. (4).

$$\begin{cases}D_1(x,y)=|I_{k-1}-I_{k-2}|\\ D_2(x,y)=|I_k-I_{k-1}|\end{cases} \tag{4}$$

Binarizing the two neighbor-dual differences, the expression is shown in Eq. (5).

$$\begin{cases}T_1=d_1+\beta\delta_1\\ T_2=d_2+\beta\delta_2\\ D_1(x,y)=\begin{cases}255 & d_1\geq T_1\\ 0 & d_1<T_1\end{cases}\\ D_2(x,y)=\begin{cases}255 & d_2\geq T_2\\ 0 & d_2<T_2\end{cases}\end{cases} \tag{5}$$

In Eq. (5), d is the mean value of the difference map, δ is the standard deviation of the difference map, and T represents the threshold value. The two binarized images are subjected to or operation to obtain the motion target information, as shown in Eq. (6) below.

$$DI(x,y)=D_1(x,y)\oplus D_2(x,y) \tag{6}$$

The above equation is used to process the image, but in the process of research, it is found that there are still some limitations, and the motion target in the extraction will produce a hole phenomenon, for this problem, the study combines the background subtraction method and the three-frame difference method. The Eq. (7) is obtained from the background subtraction method.

$$DB(x,y)=|I_k(x,y)-B(x,y)| \tag{7}$$

In Eq. (7) $I_k(x,y)$ is the current degree value, $B(x,y)$ is the back pixel gray value. $d(x,y)$ is the difference absolute value image $DB(x,y)$ value image $DB(x,y)$ pixel point of $\bar{d}$ and standard δ, with and standard deviation set to a value of T, into the binarization, to obtain the expression (8).

$$\begin{cases}\bar{d}=\dfrac{\sum_{x=0}^{x<w}\sum_{y=0}^{y<h}d(x,y)}{wh}\\ \delta=\sqrt{\dfrac{\sum_{x=0}^{x<w}\sum_{y=0}^{y<h}\left[d(x,y)-\bar{d}\right]^2}{wh}}\\ T=\bar{d}+\alpha\delta\\ DB(x,y)=\begin{cases}225 & d\geq T\\ 0 & d<T\end{cases}\end{cases} \tag{8}$$

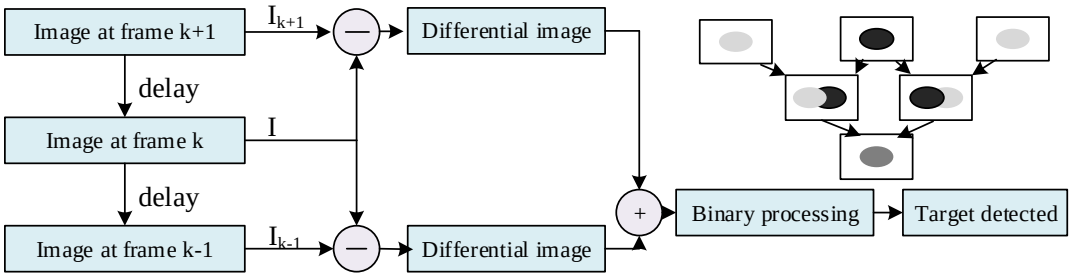

Fig. 1. Diagram of three frame difference method.

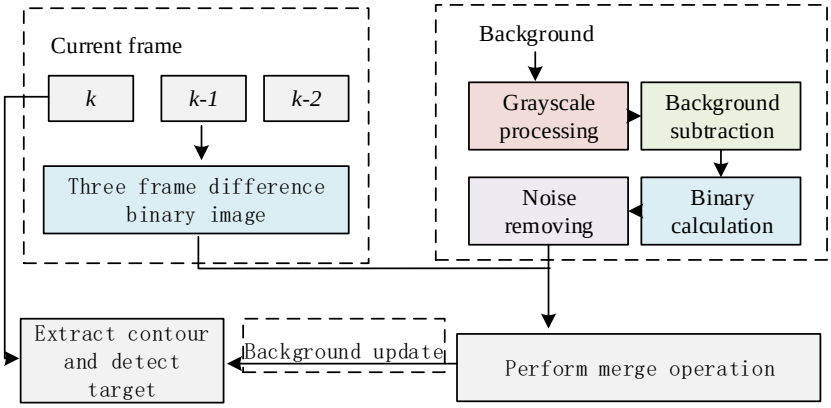

Fig. 2. Flow chart of fusion algorithm for target detection.

Combining the above, the study implements the pre-processing of volleyball sports action video images, and detects and extracts the image targets using an algorithm fused with the three-frame difference method and the background elimination method.

B. Target Tracking based on Improved CamShift Algorithm

After the target is detected and extracted, it needs to be tracked so that the subsequent violation movements during volleyball sports can be identified in real time. The CamShift is a successive self-adaptive Meanshift algorithm. The Meanshift algorithm that finds and iterates over the pixels of a single image for optimal results, the CamShift mainly iterates processing of video sequences, where each frame in the video is called using the Meanshift algorithm. The Meanshift algorithm belongs to the kernel density estimation method, which describes the model of the target and the candidate model by the probability of the pixel feature values in the region and in the candidate region. Given a sample of points in the d-dimensional space Rd, the Meanshift vector of points has the basic form shown in Eq. (11).

$$M_h(x)=\frac{1}{k}\sum_{x_i\in S_h}(x_i-x) \tag{11}$$

In Eq. (11), x_i is the sample points in the interval, k denotes the sample falling into the S_h region, and S_h is the high-dimensional spherical region of radius h, which is the set of y points satisfying the relationship in Eq. (12) below.

$$S_h(x)=\left\{y:(y-x)^T(y-x)\le h^2\right\} \tag{12}$$

The study extended the basic Meanshift form by introducing a kernel function in order to take into account the effect of the distance of each pixel point during the calculation, as shown in Eq. (13) below.

$$M_h(x)=\frac{\sum_{i=1}^{n}G\left(\frac{x_i-x}{h}\right)\omega(x_i)(x_i-x)}{\sum_{i=1}^{n}G\left(\frac{x_i-x}{h}\right)\omega(x_i)} \tag{13}$$

In Eq. (13) $G(x)$ is a unit kernel function and ω is the weight assigned to the sampled points. Using Eq. (13) for iteration, the following Eq. (14) is obtained.

$$m_h(x)=M_h(x)+x \tag{14}$$

After calculating the value of $m_h(x)$, assign it to x, and then calculate $M_h(x)$ again. If the absolute value of $M_h(x)$ is less than the tolerance error, end the cycle to get the final target position. If not, continue the calculation. The meanshift tracking algorithm is shown in Fig. 3.

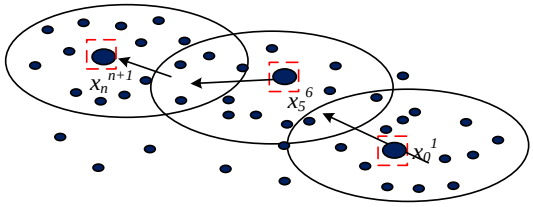

Fig. 3. Schematic diagram of MeanShift tracking algorithm.

The CamShift utilizes the modalities of the probability distribution image detected by the Meanshift algorithm in continuous video detection by introducing a feedback loop in which the previous detection result is used as input to the next detection process and restricting the search area to the surroundings of the latest known target location. After the color tracking probability model is established, the video image is transformed with a color probability distribution map. The model initializes a search window (SW) in the first frame of the image, and adjusts the window size and position according to the tracking target in the way shown in Eq. (15).

$$\begin{cases} \hat{\bar{p}}_k(W) = \dfrac{1}{|W|}\sum_{j\in W} p_j \\ \hat{\bar{p}}_k(W) - p_k \approx \dfrac{f'(p_k)}{f(p_k)} \end{cases} \tag{15}$$

In Eq. (15) W is the SW of size s in the color probability distribution map, p_k is the initialized centroid of the SW, $f(p)$ is the Meanshift climbing gradient equation, and the new centroid $\hat{\bar{p}}_k$ is found by dynamic iteration, and this is cycled until convergence to achieve adaptive change of the window. During the study, it is found that the target tracking may be obscured during the target tracking resulting in target loss. To address this, the study adjusts the position of the SW of the CamShift by introducing the Kalman filter (Kalman) to predict the motion parameters of the target to compensate for the temporary target loss due to occlusion. In two stages, prediction and correction, the components of the predicted state vector are used to set the center position of the SW of the CamShift, and the center of gravity position output by the CamShift is used as the measurement value to correct the predicted state vector and achieve the optimization improvement of the CamShift.

C. HMM-based Foul Play Analysis Identification

Judgment of foul actions and near-foul actions in volleyball by the eyes of the referee alone usually leads to errors. The study has already used a fusion algorithm combining the three-frame difference method and the background elimination method to detect and extract the motion targets in the previous paper, and the improved CamShift has been used to achieve the tracking of motion targets in ball game sports videos. The HMM recognition process is shown in Fig. 4.

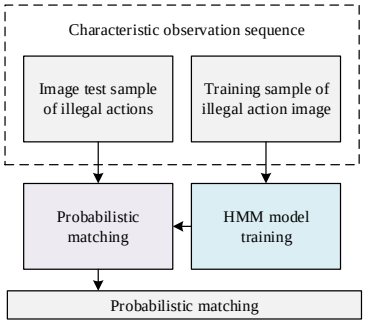

Fig. 4. HMM violation action identification process.

The data set of foul actions of volleyball players is divided into different types, firstly, each violation is modeled in a targeted way, and the study sets the amount of data for each violation to 120 and decomposes the action into n meta-actions. These meta-actions during the motion are temporal in nature, so each game violation is considered as a sequence of observations of length n. Training and learning are performed for this sequence of observations to find the best HMM parameters for each action model [17]. After finding the best parameters, the extracted observation sequence data is used as the input data of the HMM, and the probability of the best state sequence occurrence of the action in the current video under each action model is obtained using the Viterbi algorithm. The action which corresponds to the model with the largest probability of output is the identification outcome of the current observed sequence. HMM is expressed as the following Eq. (16).

$$\lambda = (A, B, \pi) \tag{16}$$

In Eq. (16), A is the state probability distribution, B is the observation probability distribution, and π is the initial probability distribution. To find the appropriate HMM parameters, the study uses the Baum-Welch algorithm to train each parameter of the HMM so that the probability of the observed sequence in the model is maximized. The state sequence data is considered as unobservable hidden data I as shown in Eq. (17) below.

$$P(O|\lambda) = \sum_I P(O|I,\lambda)P(I|\lambda) \tag{17}$$

In Eq. (17), O is the observed sequence data. The maximum expectation algorithm (EM) is used to implement the HMM algorithm for parameter learning. Q function, as follows in Eq. (18).

$$Q(\lambda, \bar{\lambda}) = \sum_I \log P(O, I|\lambda) P(O, I|\bar{\lambda}) \tag{18}$$

In Eq. (18), $\bar{\lambda}$ denotes the current estimate of the model parameters and λ is the maximized model parameters. After obtaining the value of the Q function, the parameters of the HMM were obtained by maximizing the Q function, combined with the Lagrange multiplier method. In volleyball violation recognition, after the training of the violation model is completed, the study uses the Viterbi algorithm to find the optimal solution of the HMM. For a given HMM model and observed sequence data, the optimal path $I^* = (i_1^*, i_2^*, \ldots, i_T^*)$ is solved, and T denotes the length of sequence I. Through the above operation, the analysis of the recognition of ball game violation actions is completed. The principle of the Viterbi algorithm is shown in Fig. 5.

Comprehensive above, the research uses video detail enhancement technology to detect and track the target of the game video image, and finally uses the action model for recognition, and constructs a volleyball foul action tracking recognition model based on video detail enhancement technology, which effectively makes accurate judgment on each violation action remembered in volleyball.

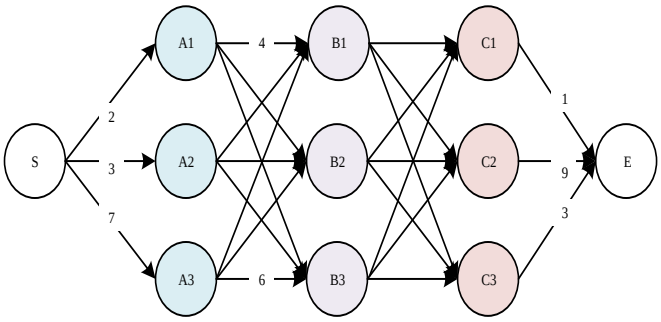

Fig. 5. Principle of viterbi algorithm.

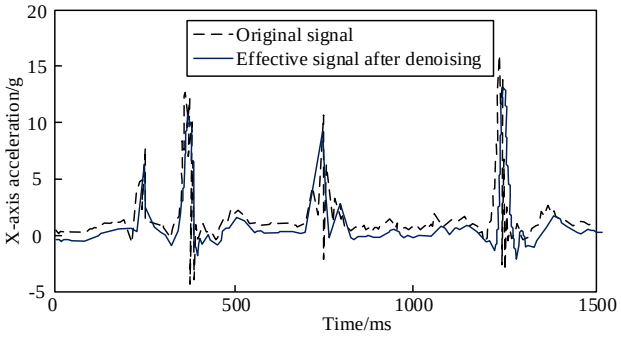

Fig. 6. Effect before and after wavelet change noise reduction.

IV. Tracking Recognition Model Performance Analysis

The study conducted performance analysis on the constructed model, using the Volleyball Dataset for training and analysis. The dataset consists of 55 videos, with 4830 keyframes annotated with athlete positions, as well as their individual and group actions. The study divided 4830 keyframes into training and testing sets, with a ratio of 8:2 between the training and testing sets. The study was conducted to improve the accuracy and recognition precision of tracking. The wavelet transform noise reduction process was performed on the image, and to verify the noise reduction process effect, the study used the same set of violating actions. The noise reduction effect of the image of the x-axis acceleration of the violation action is compared as shown in Fig. 6.

As can be seen from Fig. 6, before the use of wavelet denoising, there is a lot of redundant data and noise in the image of the offending action, which is difficult to extract and identify subsequently, after the wavelet transform noise reduction eliminates the redundant data removal in the x-axis acceleration of the foul action, and maintains the original curve direction while noise reduction, which makes the waveform graph clearer and improves the accuracy of tracking and identification of subsequent video images. To verify the tracking effect of the tracking algorithm, the research improved CamShift tracking algorithm is compared longitudinally with the currently popular and superior performance tracking algorithms: CT, TLD, IVT, and L1PAG to verify the tracking of a subset of target deformation, a subset of illumination changes, and a subset of background interference, a subset of the three interference cases using the algorithm. The distance accuracy curves (precision plot) for comparing different cases are shown in Fig. 7.

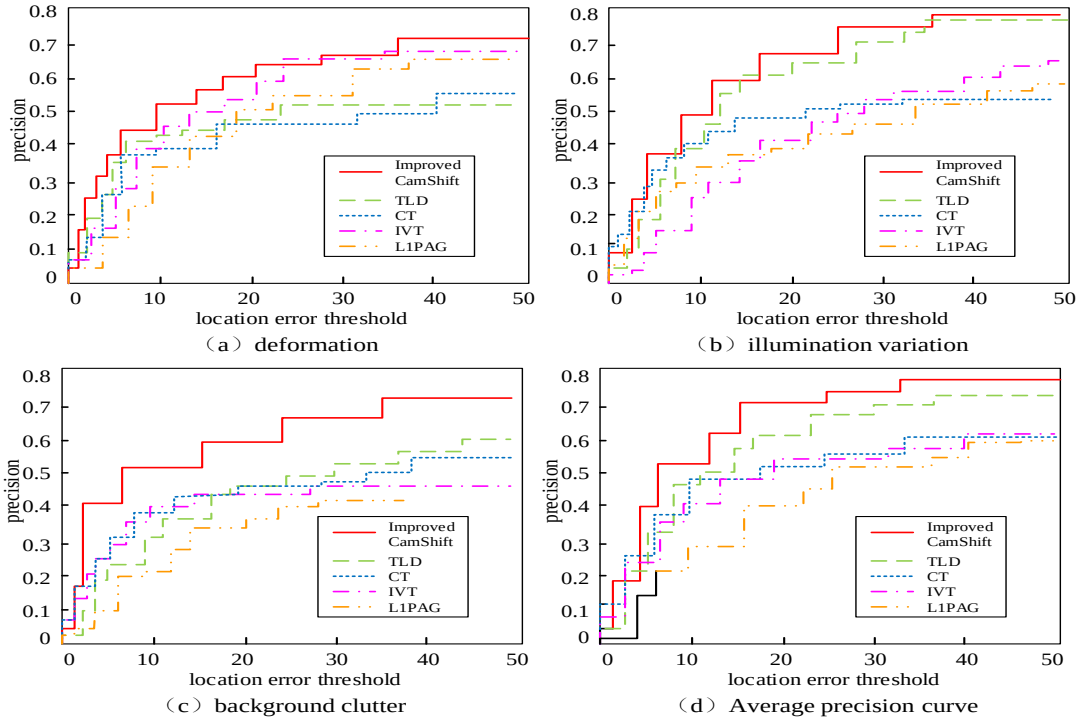

Fig. 7. Comparison of Accuracy Curves.

In Fig. 7, the tracking distance precision of each algorithm is affected by different effects under different interference scenarios, and increases with the increase of the positioning error threshold, among which the improved CamShift tracking algorithm is the least affected by interference, and the average tracking precision is 0.79 under each interference scenario at a positioning error threshold of 50. The TLD tracking algorithm is at a positioning error threshold of 50. The average tracking precision of the TLD tracking algorithm is 0.73 at a positioning error threshold of 50, which is 8.2% less than the precision of the improved structure. The average tracking precision of the CT tracking algorithm is 0.59 at a positioning error threshold of 50, which is 33.9% less than the precision of the improved structure. The average tracking precision of the IVT tracking algorithm is 0.61 at a positioning error threshold of 50, which is 29.5% less than the precision of the improved structure. The average tracking precision of the L1PAG tracking algorithm is 0.59 at a positioning error threshold of 50, which is 29.5% less than the precision of the improved structure; and the average tracking precision of the L1PAG tracking algorithm is 0.61 at a positioning error threshold of 50. The average tracking precision of the L1PAG tracking algorithm is 5.7 when the localization error threshold is 50, which is 38.6% less than the algorithm precision. The above figure shows that the proposed algorithm has better tracking effect in complex scenarios such as target deformation, light change, and background disturbance. To further verify the performance of the algorithms, the study introduces recall (Re), precision (Pr), and comprehensive performance (F-measure) to compare the performance of the five tracking algorithms under three different scenarios of multimodal background, light change, and bad weather with quantitative metrics, as shown in Table I.

TABLE I. COMPARISON OF AVERAGE PERFORMANCE OF TRACKING ALGORITHMS

Algorithm	Scene	Re	Pr	F-measure
Improved CamShift	Highway	0.90	0.88	0.89
	Fountain	0.88	0.87	0.87
	Wet Snow	0.92	0.91	0.91
TLD	Highway	0.82	0.81	0.81
	Fountain	0.84	0.80	0.82
	Wet Snow	0.76	0.79	0.78
CT	Highway	0.82	0.83	0.82
	Fountain	0.79	0.80	0.79
	Wet Snow	0.80	0.81	0.80
IVT	Highway	0.73	0.75	0.74
	Fountain	0.69	0.70	0.69
	Wet Snow	0.77	0.79	0.78
L1PAG	Highway	0.74	0.72	0.73
	Fountain	0.71	0.74	0.72
	Wet Snow	0.68	0.69	0.68

Table I shows that the improved CamShift tracking algorithm using particle filtering significantly improves the ability of the improved algorithm to handle complex backgrounds including light changes and multimodal backgrounds, and the performance index value of this algorithm is the highest in all scenes. The improved CamShift tracking algorithm has an average Re value of 0.90, an average Pr value of 0.89, and an F-measure value of 0.89 for the three scenes. Slightly higher than the TLD tracking algorithm, which has an average Re value of 0.81, an average Pr value of 0.80, and an F-measure value of 0.80. The CT tracking algorithm has an average Re value of 0.81 The average Re value of the CT tracking algorithm is 0.81, the average Pr value is 0.81, and the F-measure value is 0.81; the other two algorithms have lower values of quantitative indicators. The comprehensive table above shows that the research improved CamShift tracking algorithm has high comprehensive performance and good robustness. To verify the recognition effect of the volleyball violation recognition model (model 1) constructed by the study, the recognition models constructed by convolutional neural network (CNN) (model 2) and support vector machine (SVM) (model 3) and BP neural network (model 4) were used for violation recognition using volleyball sports videos from video websites, and the recognition effect was compared for training and testing, and the results are shown in Fig. 8 is shown.

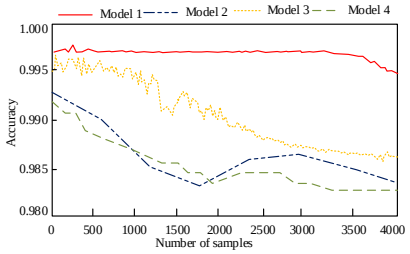

Fig. 8. Comparison of recognition effects of four recognition models.

In Fig. 8, the recognition rate of the offending actions in models 1, 2, 3, and 4 decreases as the sample size increases. When the sample reaches at 500, the precision of model 1 was 99.76%, the recognition precision of model 3 was 98.68% lower than that of model 1 by 0.08%, the recognition precision of model 2 was 99.21%, lower than that of model 1 by 0.55%, and the recognition precision of model 4 was 98.89%, lower than that of model 1 by 0.87%. When the sample was increased to 4000, the precision of the four model precision all decreased, but model 1 decreased the least, the recognition precision was 99.52%, model 3 recognition precision was 98.39%, 1.13% lower than model 1, model 2 recognition precision was 98.49%, 1.03% lower than model 1, model 4 recognition precision was 98.32%, 1.20% lower than model 1. The comprehensive content of Fig. 6 shows that model 1 has high stability and the highest recognition precision among the four recognition models. To further verify the recognition model performance, the recognition errors of the four models are recorded and compared, as shown in Fig. 9.

From Fig. 9, the highest error of recognition error curve of model 1 is 0.009, the lowest error is 0.001, and the average error is 0.003. It can be seen that the overall curve of this model is lower than the other three models, among which the highest recognition error of model 3 is 0.012, the highest recognition error of model 4 is 0.018, and the highest recognition error of model 2 is 0.016. The volleyball violation tracking recognition model constructed by the study can effectively track and identify the violations during the game, providing a strong guarantee for the fairness of the game.

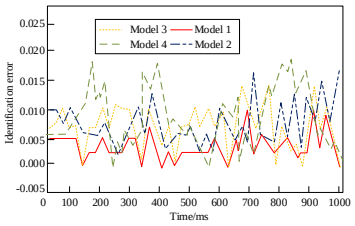

Fig. 9. Identification error of four models.

V. DISCUSSION

Through the above experimental results, it was found that wavelet transform denoising of images can effectively eliminate redundant data and noise, maintain the original curve direction, and make the waveform clearer. This is crucial for the tracking and recognition of subsequent video images. Compared with popular tracking algorithms such as CT, TLD, IVT, and L1PAG, the improved CamShift tracking algorithm has higher tracking accuracy and robustness in subsets of illumination changes, background interference, and target deformation. Similarly, accuracy, recall, and comprehensive performance indicators also demonstrate its superiority. Compared with popular tracking algorithms such as CT, TLD, IVT, and L1PAG, the improved CamShift tracking algorithm has higher tracking accuracy and robustness in subsets of illumination changes, background interference, and target deformation. Similarly, accuracy, recall, and comprehensive performance indicators also demonstrate its superiority. The recognition error of Model 1 is the smallest, indicating that the model has high robustness and accuracy, and is suitable for use in actual competitions to accurately track and identify violations that occur during the competition. In summary, the improved CamShift tracking algorithm and volleyball violation recognition model constructed in the study have been compared and tested, showing high accuracy and stability. This will have a significant impact on the fairness of actual matches and open up a path for subsequent research, indicating that important positions should be given to data preprocessing and model optimization in such research.

VI. CONCLUSION

In volleyball, referees are very prone to subjective judgment errors due to different observations from different angles during the game viewing process. For this reason, the study uses RGB to grayscale the image and wavelet variation algorithm to noise reduce the video image, and then uses the fusion algorithm of three-frame difference method and background subtraction method to detect and extract the motion target in the image. After the motion target is detected and extracted, the CamShift tracking algorithm is used to track the motion target and it is found that the target may be lost due to the occlusion of the tracking target during the target tracking. To address this, the study adjusts the position of the SW of the CamShift by introducing the motion parameters of the Kalman prediction target to achieve improved optimization. The set of images that have been extracted from the offending actions are input to the HMM recognition model to track and identify the offending actions that appear in the motion video. As a result, the study constructs a volleyball motion violation tracking recognition model based on video detail enhancement. Through the analysis of experimental verification, the recognition accuracy of the research constructed tracking recognition model is 99.76%, and the average error is 0.003, which can effectively realize the tracking recognition of players' illegal actions during volleyball sports and realize the fairness of the penalty in the court game. At present, the model is only used in sports competitions and has not been put into more fields, which can be further explored in the future research.

REFERENCE

[1] Hessert B. The protection of minor athletes in sports investigation proceedings. The International Sports Law Journal, 2021, 21(1):62-73.

[2] Bao W, Lai W S, Zhang X, Gao Z, Yang M H. Memc-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(3):933-948.

[3] Liu R, Fan X, Zhu M, Hou M, Luo Z. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(12):4861-4875.

[4] Wang W, Chen Z, Yuan X, Wu X. Adaptive image enhancement method for correcting low-illumination images. Information Sciences, 2019, 496:25-41.

[5] Hou R, Zhou D, Nie R, Liu D, Xiong L, Guo Y, Yu C. VIF-Net: an unsupervised framework for infrared and visible image fusion. IEEE Transactions on Computational Imaging, 2020, 6: 640-651.

[6] Xue T, Chen B, Wu J, Wei D, Freeman W T. Video enhancement with task-oriented flow. International Journal of Computer Vision, 2019, 127(8): 1106-1125.

[7] Guan Z, Xing Q, Xu M, Yang R Liu T, Wang Z. MFQE 2.0: A new approach for multi-frame quality enhancement on compressed video. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(3): 949-963.

[8] Wang L, Guo Y, Liu L, et al. Deep video super-resolution using HR optical flow estimation. IEEE Transactions on Image Processing, 2020, 29: 4323-4336.

[9] Zheng P, Zhang J. Quantitative nondestructive testing of wire rope based on pseudo-color image enhancement technology. Nondestructive Testing and Evaluation, 2019, 34(3): 221-242.

[10] Tang G, Zhao H, Claramunt C, et al. FLNet: A Near-shore Ship Detection Method Based on Image Enhancement Technology. Remote Sensing, 2022, 14(19): 4857-4857.

[11] Kim J, Chi S. Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles. Automation in Construction, 2019, 104: 255-264.

[12] Jaouedi N, Boujnah N, Bouhlel M S. A new hybrid deep learning model for human action recognition. Journal of King Saud University-Computer and Information Sciences, 2020, 32(4): 447-453.

[13] Angelini F, Fu Z, Long Y, Shao L Naqvi S M. 2D pose-based real-time human action recognition with occlusion-handling. IEEE Transactions on Multimedia, 2019, 22(6): 1433-1446.

[14] Zhang Z, Lv Z, Gan C, Zhu Q. Human action recognition using convolutional LSTM and fully-connected LSTM with different attentions. Neurocomputing, 2020, 410: 304-316.

[15] Ge H, Yan Z, Yu W, et al. An attention mechanism based convolutional LSTM network for video action recognition. Multimedia Tools and Applications, 2019, 78(14): 20533-20556.

[16] Liu N, Liu P. Goaling recognition based on intelligent analysis of real-time basketball image of Internet of Things. Journal of supercomputing, 2022, 78(1):123-143.

[17] Wang Y, Song Q, Ma T, Yao N, Liu R, Wang B. Research on human gait phase recognition algorithm based on multi-source information fusion. Electronics, 2022, 12(1): 193-193.