Complex & Intelligent Systems (2022) 8:4993–5010
https://doi.org/10.1007/s40747-022-00752-3

ORIGINAL ARTICLE

Automatic data volley: game data acquisition with temporal-spatial filters

Xina Cheng[1] · Linzi Liang[2] · Takeshi Ikenaga[2]

Received: 7 December 2021 / Accepted: 8 April 2022 / Published online: 29 April 2022

Abstract
Data Volley is one of the most widely used sports analysis software for professional volleyball statistics analysis. To develop the automatic data volley system, the vision-based game data acquisition is a key technology, which includes the 3D multiple objects tracking, event detection and quality evaluation. This paper combines temporal and spatial features of the game information to achieve the game data acquisition. First, the time-vary fission filter is proposed to generate the prior state distribution for tracker initialization. By using the temporal continuity of image features, the variance of team state distribution can be approximated so that the initial state of each player can be filtered out. Second, the team formation mapping with sequential motion feature is proposed to deal with the detection of event type, which represents the players' distribution from the spatial concept and the temporal relationship. At last, to estimate the quality, the relative spatial filters are proposed by extracting and describing additional features of the subsequent condition in different situations. Experiments are conducted on game videos from the Semifinal and Final Game of 2014 Japan Inter High School Games of Mens Volleyball in Tokyo Metropolitan Gymnasium. The results show 94.1% rounds are successfully initialized, the event type detection result achieves the average accuracy of 98.72%, and the success rate of the events' quality evaluation achieves 97.27% on average.

Keywords Sports video analysis · Tracker initialization · Event detection · Quality evaluation

Introduction

With the high-speed development of the vision sensing devices, the vision-based sports analysis technologies contribute in more and more fields, such as TV broadcasting contents, strategy development and coaching system. In game broadcasting, precise and real-time game data will provide much more attractive contents to the audiences to understand the game status. In the strategy development and coaching system, abundant game data help the teams and players to know their strengths and weaknesses, so that the proper strategy and personal training plans can be developed efficiently.

The performance of the sports analytic tools and the reliability of the game strategy development are depending on the quality and quantity of the game data. Therefore, the big data of the network are expected to support the sports analysis and other applications. [1,2] Some researches pay attentions to the applications of sports analysis based on big data [3,4], which will undergo major changes thanks to the utilization of the big data. However, the social network for utilizing big data is mature while how to obtain the data becomes key problem in real applications.

For the data acquisition in real game, only the 3D physical data (especially the position and speed information) make sense in data analysis and strategy development. However, most existing practical products for data acquisition only extract and present the game data (such as the position information) in the image coordinate system. From this point of view, volleyball is a typical object of 3D sports analysis, since both the ball and the players require the 3D concept to describe their motions. Once the volleyball game data can be acquired by computer vision method, other game data can also be obtained in similar way.

For the data acquisition and analysis of volleyball game, Data Volley [5] is the most widely used software for professional statistics analysis of volleyball games. This software

✉ Xina Cheng
xncheng@xidian.edu.cn

[1] School of Artificial Intelligence, Xidian University, No. 2 South Taibai Road, Xi'an 710071, China

[2] Graduate School of Information, Production and Systems, Waseda University, Kitakyushu City 808-0135, Japan

Fig. 1 The software interface of Data Volley

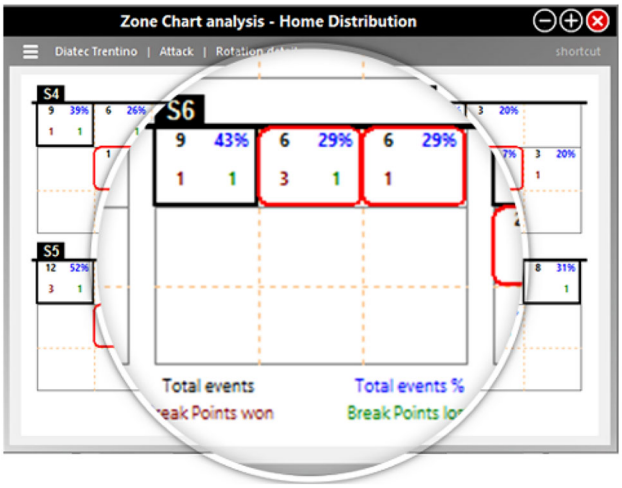

Fig. 2 The rule of count zone division for players' location analysis in Data Volley

can not only record the technical and tactical playing data of the players from both teams by a convenient interface, but also get a variety of statistical analysis data immediately in the statistical process, which helps the coaches to conduct real-time analysis and on-the-spot guidance for the game. In Data Volley, all input data are observed and judged by peoples through watching the game. This input process not only costs large human labor and time but also lacks of data accuracy since human eyes are weak at measuring the distance, velocity and time. Therefore, this article targets on the automatic and precise game data acquisition method to development the automatic Data Volley system with high reliability and efficiency of the volleyball game analysis.

Introduction of data volley

In order to achieve the development of the automatic Data Volley system, the data categories and functions utilized in Data Volley should be made clear. Thus, the requirement of expected automatic data acquisition method can be decided. The software interface of Data Volley (as Fig. 1 shows) is simple and clear, so that the game data can be quickly input through the keyboard and mouse operations. In Data Volley software, several types of data are available to be recorded and presented by some basic code. For each event, the input format sequence of the basic code is "*the player number + event code + the start of the ball's trajectory + the evaluation of the event*". The simple code sequence consists of information related to all the players and the ball.

- The location area of each player is required to decide the code of "*the player number*". As Fig. 2 shows, each half-court is divided into nine zones, which are used to present the position information of each player.
- The "*event code*" denotes the event information. In Data Volley, seven types are defined to represent the function of every event. Based on some introductions [6] and volleyball game rules, the descriptions of all the events are summarized in Table 1.
- The roughly trajectory of the ball with whose starting position is the required "*start of the ball's trajectory*".The number of the zone is used as the basic code.
- The "*quality of each event*" is the comprehensive judgment based on the related players and the ball is required. For each type of event, symbols are used to represent the quality level like it is shown in Table 2.

Although Data Volley has demonstrated its great ability in the convenient interface, effective cooperation, and the comprehensive analysis functions, it still has limitations in practical applications. First, it is unfriendly for beginners to learn the input key and basic code. Second, the data observed by human eyes is rough, because the human eyes are weak at measurement of position, velocity and time. Third, the data accuracy is affected by individual difference of different analyses' judgment and operations. At last, it is difficult to synchronize the game video with the acquired game data.

Automatic game data acquisition

Based on the statement above, the research target of this article is the vision-based game data acquisition for the automatic Data Volley system. With the game videos, the computer vision technology will provide more accurate data than manually input, especially for the position and velocity information, so that the reliability and accuracy of the game data analysis will increase. According to the data being used in the Data Volley software and considering other game data which are useful for the analysis of game strategy, the requirement of automatic data acquisition from game video are stated as following. Firstly, the available game playing period should be detected from the entire game video, which also includes the rest time, pause time and other parts besides

Table 1 Description of events in Data Volley

Type	Description
Serve	Serve the ball at the start of one round
Receive	First pass after the ball crosses the net
Set	Put the ball in the air to decide attacker
Attack	Attempt to score a point against defense
Block	Interception of the ball coming from opponent over the net
Dig	Prevent ball from hitting the ground after opponent's attack or on an emergency
Free Ball	Pass the ball over the net when attack is impossible

Table 2 Quality marks of each event

Event		Sever	Receive	Attack	Block	Dig	Set	Free Ball
	Bad	=	=	=	=	=	=	=
	↓	/	/	/	/	/	/	/
Quality	↓	−	−	−	−	−	−	−
code	↓			!				
	↓	+	+	+	+	+	+	+
	Good	♯	♯	♯	♯	♯	♯	♯

playing. Secondly, all the players should be distinguished from each other, and at each event, the locations of players are required. Thirdly, as the evaluation criteria required, the trajectory including the position and velocity of the ball is also important data. Fourthly, the game events are required to present the game status. At last, based on the information of the ball and players in one event, the evaluation of the event is required.

In this paper, we propose data acquisition methods to automatically collect the game data required by the Data Volley software from complete volleyball games. Our contributions are summarized as follows:

- A new initialization method of multiple player tracking is proposed for the team sports videos. The team state distribution and the sub-distribution are designed representing the entire team and each player based on game rules and physical basis. Instead of detecting all the players separately based on the image features, the sub-distribution of each player is filtered out through the proposed time-vary fission filter, which can be used to initialize the tracker directly. Combing the temporal and spatial features, this algorithm is robust for the occlusion and similar appearance of targets.
- An event detection method based on the team formation mapping and the sequential motion feature are proposed. The team formation mapping represents the players' distribution of each event, which reveals the moving tendency of the team, so that the intra-class events are easier to be distinguished. The sequential ball motion state feature is designed to describe the relationship between the current event and the former one.
- For the target of evaluating the quality, the event series feature and relative spatial filter are designed. To describe different situations, event series feature utilizes the relationship of event types and the game process. The relative spatial filter is proposed to extract the information of the following events so that the overall quality of each event can be evaluated.

Related works

In this section, the related works are discussed for different tasks. The target of this article is automatic Data Volley for the volleyball game analysis. With the similar target, the vision-based game analysis methods are discussed at first. Then the data acquisition for automatic Data Volley can be divided into several tasks, the detection of play scene, the multiple player tracking, the ball tracking, the event detection and event evaluation. For each task, we use one subsection to discuss the related work. Among these tasks, the detection of play scene and the ball tracking have been achieved by the conventional works. And the left ones are the main works in this article.

Vision based sports video analysis

There have been several researches targeting on the development of the automatic game analysis system. Work [7] proposes a computationally efficient hybrid method for automatic sports highlights generation to make contributions for the broadcasting applications. Method [8,9] proposes a trajectory and action recognition of the player to analyze soccer

training videos. This work has strict limitations on the environment and it only can be used in single player training scenario. Work [10] predicts the team events by analysis of the player motions and performance in basketball and water polo. This work analyzes the data by transferring the input video to overhead view, so that only 2D team sports can be used. Work [11] estimates the team tactics in soccer game videos based on the deep learning method and unique characteristics of tactics. Work [12] presents techniques for automatically classifying players and tracking ball movements in game video clips to analyze basketball movements and pass relationships.

Targets and results of above methods are far from the requirements of automatic Data Volley. Therefore, there is no state-of-art method can be used directly to achieve our goal. Since the automatic Data Volley system can be broken down into multiple tasks, the subsequent contents discuss about the related works for each task.

Detection of play scene

With the similar target for detecting the available game playing sequence, work [13] analyzes the soccer game structure to classify shot-views and segment play-back. This work is applied for the broadcasting videos based on the editing in broadcast video without using game status. There is another work utilizing player feature based method to detect the start scenes [14] in actual badminton matches. The input of this system is the video captured by the ceiling camera from the top of the court. By simultaneous extraction of spatial-temporal features from the motion images, features of players' postures and motions are extracted, which is used to detect the serving. However, this work cannot handle the complex situation in volleyball videos, in which the feature of the server player is difficult to extract and the background noise is heavy.

Multiple players tracking

For multiple people and object tracking, there are large amount of works [15–17]. Most of them cannot be directly used in sports videos due to the special features of the sports scene, such as the severe occlusion between players with similar appearances (features), the complex background, and the fast motions of the target player. To deal with the occlusion and similar appearances problems, utilization of the temporal feature [18] is a feasible solution. Targeting at the multiple player tracking in sports video, traditional work in computer vision analyzing sports videos [19] has focused on tracking players [20]. In order to distinguish players from each other, Yamamoto et al. [21] performs brute-force SIFT features matching between learned features and extracted features of player's jersey number, which is weak at occlusion and complex background noise. Work [12] proposes a player identification method based on jersry number detection and player tracking based on the Yolo framework [22]. To handle the severe occlusion problem, Ikoma et al. [23] proposes a 2D elimination method which removes all other objects' regions in the frame. And for the tracking after occlusion/overlapping, Huang et al. [24] utilizes the motion vector of positions at two previous time points to predict player's position after occlusion.

In addition, most above multiple players tracking algorithms are initialized manually. The research targeting on automatic initialization of tracking [25,26] are based on object detection results. These detection based methods are weak at the occlusion and similar appearance problems, which often occur in volleyball game.

3D ball tracking

Work [27] summarizes the challenges of the ball tracking in sports video: fast speed, small size, and influence of other items in the court. A large number of research works [28–30] targeting at the ball tracking based on the 2D images and the results are presented in image coordinate system. As for the 3D ball tracking, Chen [31] proposed an automated system to approximate the 3D trajectory of the ball. However, the lack of multiple space information makes their approximated result unreliable and the tracking accuracy is not high. Work [32] and Takahashi [33] proposed a multi-view based 3D ball tracking in volleyball game to obtain the physical 3D data. Compared with Takahashi's work, our framework avoids the 3D reconstruction process (which causes large error) in the 3D tracking framework so that the result achieves higher accuracy.

Event detection

Many similar works for event detection and recognition focus on broadcasting video analysis. These works refer to some post-process information like text feature on the screen and inserted audio in work [34]. Wang [35] proposes a framework using visual feature and audio feature to detect the event for soccer. But for the purpose of automatic data acquisition, event data are expected to be collected based on the game content itself. Based on the pure vision information, [36–38] propose methods based on image clustering technologies to detect the events for soccer games. Yang et al. [39] and Guo et al. [40] classify the game videos based on the detection of the players' actions, which cost large calculation on the detail information of individual players. Work [41] proposes a ball event detection method while ball tracking. This work classifies the event into four categories based on the ball trajectory, which is different from the requirements of Data Volley.

Event evaluation

In order to analyze the quality or evaluate the performance of the sports, some researches [42,43] focus on the method of statistical analysis. By accessing historical data, these works analyze different factors of the overall games and draw up evaluation report based on certain criteria. These researches focus on the performance of the whole team and do not pay much attentions on a certain action or event. To obtain quality information of the receive event, we proposed a framework [44] for qualitative action recognition for volleyball game analysis. This work evaluates the quality based on the return ball quality and the posture quality, which is different to the definition of event evaluation in Data Volley. In general, there is few research targeting on the event quality evaluation. Since the event quality is defined according to specific game rule, the very few existing method cannot be used as a comparison.

Automatic Game data acquisition

Overall framework

The conceptual setup of the entire automatic data volley system is designed as Fig 3. The equipment consists of multiple high resolution cameras and a computer to collect and analyze the game data.

The multiple cameras are used to record the game from different view-angles so that we can obtain multi-view videos as the input. The reason multi-view videos are used in this system is because it is difficult to construct precise 3D coordinate from single view information. Although there are some works [29,31] estimating the 3D coordinate only using single video, it requires heavy algorithms to compensate the reconstruction error. In addition, the multi-view videos are robust for occlusion situation. In volleyball game, there are always twelve players in the court who share same appearances and are overlapped by each other. In order to ensure a high data precision, multiple cameras are used to reduce the difficulties of the occlusion problem.

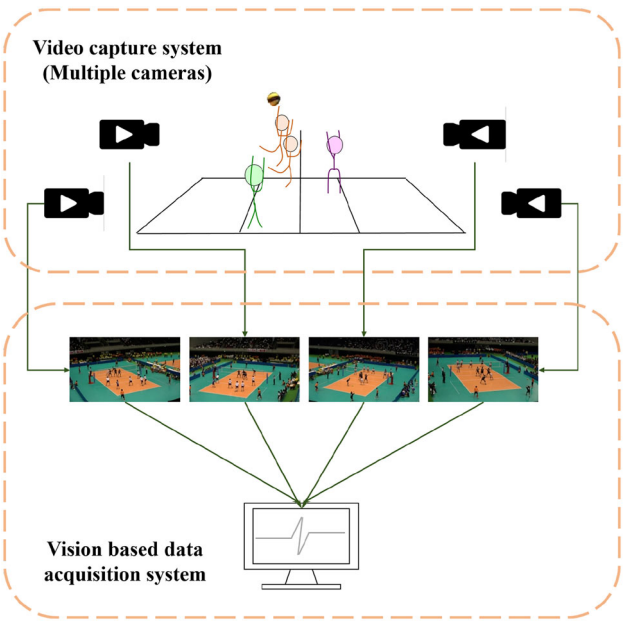

Fig. 3 The conception of the entire automatic data volley system setup

With the input multi-view game video, the vision based data acquisition algorithms are implemented on the computer. In this system, the overall framework and tasks are shown in Fig 4.

First, the preprocessing consists of the multi-video synchronization, camera calibration, and the play scene detection. The video synchronization is for aligning different views to fully utilize the image information. The camera calibration is the key to create the projection relation between each image to the real physical world. With the input videos, the play scene detection method outputs the time at which one round begins and the subsequent data acquisition process starts.

Second, the basic data acquisition consists of the physical data tracking and the event detection. For the physical data tracking, 3D ball tracking [32] and multiple players tracking [45] are implemented to obtain the 3D trajectories of the ball and players. Here, we proposed a time-vary fission filter to approximate the initial distribution of the whole team to automatically initialize the tracker.

Third, for the event detection, based on the collected historical trajectories and the image features, a simultaneous tracking and event detection framework is used. Here, we proposed a team formation mapping with sequence motion based event detection method.

At last, for the evaluation process of each event, we propose a relative spatial filter based quality evaluation method.

By connecting combining the output of all the steps, the required data of automatic Data Volley are collected and the following processing of data analysis and strategy development can be applied. The detail algorithm of each proposal is described in the following sections.

Temporal-spatial filters

In this work, three methods are proposed for the acquisition of different game data, based on the same core concept: the fusion of the temporal and spatial information. The temporal correlations of the position states and event states make contribute to predict the state and extract objective features. And the spatial information consists not only the image contents, but also the physical information in the 3D world, such as the

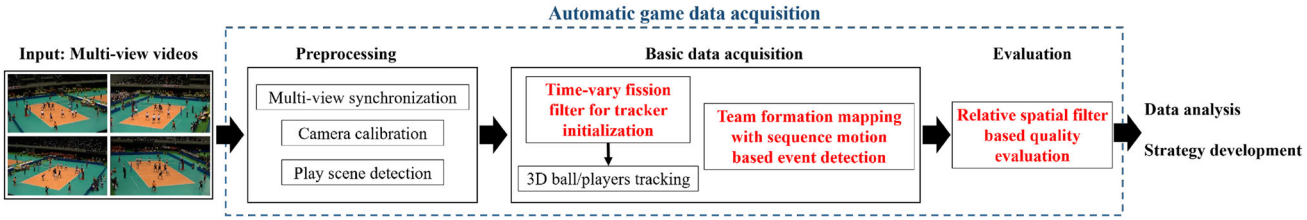

Fig. 4 The automatic Data Volley system consists of two parts: the automatic game data acquisition and the data analysis/strategy development. The overall framework and tasks of the automatic game data acquisition is marked with rectangle, in which the proposals are denoted with red color

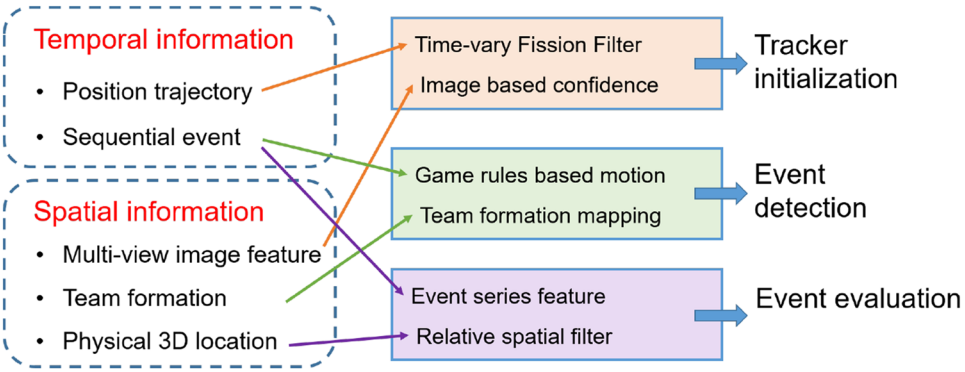

Fig. 5 The concept of the temporal-spatial filters

player's height and the team formation. Fig. 5 shows how the concept of the temporal-spatial filter works in each proposal.

First, with the target of automatic tracker initialization, the distribution of each player is fissured from the team state distribution. The image feature is extracted to calculate the confidence value of each player and the temporal feature of trajectory is utilized to distinguish the players from each other. Second, the event detection is achieved through combination of the sequential motion and the team formation mapping. The sequential motion refers the order of the event specified by the game rules. The team formation mapping describes the event feature from the perspective of the whole court space, not just looking at the state of individual player. Third, the temporal and spatial features are extracted respectively for different evaluation criteria. The event series feature refers the subsequent events, which representing the consequence of the event, while the relative spatial filter utilizes the additional 3D location information. Therefore, the overall automatic data acquisition system combines the advantage of both the temporal and spatial information.

Time-vary fission filter based automatic initialization

In order to initialize the player tracker, the 3D positions of all the players are required, and the players in each team should be distinguished with each other. To estimate the location states of the players in one team, we propose a time-vary fission filter, which approximates the process from one distribution separating to six of sub-distributions, that means simulating the state of one team to six players.

First of all, let us give a description of the player motions at the game start scene. We name the two teams as team-defense and the team-offense. One round game starts from the team-offense serves the ball. Before the serve event, both teams wait at standby formation, which is based on the strategy of each team and there are some common formations. At the common formation, each player stands at the fixed area. When the server of the team-offense hits the ball, which shows the game starting, based on the movements of the ball, two teams will move and adjust the formation. After the ball moves over the net, the team-defense adjusts the formations to organize one attack and the team-offense prepares to change the formation to defense.

Based on this fact, we define the team state distribution to present the position state of the team and propose the time-vary fission filter to estimate the distribution of each player to initialize the player's tracker. The initial team state distribution is the rough position distribution of the whole team and the distribution of each player is the sub-distribution fissured from the team state distribution. At the start scene, no matter where the individual players is locating, the team state distribution only presents the probability of the team covering the court. As the game goes on, the team adjusts the formation and all the players are in dynamic states. Based on the players feature extracted from the multi-view images, the distribution can be filtered by a weight of the player probability. As long as the player feature is reasonable, the distribution in the court space would present several peaks, which repre-

sent the high probability of this area with players existing. According to the amount of the peaks, the team distribution is separated into several sub-distributions and each distribution represents at least one players. This process is repeated at each time step until the six sub-distributions are filtered out. At this time, one sub-distribution represents the position state of one player, and the sub-distribution can be used directly to initialize the player tracker.

Compared with other tracker initialization method, the advantage of the proposed time-vary fission filter is presented on handling the occlusion and same appearance problems of players. The conventional works detect the targets by trained model and use the detected results to initialize the tracker. In volleyball game, all the players wear the same uniform and are often occluded by others. The occlusions and similar appearance of targets limit the performance of detection. To deal with the specific problems in volleyball, the proposed time-vary fission filter use the temporal continuity of image feature to generate the distribution of each player. The distribution of one player is filtered out at the moment when the player is not occluded. And the filtered order is used as label to distinguish the player from others.

The detailed algorithms are introduced as below.

State definition and initialization

First, the team state X is defined as

$$\mathrm{X} = (\mathbf{z}, n, s)\,, \tag{1}$$

where, the coordinate $\mathbf{z} = (x, y, z)$ represents the 3D position in the court space. n represents the account number of the players and s is the serial number of the distributions. The value of n and s are the integers from 1 to 6.

$$\sum_{s=1}^{6} n = 6\,, \tag{2}$$

since there are six players in each team. It can be assumed that one player, whose center position is located at (x, y, z), belongs to the s_{th} sub-distributions and there are n players in the s_{th} sub-distributions. Since the target distribution of the team state consists of both continuous elements and discrete elements, it is difficult to use one simple model to represent the high-dimensional state. In the proposed algorithm, a variety of the team state samples are generated to approximate the distribution as precise as possible. Therefore we use the samples set $\mathbf{z}^{(i)}$ to describe the team state.

Initial distribution of the team state at time k=0 follows

$$\mathbf{z}_0 \sim N(\bar{\mathbf{z}}, \Sigma_0),\ \Sigma_0 = diag(\tau_x^2, \tau_y^2, \tau_z^2)\,, \tag{3}$$

where $\bar{\mathbf{z}}$ is the given mean value of the distribution, which follows the approximated distribution of the standby team formation. τ_x, τ_y and τ_z are the Gaussian noise variances for position term at different directions.

The value of $\bar{\mathbf{z}}$ is decided referring the game rules and statistic game data. In order to initialize the value of $\bar{\mathbf{z}}$, a sequential integer number m ($m \in [0, 5]$) is generated. Since there are 6 players in each team, the variable m, which is from 0 to 5 is capable to number them. In this research, we use a Gaussian distribution H_m to represent the spatial probability of the m_{th} player. As preparation, some 3D players positions data are manually collected and divided into 6 groups based on different player roles. As for the m_{th} player, the mean coordinate and the variance of the data is defined as the mean value C_m and variance $[\sigma x_m, \sigma y_m, \sigma z_m]^T$ of distribution H_m:

$$H_m \sim N(C_m, \Gamma_m),\ \Gamma_m = diag\left(\sigma x_m^2, \sigma y_m^2, \sigma z_m^2\right)\,. \tag{4}$$

After the value of $\mathbf{z}_0$ is initialized, we assign $n_0 = 6$ and $s_0 = 0$. The physical meaning of the initialization is that at time $k = 0$, there is no prior information of the positions and identity labels for all the players. At this situation, all the players has no difference with each other. And we only know they belong to the same team, so that all of them belong to the 0_{th} sub-distribution and there is only one sub-distribution. The positions of players are generated according to the standby team formation.

Up to here, a prior team distribution is obtained.

Confidence calculation

Our goal is to filter (separate) the prior distribution via the temporal video so that the precise distributions of all the players in the team can be obtained. The distribution, which is represented by a set of team state samples, is filtered based on the confidence value of each samples. The confidence value is calculated according to the image content from different views. We define the confidence $\mathbb{I}_k$ as a collection of image frames at discrete time k:

$$\mathbb{I}_k = \left\{\mathbb{I}_k^1, \mathbb{I}_k^2, ..., \mathbb{I}_k^m, ..., \mathbb{I}_k^M\right\}\,, \tag{5}$$

where M is the total number of views. Thus, the confidence $C\left(\mathrm{X}_k^{(i)}\right)$ of each sample is indicated as:

$$C\left(\mathrm{X}_k^{(i)}\right) = g\left[C\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k^1\right), ..., C\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k^m\right), ..., C\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k^M\right)\right]\,. \tag{6}$$

Here, the $C(\mathrm{X}_k^{(i)}; \mathbb{I}_k^m)$, which is defined as "image confidence" of the i_{th} sample estimated from the observation

region in the m_{th} view at the state $\mathrm{X}_k^{(i)}$. $g(x)$ is a function to combine each image likelihood obtained from each camera. Here we use three image features of the players: the color feature, the motion feature, and the body part categorization based feature.

For the color feature and the motion feature, we use the same one that the player tracking used. With the prepared image templates, the distance between the observation region and the templates are used to represents confidence of color. Motion feature $C_{\text{motion}}\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k^m\right)$ is obtained according to the background subsection.

In the team distribution, more than one players exist. To evaluate this confidence, the factors related to the player amounts are required. To avoid the severe occlusion of players, the amount and the probability of the body parts are used.

Therefore, the body part category based confidence is calculated as

$$\begin{aligned} C_{body}\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k\right) = d\left(n_k^{(i)}, N_{\text{hand}}\right) {P_{\text{head}}}_k^{(i)} \\ \times d\left(2n_k^{(i)}, N_{\text{head}}\right) {P_{\text{hand}}}_k^{(i)} \\ \times d\left(2n_k^{(i)}, N_{\text{foot}}\right) {P_{\text{foot}}}_k^{(i)}, \end{aligned} \tag{7}$$

where, the N_{hand}, N_{head}, and N_{foot} are the total amount of the detected efficient body parts of the heads, hands and feet. The function $d()\dot{}$ is to calculate the confidence of the amount depending on the deviation from the predicted amount $n_k^{(i)}$ to the detected one. The reason these body parts are chosen is because the head, hand, and feet number directly decide the players' number. As for other body parts, such as the torso, shoulder, arm and leg, which are always overlapped and confused with each other when two or more players move close.

The ${P_{\text{head}}}_k^{(i)}$, ${P_{\text{hand}}}_k^{(i)}$, and ${P_{\text{foot}}}_k^{(i)}$ are the probability of the detected head, hand and foot, which are calculated as:

$${P_{\text{head}}}_k^{(i)} = \sqrt[M]{\prod_{m=0}^{M-1} P_{\text{head}}(\mathbb{I}_k^m)}, \tag{8}$$

$${P_{\text{hand}}}_k^{(i)} = \sqrt[M]{\prod_{m=0}^{M-1} P_{\text{hand}}(\mathbb{I}_k^m)}, \tag{9}$$

$${P_{\text{foot}}}_k^{(i)} = \sqrt[M]{\prod_{m=0}^{M-1} P_{\text{foot}}(\mathbb{I}_k^m)}, \tag{10}$$

where, the $P_{\text{head}}(\mathbb{I}_k^m)$, $P_{\text{hand}}(\mathbb{I}_k^m)$, and $P_{\text{foot}}(\mathbb{I}_k^m)$ are the probabilities of detected efficient head, hand, and foot in the m_{th} view. Body part categorization is based on the hypothesis that the perspective image contents of one body part in different views should be detected as the same body part category, even the appearances of them are different. In each view, we can obtain a region of interest of each team state sample by projecting the 3D position to the image.

Here we train a detector using convolutional network to detect the body parts and the probability. 10,000 labeled images are used. Based on the output probability, we could obtain a Heatmap of the body part. Based on the Heatmap, the position and the probability of the detected body part can be obtained. Then, we count the detected body parts in the distribution area to obtain the value N_{hand}, N_{head}, and N_{foot}.

The total confidence $C\left(\mathrm{X}_k^{(i)}\right)$ is calculated as:

$$\begin{aligned} C\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k\right) = C_{body}\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k\right) \times C_{\text{color}}(\mathrm{X}_k^{(i)}; \mathbb{I}_k) \\ \times C_{\text{motion}}\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k\right), \end{aligned} \tag{11}$$

and

$$C_{\text{color}}(\mathrm{X}_k^{(i)}; \mathbb{I}_k) = \sqrt[M]{\prod_{m=0}^{M-1} C_{\text{color}}\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k^m\right)}, \tag{12}$$

$$C_{\text{motion}}(\mathrm{X}_k^{(i)}; \mathbb{I}_k) = \sqrt[M]{\prod_{m=0}^{M-1} C_{\text{motion}}\left(\mathrm{X}_k^{(i)}; \mathbb{I}_k^m\right)}. \tag{13}$$

Time-vary fission filter

Filtering at $k = 0$

With the confidence of each sample is obtained, the posterior distribution can be filtered to approximate the true distribution of the team. Here, the normalized confidence weight of each sample is denoted as:

$$\mathbf{w}_0^i \sim p(\mathbb{X}_0 \mid \mathbb{I}_0) = C\left(\mathrm{X}_0^{(i)}; \mathbb{I}_0\right). \tag{14}$$

The filtering model is based on the Monte Carlo method and Bayesian equation [46]. We randomly generate a number on (0, 1) and look up the corresponding state sample through the cumulative probability distribution of the confidence weight of the state team distribution. The theoretical basis of this step is that the team state sample with larger confidence has higher probability to be filtered out for the posterior distribution. By repeating the random number generation and the team state sample filtering enough times, we can obtain a new distribution $p(\mathbb{X}_0; \mathbb{I}_0)$. Yet this is not what we expected for the posterior distribution.

Up to here we only filter the position state, and the player identity states are filtered as following algorithm. First, all the samples are clustered by the Mean Shift [47] according to their spatial 3D coordinates. Based on the clustering

result, the team state distribution is separated to several sub-distributions $\mathbf{D}^{K_\alpha}$, and the K_α is the serial number of the cluster.

For the $K_{\alpha\text{th}}$ sub-distribution, we assign the value of $s^{(i)}$ as:

$$s^{(i),K_\alpha} = K_\alpha \,, \tag{15}$$

which means this samples belonging to the $K_{\alpha\text{th}}$ sub-distribution, and the $s^{(i)}$ value is assigned as K_α.

Also, for the $K_{\alpha\text{th}}$ sub-distribution, the projected image region can be decided based on the 3D positions. We check the body part classification result, which is mentioned in section *Confidence calculation*. Based on the detected head number, foot number and hand number, it can be know how many players are there locating in this area. And the player numbers N_{K_α} are assigned to the value $n^{(i)}$:

$$n^{(i),K_\alpha} = N_{K_\alpha} \,. \tag{16}$$

System model at $k = k + 1$

When the team state distribution moves to the next time step, which means

$$k = k + 1 \,, \tag{17}$$

we simply assume that the players move randomly, and transfer the position state with Gaussian noise. So the system model is

$$\mathbf{z}_k = T \cdot \mathbf{z}_k + W_k \,, \tag{18}$$

where $\{W_k, k \in N\}$ is the Gaussian noise term, which is defined as

$$W_k \sim N(0, \Sigma_0), \Sigma_0 = diag(W_x^2, W_y^2, W_z^2) \,. \tag{19}$$

As for the player identity state,

$$n_k = n_{k-1}, s_k = s_{k-1} \,. \tag{20}$$

Filtering at $k > 0$

For the sub-distribution in which the $n_k = 0$, we only filter the position state based on the samples confidence and keep the same identity information, since the $n_k = 0$ represents this player has already be distinguished from other players.

Then, for the sub-distribution in which the $n_k \neq 0$, both the position state and the identity state are filtered as same algorithm when $k = 0$.

And for each frame, the team state distribution are iteratively processed according to the system model, the confidence calculation and the filtering process until all the player number n_k of all the sub-distributions is 0. And the set of distributions $\mathbf{D}^{K_\alpha}$ can be used to initialize the multiple player tracker.

Team formation mapping with sequence motion based event detection

Simultaneously tracking and event detection

As we mentioned at the Sect. 3, the multiple players tracking [45] and the 3D ball tracking [32] is implemented to acquire the 3D physical data. Simultaneously, the ball event is also estimated. The rough algorithm is introduced as following:

First, for the player tracking, the serial number s obtained at player initialization is used to manage multiple targets in one team. C is the set of all numbers. Then each player's state is defined as $\mathbf{z}_k^c$, and the state is transited as

$$\mathbf{z}_k^c = \mathbf{z}_{k-1}^c + \Omega_k \cdot \Delta T \,, \tag{21}$$

where Ω_k is a three-dimensional noise in the prediction model combining a Gaussian model and a least square fitting prediction model. ΔT is the sampling time interval of the video. Observation of each player includes color likelihood and sobel gradient likelihood. The estimated player positions are indicated as $\hat{\mathbf{z}}_{0:k}^c$.

Second, the 3D ball tracking and event change detection are employed in one model. the 3D ball state (including the event state) at a discrete time k is denoted as

$$\mathbf{E}_k = (\mathbf{y}_k \,,\; d_k \,,\; e_k) \,, \tag{22}$$

where $\mathbf{y}_k$ is a pair of ball physical states including the three dimensional position and velocity. d_k is the motion of the ball and the last term e_k is the event type. Dynamic of the ball physical state is modelled in a time difference equation:

$$\mathbf{y}_k = \begin{bmatrix} \mathbf{I}_3 & \Delta \mathrm{T}\mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{y}_{k-1} + \begin{bmatrix} \frac{\Delta \mathrm{T}^2}{2}\mathbf{I}_3 \\ \Delta \mathrm{T}\mathbf{I}_3 \end{bmatrix} \mathbf{w}_k \,, \tag{23}$$

where system noise term $\mathbf{w}_k$ is three-dimensional random vector that represents external force added to the ball, which is a mixture of general noise and abrupt noise. The mixture of noise $\mathbf{w}_k$ and the transition of e_k is decided by the prediction of d_k. The observation is calculated based on the multi-view images and the past ball trajectory features. At each sampling time, the estimated state $\hat{\mathbf{y}}_{0:k}$ by particle filter is the tracked 3D ball trajectory and event. The estimated state $\hat{d}_{0:k}$ updates when the event changes. As for the observation and estimation method for the event type $\hat{e}_{0:k}$, we proposed the team

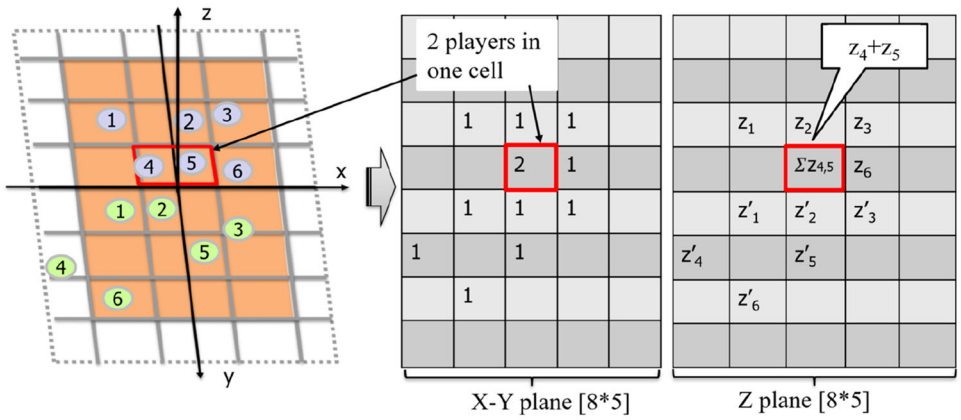

Fig. 6 The conception of the team formation mapping method

(a) Court dividing **(b)** Mapped matrix: $\boldsymbol{T}$ [8*10]

formation mapping method and the sequential ball motion feature, which are introduced in the following subsections.

Team formation mapping

Team formation refers to how the team is organized to strengthen the power of their defensive or offensive activities. There are two situations for the defensive activity. When the team is going to handle opponent's serving, the receiver will move to the position against the ball's moving direction, and the others will stand around to get ready for saving the ball while the setter stands by in front of the net. When the team tries to stop the opponent's Attack, blockers will gather near the net and the others will fill the blank of the back court to dig the ball if necessary. As for the offensive situation, that means the team is going to perform the Attack, potential attackers will move towards the net, and the setter usually hits the ball near the net to decide the attacker. Based on the discussion above, when different events happen, the team will perform certain team formations on the court.

The team formation mapping method describes the feature of team formation when certain event happens. Fig. 6 shows the details of this method. The court is divided into zones of five rows and eight columns on the $x-y$ plane as the Fig. 6(a) shows. The principle of court division is based on the design of the court line and the game rules, which is summarized as follows.

- The size of a half court is a square of 9-meters×9-meters, and a 3-meter line is set to restrict the positions of attackers. The court inside the sideline and end-line are divided into 6×3 grids with the sides of 3 meters.
- As the volleyball game allows hitting the ball outside the boundary, especially for some emergency, so the area outside the court is also taken into consideration and divided by every 3 meters along the court line.
- In the vertical scale, the player amount denotes the formation status and the height of player reflects the players' action. Both of them are important features of the event.

The mapped matrix Υ_k is the descriptor corresponding to the three dimension of the players' positions, consisting of two parts: the X-Y plane matrix E^{XY} and the Z plane matrix E^{Z}. The element in X-Y plane matrix denotes how many players are there in the corresponding zone of the court. And the element in Z plane is the sum of these players' height like the Fig. 6(b) shows. The mapped matrix is obtained by connecting these two matrices by row.

For the positions of the players from both team $\hat{\mathbf{z}}^{c}_{0:k}$, where $c \in [1, 2, ..., 12]$, the element of the two sub-matrices is defined as $E^{XY}_{a,b}$ and $E^{Z}_{a,b}$ respectively, where $a \in [1, 2, ..., 8]$ and $b \in [1, ..., 5]$. The process of mapping is described as follow:

First, the initial values of $E^{XY}_{a,b}$ and $E^{Z}_{a,b}$ are set to zero. Then, for the s_{th} player, if the position $\hat{\mathbf{z}}^{s}_{0:k} = (\hat{x}, \hat{y}, \hat{z})$ locates in the zone (a, b), the two elements are calculated by:

$$E^{XY}_{a,b} = E^{XY}_{a,b} + 1, \tag{24}$$

$$E^{Z}_{a,b} = E^{Z}_{a,b} + \hat{z}. \tag{25}$$

As the feature vector which is a part of the input to the classifier for recognizing the event type, the X-Y plane and the Z plan are connected by row. Then the overall mapped matrix will be reshaped to a one-dimension vector by connecting the end of one row to the head of the next row. Thus the reshaped feature vector is represented as:

$$\Upsilon_k{}' = \left[E^{XY}_{1,1}, E^{XY}_{1,2}, ..., E^{XY}_{1,5}, E^{Z}_{1,1}, ..., E^{Z}_{1,5}, E^{XY}_{2,1}, ..., E^{Z}_{8,5}\right]. \tag{26}$$

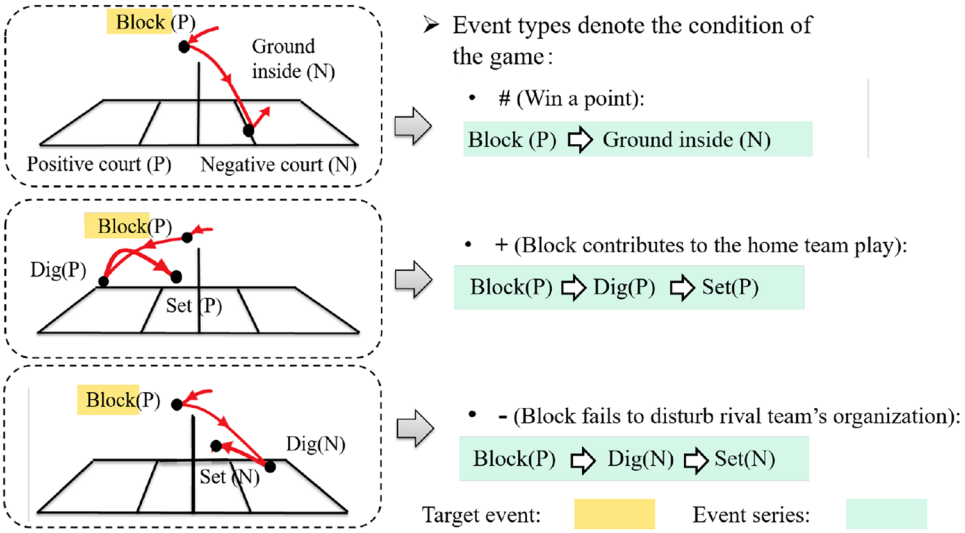

Fig. 7 The definition of event series feature

Sequential ball motion feature

According to the rules of volleyball game, the order of each event is fixed. Based on the discussions in [48], the status of volleyball game is divided into three categories: Attack Process, Counterattack Process and Emergency.

1. Attack Process: The process for a team to deal with opponent's Serve. This process starts from Serve, Reception, to Set and then to Attack.
2. Counterattack Process: A series of events starts from opponent's Attack and ends when the team returns the ball. In some situations, Block cannot be performed successfully so it will start from a Dig.
3. Emergency: A receiver or digger failed to control the direction of the ball. Players usually try to dig the ball when the emergency happens.

Since the team formations of all the events have some regular patterns like the Attack and Block are always performed near the net, and the preference of hitting position always occurs at dense area. So the ball motion of certain event highly relates to that of the former events. Therefore, the sequential ball motion feature is described as follows.

S_j is defined as the ball motion state of the j_{th} event counting from Serve of each round. Since $\hat{\mathbf{y}}$ represents the estimated physical state of the ball, and $\hat{\mathbf{y}}_k = \begin{bmatrix} \mathbf{x} \\ \dot{\mathbf{x}} \end{bmatrix}$, where $\mathbf{x}$ and $\dot{\mathbf{x}}$ are position and velocity of the ball in the three dimensional space. So the S_j is defined as:

$$S_j = \hat{\mathbf{y}}_j. \tag{27}$$

Thus the sequential ball motion state feature S_{seq} for the j_{th} event of one round is defined as

$$(S_{\text{seq}})_j = [S_{j-1}, S_j]. \tag{28}$$

For the occasion when the target event is Serve, since it is the first event of one round, the former ball motion state are set to zero.

Relative spatial filter based quality evaluation

According to the summary of the evaluation basis in Table 1, although the bases of all the event are different from each other, they can be roughly divided into two types. One refers the subsequent event series to evaluate the event quality; the other one requires additional information to make judgment. Therefore, this work gives two proposals to obtain the overall evaluation results.

Event series feature

For each event, there are three potential results for the target event: *hit by home team*, *hit by rival team* and *hit the ground*. Based on these three results, we define two corresponding factors in event series feature: the side of the court and the the ball hitting the ground.

The definition of event series feature is shown in Fig. 7. In this figure, the target event to be evaluated is the Attack. And on the court where the attacker is recognized as the positive court. For this situation, the Attack is blocked by the opponent and the ball directly hits the ground, which indicates losing a point for the attacker's team. Here, the event series which contain the event type and their belonging court describes the condition whether the Attack is blocked. Thus the judgment can be made referring the game rules.

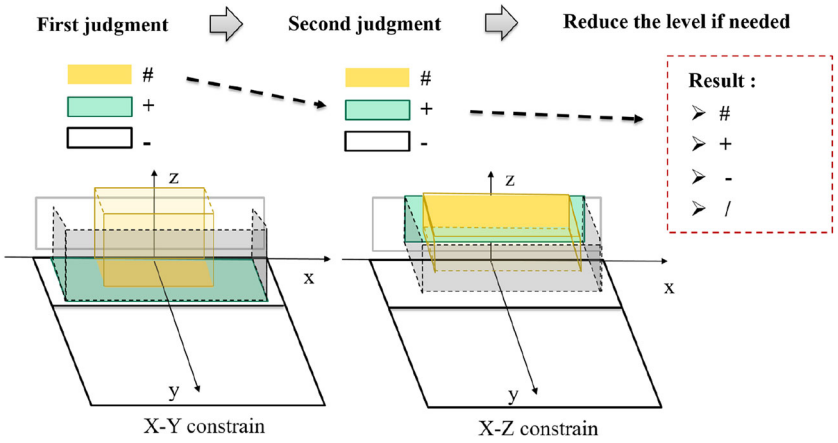

Fig. 8 The process of judging Receive quality by setting point

Relative spatial filter

The relative spatial filter is designed for specific categories for Receive and Set quality evaluation, since evaluation of these event required additional information besides event series. According to the summarise in Table 1, the judgment of Receive is based on the organization of Set. While the evaluation of Set needs to get the number of available blockers from opponents who play against the following Attack. Therefore, two filters are proposed to deal with these two events: the filter for Receive quality evaluation and filter for Set quality evaluation.

Filter for Receive quality evaluation

To evaluate the Receive quality, the hitting position of the following Set is the main basis. There are two factors making difference to the setting condition. First, since the setting position is decided by how the receiver pass the ball, it's clearly relative to the quality of Receive. When the ball is hit close to the center of the court, the possibility of the setter passing the ball to different directions tends to be equal. At this condition, the setter's team has a higher possibility to score a point. It is difficult for the blockers of rival team to judge which Attack should block against until the ball leaves the setter' hands. So this will disturb the judgment of the blockers of the rival team. Second, the pose of the setter is relative to the quality level. There are two poses generally used for setting the ball, the Bump and Overhand. Basically, when using the pose Overhand, the setter has a better control on the ball so that the Receive is considered to be better if the setter hits the ball using Overhand. It can be easily recognized that the hitting points of this two poses' tends are usually different. To be specifically, the hit point of Overhand is commonly higher than that of Bump.

Based on the two factors that are relative to the setting organization, a three-step spatial filter is used to evaluate the Receive quality as shown in Fig. 8. Firstly, the hitting point of Set is judged by the $x-y$ constraining with three levels, which refers to the discipline that the closer the hit is to the court center, the better it is. Secondly, the hitting point is judged by the horizontal constraint, which refers to the relationship of Overhand and Bump's hitting point. This step also has three levels decided by the x-z constraint. Thirdly, to get the final judgment result, we take the first judgment as the basis, and if the level in step 2 is worse, the result will decrease a corresponding level. So the final result will be the quality of the corresponding Receive.

Filter for Set quality evaluation

To obtain the quality of Set, the number of opponent's blockers against the following attack is the key factor. The method to find out available blockers based on the assumption: the blocker should be close enough to the ball and jump in front of the net at certain moment to make it possible to block the ball. If the motion of the player does not meet the conditions, it is assumed that this player is not an available blocker.

Based on this assumption, the moment to search for the blocker should be make sure at first. There are two situations after the ball is attacked. One is the condition that the blocker hits the ball, the moment is chosen as the Block hit point. The other one is the condition when the ball crossed the middle line and is not far from the net is chosen. After deciding the moment for searching, available blockers will be filtered out by their height of torso, which refers to the value of their z-direction positions, and their relative distance to the ball.

Experimental results

Experiment environment and data set

The resource videos of the experiments in this work records the Semifinal Game and Final Game of 2014 Japan Inter High School Games of Men's Volleyball in Tokyo Metropolitan

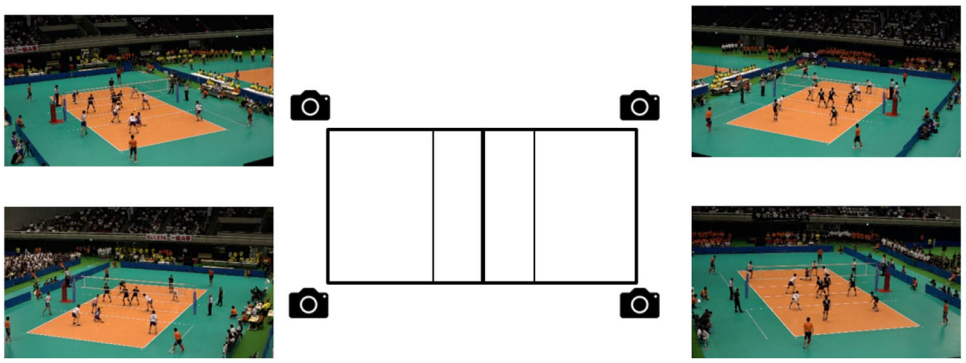

Fig. 9 Camera position of the data set

Table 3 Data sets of events

	Serve	Receive	Set	Attack	Block	Dig	Free Ball
Semifinal	14	23	68	77	33	41	5
Final	10	13	42	41	14	41	3
Total	24	36	110	118	47	82	8

Gymnasium. The camera views chosen for the test data set are the four corners of the court as it is shown in Fig. 9. The resolution of all videos is 1920×1080, with the frame rate of 60 frames per second and the shutter speed of 0.001 second.

The experiment is executed with the following environment setting: the CPU is Intel Core i7-3770, the RAM is 8GB, the complier is Visual Studio 2017, and the external library includes OpenCV-3.4.1 and GSL-1.6. In the implementation, for all the test samples we set the same parameters to create a universal method with high fitness to different conditions.

Data set for player tracker initialization

The automatic initialization is the post-process after the detection of the play scene. Therefore we cut the entire game sequence manually into different rounds, and each round starts at the detected frame of the player scene. Generally, most sequence starts at the serve event when the server throws the ball and is going to hit the ball. As for the distribution of the player formation, besides the server, all the players stand by following a fixed team formation. After the server hits the ball, all the players start to move to change the team formation. In our sequence, there are totally 286 rounds from 5 sets (There are three sets in the semi-final game and two sets in the final game).

Data set for event detection and evaluation

For each round, the trajectory information are acquired by the tracker initialization, ball tracking [32], players tracking [45] works. Then, among the correct tracking rounds, we labeled the event type manually. The overview of the data set is summarized in Table 3. The totally number of available event *Serve, Receive, Set, Attack, Block, Dig and Free Ball* are 24, 36, 100, 118, 47, 82, and 8. One thing to be mentioned is that the data set includes little numbers of the event Free Ball, so it is hard to conduct experiments. Besides, as it is discussed in previous chapters, this kind of event is not typical in general volleyball analysis. So the experiment does not take this event into consideration either.

Corresponding to the overall framework, the experimental results contain three parts: the initialization of player tracker, the event type detection and the quality evaluation. Since these three parts are based on different kinds of methods, their evaluation criteria are different as well.

Experiment result for player tracker initialization

Evaluation criteria

As the discussion in Sect. 2, the conventional works of tracker initialization are detection based methods, which only process on one frame and are evaluated by the detection accuracy. However, the proposed algorithm in this article should be evaluated from both the time and the space scale. Therefore, the conventional evaluation method of tracker initialization cannot be used to evaluate our algorithm. In the experiments, the evaluation criteria are defined as following.

First, from the time scale, the time efficiency of the algorithm is evaluated by the number of processing frames T_{frame}. we use the number of used frames T_{frame} to evaluate the time efficiency of the algorithm. T_{frame} represents how many frames are used before the tracker is initialized. In theory, the smaller number of the used frames, the higher performance of the algorithm on time scale. Second, from the space scale, we calculate how many players are correctly initialized after the initialization process finished. As long as the main body part of the player can be distinguished to initialize the

Table 4 Experiment results for time-vary fission filter based automatic tracker initialization

Evaluation	Number	Proportion
Success ($T_{\text{frame}} \leq 30$)	83	29.02%
Success ($30 > T_{\text{frame}} \leq 60$)	124	43.36%
Success ($60 > T_{\text{frame}} \leq 90$)	47	16.43%
Success ($90 > T_{\text{frame}}$)	15	5.2%
Failure	17	5.9%

tracker, we assume this player is correctly initialized. For each sequence, there are 12 players participate so that the total players required to be initialized are $12 \times 286 = 3432$.

Experimental result

Data in Table 4 are the experimental results of the proposed time-vary fission filter based automatic initialization for player tracker. For each round of game, the initialization of tracker is evaluated by two factors. First one is whether the tracker is initialized successfully. When the iteration of time-vary fission filter is finished, the twelve trackers of all the players are initialized and the tracking process continues smoothly. We assume this round is "successful round". Otherwise, this round is defined as "failed round". The second evaluation factor is based on the precondition of "successful round". According to the value of T_{frame}, the "successful round" can be classified into four types.

From Table 4, it can be known that there are only 5.9% rounds are failed, and more than 70% rounds can be initialized within 1 second (that means the value of T_{frame} is less than 60 frames). Due to the high frame rate, the players' positions are almost unchanged. The long time of the processing/iteration time is because of the sever occlusion between players, who shares the similar appearance. In order to reduce the processing time and increase the performance of the automatic tracker initialization, there are two possible ways. First, since the moving state of the players is not random, there is still large room for us to improve the current model. Our subsequent study focuses on learning a semantic distribution for representing the players moving state. Second, we could improve the model of confidence calculation to make it robust to the similar appearances of the targets.

In order to show the detailed processing and the intermediate results of the time-vary fission filter, we draw all the sampling point on one view. The results are shown as Fig. 10, we choose four frames during the process. At $T_{\text{frame}} = 0$, as Fig. 10a shows, there is only one distribution represents the state of the whole team. The different colors represent different teams. As Fig. 10b shows, the distribution is fissured into sub-distributions according to the confidence value. From Fig. 10c it can be known that even the occlusion

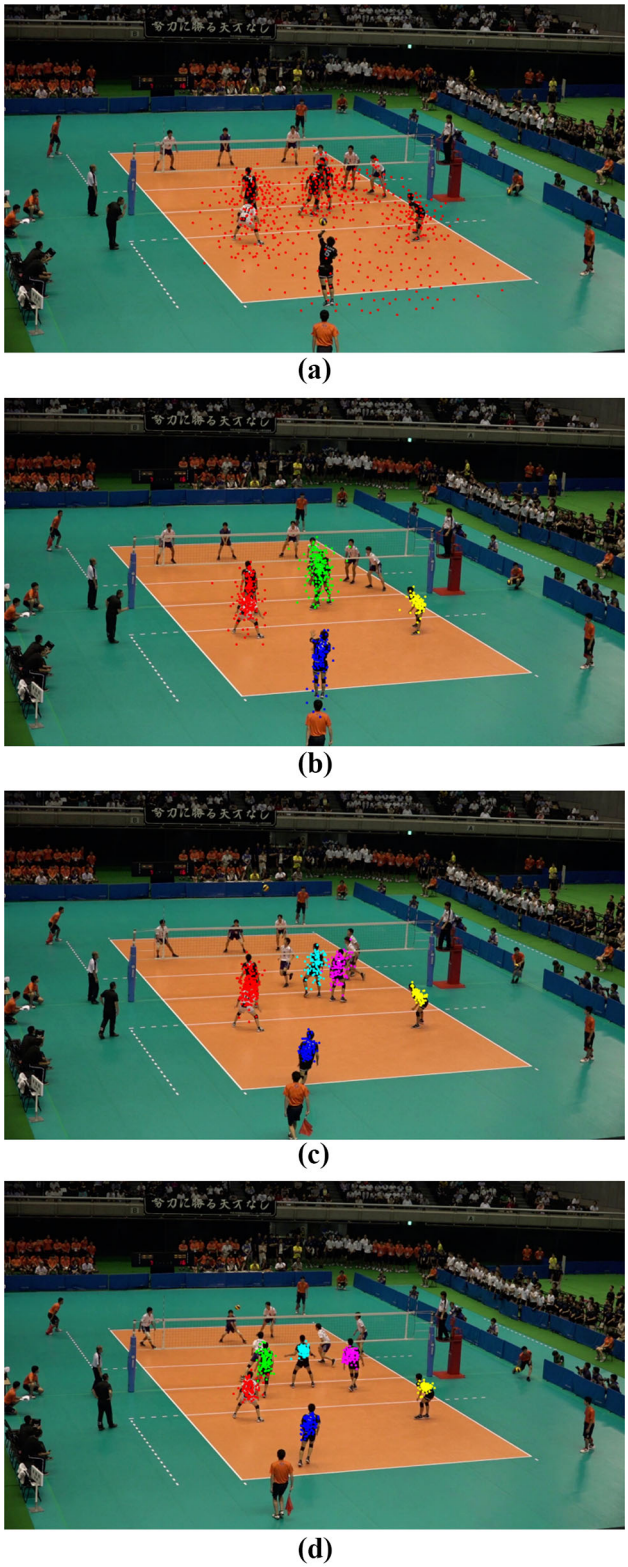

Fig. 10 Detailed processing and middle results of the time-vary fission filter

Table 5 Experiment results for event type detection

Method	Event	Precision	Recall	Accuracy
Conventional work [41]	Serve	100%	100%	100%
	Receive	66.67%	80.00%	86.52%
	Set	86.67%	70.72%	83.87%
	Attack	98.33%	95.16%	96.58%
	Block	77.27%	70.83%	84.62%
	Dig	79.41%	65.85%	77.42%
+ Team formation mapping	Serve	100%	100%	100%
	Receive	82.61%	95.00%	94.38%
	Set	82.61%	89.06%	87.33%
	Attack	98.25%	92.32%	94.02%
	Block	94.44%	70.83%	85.37%
	Dig	89.74%	85.37%	89.25%
+ Sequential ball motion feature	Serve	100%	100%	100%
	Receive	100%	100%	100%
	Set	96.97%	96.97%	97.33%
	Attack	100%	98.36%	98.28%
	Block	100%	100%	100%
	Dig	97.62%	100%	96.70%

happens, the fissured sub-distributions won't confused with each other. In Fig. 10d, all the players are distinguished and then the tracker has sufficient conditions to be initialized and work.

Although the comparison with other algorithm is not shown in the experiments due to the lack of suitable conventional work, the advantages of the proposed algorithm can still be observed. In Fig. 10(a), severe occlusion occurs between players with similar appearance. In this situation, it is hard to distinguish each players even by human eyes, let alone by other algorithms. From the conceptual aspect and the example, the proposed algorithm presents better on the occlusion and same appearance problems than the detection based methods.

Experiment result for event detection

The experiment of this step includes three parts. First, the idea of conventional work [41] is implemented and applied to the data set for recognizing six kinds of event types. Then the experiment combining conventional method and proposed method, the team formation mapping method, is conducted. Finally, the sequential motion state feature is combined with the former two methods to get a better result. The experiment results of the three parts are shown in Table 5.

Compared with the conventional work, who performs low ability in distinguishing *Receive, Set* and *Dig*, a much better result is achieved by combining the team formation feature. This proves that the team formation mapping method offers a good description for players' moving tendency of both teams on the court, which is relative to the purpose of each event. However, this method does not fully solve the problem to distinguish those similar events so a further experiment is conducted. For the result of combining the two methods, the improvement of each criteria is up to 20% comparing with the conventional work. This experiment also achieves a better result than that with formation mapping, especially for Set and Block. That means the sequential feature has the ability to distinguish events of similar hitting point thanks to the idea of referring to the former event corresponds to the principle for classifying events.

Experimental of event quality evaluation

For the event quality evaluation, there is no conventional work even for the similar target as we stated in Sect. 2. To evaluate the proposed algorithm, we define the concept of success rate as:

$$success\_rate = \frac{Correct\_detection}{Total\_events} \times 100\% . \tag{29}$$

The $Total\_events$ means the amount of the patterns for the target event type. As it is explained before, the quality evaluation not only relies on the following event, but also relates to some addition information including available blockers and setting positions. So for those cases that can be judged by the following events, the correct detection refers to the number of detected event series. Here, the detected event series means that if all the events, including the target event, of the series are correctly detected, this event series

Table 6 Experiment results for quality evaluation

Target event	Correct detection	Total amount	Success rate
Serve	24	24	100%
Receive	30	31	96.77%
Set	105	108	97.22%
Attack	112	117	95.72%
Block	46	47	97.87%
Dig	73	76	96.05%

is regarded as a correct detection. For the quality of Receive that judged by the setting point, we assume that if the judging result is the same with the ground truth, this case will be a correct detection. As for the quality of Set that relates to the number of available blockers, when the number of available blockers is correctly detected, we assume this event is a correct detection.

The experiment result for the quality evaluation part is shown in Table 6. It can be seen that the success rate of obtaining the quality of each target event is over 95%.

Analysis and discussion

Analysis of the algorithm performance

Firstly, for the evaluation of the automatic player initialization method, there are 5.9% sequences being failed. We analyze the failed sequences and summarized them into two big categories. One is that the round is very short, such as the serve ball cannot go cross the net or flies out of the court. In this situation, the team formation changes too little so that some players standing close with each other cannot be distinguished. For this kind of game round, as long as we know who is the server, the analysis of the player trajectories and event is not necessary. The other one is that two or three players cannot be distinguished. The main reason is our algorithm for calculating the player confidence failed, especially when the body part categorization provides a wrong detection result.

Secondly, for the result of event type detection, although the average accuracy reaches over 98%, there are still some abnormal cases that current methods failed to recognize correctly. These cases mainly happen between the Set and Dig. In this case, the feature of the Dig is very close to Set as the player only tries to adjust the direction of the ball. The reason to define it as Dig is that it is followed by a Set. Since the methods of this work only refer to current and historical information, the wrong historical message will negatively affect the performance. Thus, this problem is not so crucial in implementation of automatic system.

Thirdly, as the event series feature relies on the event type detection result, the wrong event detection also has a negative effect on the result of quality evaluation. The basis of judging the type is that the action of player trying to save the ball obviously lose the control from the image. This kind of cases belongs to the situation of emergency which is included in the definition of Dig. For current features only refer to previous trajectories of the ball and the player, it is hard to recognize the potential failure of the events. However, the pose of the player denotes the situation of mandatory hit, so it is potential to utilize the player pose to strengthen the ability in recognizing these cases.

Analysis of the processing time performance

Since the automatic Data Volley system is expected to be applied in the real game, the time performance is also an important factor of the system efficiency. In this article, all the algorithms are implemented without considering the time efficiency since the accuracy has the highest priority. Based on current platform (we don't use the powerful server), the processing time of the time-vary fission filter is 20∼30 seconds per time step. Here at each time step, four images (whose size is 1920×1080) are processed. For the event detection and evaluation, the processing time is very fast whereas there are large delay since the features must be extracted after this round finished. For this task, there is little space to improve the time performance. Therefore, to achieve real-time implementation of the tracker initialization is the further topic. We have implemented real-time multiple-player tracking on GPU, whose processing time is 14.43∼16.01 milliseconds per time step. Based on this, the proposed algorithm has a large potential to achieve the real-time system.

Analysis of the generalization performance

Besides the automatic Data Volley, the data acquisition methods proposed in this article can also be generalized to other sports or tasks. Firstly, the automatic tracker initialization method can be used for players tracking of any sports. In addition, by modifying the image feature for the target, the time-vary fission filter can be generalized to any other multiple objects tracking. The only constraint condition is the amount of tracking objects must be fixed. Secondly, the idea of team formation mapping can be used for event detection of the sports whose event definition are related to the team formation.

Conclusion

To achieve the automatic data volley system, this paper combines the spatial and temporal features, proposes the

time-vary fission filter based tracker initialization, the team formation mapping with sequence motion based event detection, and the relative spatial filter based quality evaluation. Firstly, by approximating the temporal variation of the team (and players belonging to which) distribution, the tracker can be automatic initialized without using absolute player positions. Secondly, the team formation mapping with the sequential motion represents the event feature using both the temporal event relation feature and the spatial feature of the team organization. At last, using of the relationship of event types and the volleyball game process, the event quality are evaluated referring relative position relationships. The experimental results show there are 94.1% rounds successfully initialized, the event type detection result achieves the accuracy of 98.72%, and the success rate of obtaining events' quality achieves 97.27% in average.

Our future target will dive into two aspects. First, combining current work with the game strategy knowledge and rules, a comprehensive automatic Data Volley system consisting of game data acquisition and tactics development is our future topic. Furthermore, based on the achievements of the GPU real-time acceleration [45,49], an real-time and low-delay automatic data volley analysis system is expected for the supporting of TV content broadcasting in international big events such as Olympic Games.

Acknowledgements This work was jointly supported by the National Natural Science Foundation of China 62006178 and Waseda University Grant for Special Research Projects (2019Q-055).

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Thai My T, Weili Wu, Xiong Hui (eds) (2016) Big Data in Complex and Social Networks. CRC Press, Boca Raton
2. Gadekallu RT, Gao Xiao-Z (2021) An efficient attribute reduction and fuzzy logic classifier for heart disease and diabetes prediction. Recent Adv Comput Sci Commun 14(1). https://doi.org/10.2174/2213275911666181030124333
3. Pouli V, et al. (2015) Personalized multimedia content retrieval through relevance feedback techniques for enhanced user experience. In: 2015 13th International Conference on Telecommunications (ConTEL). IEEE
4. Almujahed S, et al. (2013) Sports analytics: designing a volleyball game analysis decision-support tool using big data. In: 2013 IEEE Systems and Information Engineering Design Symposium. IEEE
5. DataVolley. https://www.dataproject.com/Products/EN-/en/Volleyball/DataVolley4
6. Glossary of Volleyball Lingo, Slang & Terms. https://www.sportslingo.com/volleyball-lingo-glossary
7. Javed A, Bajwa K, Malik H, Irtaza A (2022) An efficient framework for automatic highlights generation from sports videos. IEEE Signal Process Lett. https://doi.org/10.1109/LSP.2016.2573042
8. Xiong J, Lu L, Wang H, Yang J, Gui G (2019) Object-level trajectories based fine-grained action recognition in visual iot applications. IEEE Access 7:103629–103638
9. Theagarajan R, Bhanu B (2021) An automated system for generating tactical performance statistics for individual soccer players from videos. IEEE Trans Circ Syst Video Technol 31(2):632–646. https://doi.org/10.1109/TCSVT.2020.2982580
10. Felsen P, Agrawal P, Malik J (2017) What will happen next? Forecasting player moves in sports videos. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, pp. 3362–3371
11. Suzuki G, Takahashi S, Ogawa T, Haseyama M (2019) Team tactics estimation in soccer videos based on a deep extreme learning machine and characteristics of the tactics. IEEE Access 7:153238–153248
12. Yoon Y et al (2019) Analyzing basketball movements and pass relationships using realtime object tracking techniques based on deep learning. IEEE Access 7:56564–56576
13. Fani M, Yazdi M, Clausi DA, Wong A (2017) Soccer video structure analysis by parallel feature fusion network and hidden-to-observable transferring markov model. IEEE Access 5:27322–27336
14. Sing LT, Paramesran R (2011) Detection of service activity in a badminton game. In: TENCON 2011 - 2011 IEEE Region 10 Conference, Bali, pp. 312–315
15. Dardagan N, BrĐanin A, Džigal D, Akagic A (2021) Multiple object trackers in OpenCV: a benchmark. In: 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), pp. 1–6. https://doi.org/10.1109/ISIE45552.2021.9576367
16. Sun S, Akhtar N, Song H, Mian A, Shah H (2021) Deep affinity network for multiple object tracking. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 104-119. https://doi.org/10.1109/TPAMI.2019.2929520
17. Qian Y, Shi H, Tian H, Yang R, Duan Y (2020) Multiple object tracking for similar, monotonic targets. In: 2020 10th Institute of Electrical and Electronics Engineers International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 360-363. https://doi.org/10.1109/CYBER50695.2020.9279162
18. Xiao Z, Xu X, Xing H et al (2021) RTFN: a robust temporal feature network for time series classification. Inform Sci 571:65–86. https://doi.org/10.1016/j.ins.2021.04.053
19. Beetz M, von Hoyningen-Huene N, Kirchlechner B, Gedikli S, Siles F, Durus M, Lames M (2009) Aspogamo: automated sports game analysis models. Int J Comput Sci Sport 8(1):1–21
20. Sheng B, Li P, Zhang Y, Mao L, Chen CLP (2021) GreenSea: visual soccer analysis using broad learning system. IEEE Transa Cybern 51(3):1463–1477. https://doi.org/10.1109/TCYB.2020.2988792
21. Yamamoto T, Kataoka H, Hayashi M, Aoki Y, Oshima K, Tanabiki M (2013) Multiple players tracking and identification using group detection and player number recognition in sports video. In: IECON

2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, Vienna, pp. 2442–2446
22. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp 779–788
23. Ikoma N, Hasegawa H, Haraguchi Y (2013) Multi-target tracking in video by SMC-PHD filter with elimination of other targets and state dependent multi-modal likelihoods. In: Proceedings of the 16th International Conference on Information Fusion, Istanbul, pp. 588-595
24. Huang S, Zhuang X, Ikoma N, Honda M, Ikenaga T (2016) Particle filter with least square fitting prediction and spatial relationship based multi-view elimination for 3D Volleyball players tracking. In: IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA). Malacca City 2016:28–31
25. Feng J, Pu S, Zhao K, Zhang H, Du T (2019) Enhanced initialization with multi-stage learning for robust visual tracking. IEEE Vis Commun Image Process (VCIP) 2019:1–4. https://doi.org/10.1109/VCIP47243.2019.8966006
26. Liu Y, Zhang L, Chen Z, Yan Y, Wang H (2021) Multi-Stream siamese and faster region-based neural network for real-time object tracking. IEEE Trans Intell Transp Syst 22(11):7279–7292. https://doi.org/10.1109/TITS.2020.3006927
27. Maksai A, Wang X, Fua P (2015) What players do with the ball: A physically constrained interaction modeling. arXiv preprint arXiv:1511.06181
28. Chakraborty B, Meher S (2013) A real-time trajectory-based ball detection-and-tracking framework for basketball video. J Opt 42(2):156–170
29. Yan F, Christmas W, Kittler J (2008) Layered data association using graph-theoretic formulation with application to tennis ball tracking in monocular sequences. IEEE Trans Pattern Anal Mach Intell 30(10):1814–1830
30. Zhou X, Xie L, Huang Q, Cox SJ, Zhang Y (2015) Tennis ball tracking using a two-layered data association approach. IEEE Trans Multimed 17(2):145–156
31. Chen H, Tsai W, Lee S, Yu J (2012) Ball tracking and 3D trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. Multimed Tools Appl 60(3):641–667
32. Xina C, Norikazu , Masaaki H, Takeshi I (2017) Multi-View 3d ball tracking with abrupt motion adaptive system model, anti-occlusion observation and spatial density based recovery in sports analysis. IEICE Trans Fund E100-A(5): 1215-1225
33. Takahashi M, Ikeya K, Kano M, Ookubo H, Mishina T (2016) Robust volleyball tracking system using multi-view cameras. In: 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 2740–2745
34. Liu, H-Y, Tingting H, Hui Z (2007) Event detection in sports video based on multiple feature fusion. In: Fuzzy Systems and Knowledge Discovery (FSKD 2007), vol.2, pp.446-450. IEEE
35. Jinjun W, Changsheng X, Engsiong C, Xinguo Y, Qi T (2004) Event detection based on non-broadcast sports video. Image Processing, 2014. In: ICPI'04. 2004 International Conference, vol. 3, pp. 1637-1640, IEEE
36. M. K.M., et-al "VGRAPH: An Effective Approach for Generating Static Video Summaries", in IEEE International Conference on Computer Vision Workshops, IEEE, 2013
37. Asadi E, Charkari NM (2012) Video summarization using fuzzy cmeans clustering. In: 20th Iranian Conference on Electrical Engineering, (ICEE2012). IEEE
38. Ajmal M, Muhammad A, et-al. (2012) Video Summarization: Techniques and Classification. In: Computer Vision and Graphics. Springer Verlag
39. Yang X, Yang X, Liu M, Xiao F, Davis LS, Kautz J (2019) STEP: spatio-temporal progressive learning for video action detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 264–272
40. Guo H, Wu X, Li N (2018) Action extraction in continuous unconstrained video for cloud-based intelligent service robot. IEEE Access 6:33460–33471
41. Xina C, Norikazu I, Masaaki H, Takeshi I (2017) Ball state based parallel ball tracking and event detection for volleyball game analysis. IEICE Trans Fund E100-A(11): 2285–2294
42. Radhakrishnan S, Chittaranjan V, Kavi M (2018) V ScoreA data analytical versatility metric for cricket. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE
43. Zhao R (2012) Research on evaluation index system of media performance in sports events. In: 2012 Fourth International Conference on Computational and Information Sciences. IEEE
44. Xina C, Yang L, Takeshi I (2019) 3D global and multi-view local features combination based qualitative action recognition for volleyball game analysis. IEICE Trans Fund E102-A(11): 1891-1899
45. Xina C, Yiming Z, Takeshi I (2019) Representative spatial selection and temporal combination for 60fps 3d tracking of twelve volleyball players on gpu. IEICE Trans Fund E102-A(11): 1882-1890
46. Ar D, Johansen AM (2009) A tutorial on particle filtering and smoothing: fifteen years later. Handb of Nonlinear Filter 12:656–704
47. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 25(5):281–288
48. Eom HJ, Schutz RW (1992) Statistical analyses of volleyball team performance. Res Quar Exerc Sport 63(1):11–18
49. Yilin H, Ziwei D, Xina C, Takeshi I (2018) View priority based threads allocation and binary search oriented reweight for gpu accelerated real-time 3d ball tracking. IEICE Trans Inf Syst E101-D(12): 3190-3198