Waseda University Doctoral Dissertation

Automatic Data Acquisition Based on Abrupt Motion Feature and Spatial Importance for 3D Volleyball Analysis

Xina CHENG

Graduate School of Information, Production and Systems

Waseda University

June. 2018

Acknowledgements

I would like to express my gratitude to all those who helped me during my Ph. D. period.

First of all, my deepest gratitude goes first and foremost to my supervisor, Professor Takeshi Ikenaga, for all his effort to the course of my research. During these five years when I stay in Ikenaga Laboratory, Professor Ikenaga always gives me advices on my research. Not only on the research itself, Professor Ikenaga also supports me a lot mentally. Many times when I faced difficulties on the research and tried to avoid or give up, Professor Ikenaga always encouraged and supported me to facing the problem. This makes me become strong after this five years.

Second, I also owe a debt of gratitude to Professor Takahiro Watanabe and Professor Kiyoto Takahata who have guided me in improving and completing this dissertation by providing much useful and remarkable advice. I also would thank all other Professors in the integrated systems field of IPS including Professor Takaaki Baba, Professor Shinji Kimura, Professor Hirofumi Shinohara, Professor Noriyoshi Yamauchi, Professor Toshihiko Yoshimasu, Professor Takeshi Yoshimura and Professor Takashi Ohsawa. There are a lot thanks for their attendance of my dissertation presentation and providing valuable comments in their busy schedule.

I also would like to express my gratitude to Professor Masaaki Honda for his kindly supporting of my research. I appreciate Professor Honda a lot for taking time and giving advices on my Ph. D. thesis. With his help on taking sequence and providing background knowledge, my research can go on smoothly. I also owe a special debt of gratitude to Prof. Ikoma in Nippon Institute of Technology. He shows me a wide and deep possibility of the particle filter and also gives me many suggestions on my research.

To all the members of Ikenaga Laboratory, who support me in both the research and daily life, I would like to express my appreciation. I would like to thank Dr. Gaoxing Chen, Mr. Zhenyu Pei, Mr. Chenming Ding and Mr. Xu Zhao, who helped me a lot at the beginning of my research and taught me a lot. I also want to thank Mr. Xizhou Zhuang for the Master period we studied and made prograss together. I would also show my thanks to current lab members including: Ms. Tingting Hu, Mr. Songlin Du, Mr. Zijie Wang, Mr. Yiming Zhao, Mr. Yang Liu, Mr. Guixing Liang, Mr. Yuchen Yang, Ms. Haoxuan Tan and Mr. Yuhao Xu. Especially Ms. Tingting Hu, she discussed with me a lot and gave me many useful advises during the preparation of this dissertation.

At last, I am also deeply indebted to my beloved parents. Although we live in different counties, they are the very ones being most close to me in my heart and always support me, encourage me, conform me and respect my decisions.

Abstract

Sports analysis technologies have attracted increasing attention with the hosting of big sports events in Japan, such as the 2020 Olympics. In particular, vision sensor based sports data analysis technology is expected to provide copious game data for the realization of professional strategy analysis systems and sports broadcasting systems. The TV broadcasting system for swimming and the soccer strategy analysis system have been developed and put into practical use. However, these technologies work for the data acquisition of 1D and 2D movement. For expanding them to a wider range of sports such as volleyball, pingpong and tennis, the 3D sports analysis is expected. Particularly, the volleyball requires 3D data for 3D movements of both the ball and the player.

In order to realize the 3D sports analysis system for volleyball, the technologies that acquire the game data (the position and velocity of the ball and players, the motion of players and the data representing the play efficacy, etc.) and their real-time implementation are required. Since the movement of the ball is deeply involved in every situation during the play, the ball data holds the key of further data acquisition and analysis. The ball related data include the physical, event and strategy data. First, the physical data represent the 3D position and velocity of the ball. Second, the event data represent what kind

of play has been done, such as receive, toss and spike. Finally, the strategy data are values deciding the play efficacy, such as the number of available attackers and blockers, the attack tempo, the attack type, and the set zone. The physical data can be obtained by the 3D tracking technology, which suffers from the abrupt ball motion, the ball moving out of the frame and the occlusion. The event data can be obtained from the ball trajectory. However, noise in the trajectory affects the detection accuracy and the event change is difficult to predict because of the unknown time of the player's action that changes the event. To obtain the strategy data, the role of each player (attacker, blocker, setter) is required, however it cannot be automatically obtained only from the physical data and event data without manually labeling.

To solve above problems and realize a 3D sports analysis system for volleyball that enables automatic acquisition of ball related data (physical data, event data, and strategy data), the abrupt motion feature and spatial importance are proposed in this dissertation. The abrupt motion feature is defined by the velocity change in the abrupt situation and the general situation of the ball and player. By combining the abrupt motion feature in different situations or using it as the likelihood, it is possible to acquire several kinds of data accurately. In addition, the spatial importance indicates the probability distribution of the ball candidates, the event change positions and the locations of each player role. Based on these, target state of tracking and detection can be estimated accurately so that a higher performance of data acquisition can be achieved.

This dissertation is organized as follows.

Chapter 1 explains the significance of 3D sports analysis technology and the necessity of automatic data acquisition based on vision sensors. Then, the points of focus and the objectives of this dissertation are described.

In chapter 2, to acquire the physical data of the ball such as 3D position and velocity from the multi-view game data, the abrupt motion adaptive system model and the ball candidate distribution based recovery are proposed. In Deguchi's work [MVA, 2002], the system model is proposed assuming that the ball motion is free fall and bouncing. However, it is unsuitable for volleyball, in which the ball direction changes due to external factors by players. The proposed system model combines two different system noises that cover the motion of the ball both in general situation and abrupt situation. Combination ratio of these two noises and number of particles are adaptive to the estimated motion probability. Then, the ball candidate distribution is used to recover the tracking lost caused by the ball moving out of the frame. The anti-occlusion observation model solves the occlusion problem by eliminating influence of the camera with less confidence. In this dissertation (including all chapters), the experimental HDTV video sequences are the 2014 Inter High School Men's Volleyball Games held in Japan, which are captured by four cameras located at each corner of the court. Here, the experimental results show that the success rate achieved by the proposals of 3D ball tracking is 99.4% and the average position precision error is 5.32(cm).

In chapter 3, to acquire the event data based on the ball trajectory and game videos, the multiple event change feature based observation and the spatial event change probability distribution based prediction are proposed. As the conventional work of the ball event detection, Almaljai [ICIP, 2010] detects the event change by using entire 2D ball trajectory for tennis game. However, noise in the trajectory influences the detection performance. Besides using the abruptness of trajectory in which noise exists, the proposed observation also uses the area of players' skin as likelihood to evaluate the event change. When players' hands are close to the ball, the ball is more likely to be hit. So the area of players' skin in the region around the ball is calculated as the likelihood of event change. The probability distribution of event change is approximated according to the fact that the probabilities of event change in different spatial locations are different. For example, the probability becomes large in the spike and block areas. Thus event changes are accurately predicted. Experimental results show the detection rate of ball event achieves 92.4%.

In chapter 4, to automatically acquire the strategy data showing the attack efficacy from the game videos, the relative motion abruptness and relative court zone division based player role detection is proposed. For the strategy data analysis, the Data Volley system is widely used. However, its input is performed manually by visual observation, which is inaccurate for human mistake and requires much manually labeling. Therefore, there is a demand for a technology that automatically and accurately acquires the strategy data. All the strategy data are defined based on player role, which cannot be obtained directly from the physical and event data because of consisting strategy information. In the proposed method for player role detection, the relative motion abruptness, which is calculated from the ball motion and player motion, represents the effort of each player to the play. This feature is used to filter out the candidate role for each player. By dividing the court zone into setter, attacker and blocker zones based on the distance between the ball and the player, a court zone based filter is modeled to further detect the player role. Based on the detected role of each player and their physical data and event data, the strategy data indicating the attack efficacy variables can be calculated. The detection rate of the set zone, the number of available attackers, the attack tempo and the number of blockers are 100%, 100%, 97.8%, and 100%. Compared with the detection rate of manual acquisition, it achieves 8.3% average improvement.

In chapter 5, the overall dissertation is summarized and the future works are described. To realize a 3D volleyball analysis system, algorithms for the automatic acquisition of the physical data, event data and strategy data are proposed. The motion features in abrupt and general situation are combined in system model and are used as likelihood in observation. The spatial probability distribution indicates or ball and players' exiting area. They are used to achieve an accuracy of 92∼100% for the game data acquisition, which greatly contributes to the academic. These results are highly evaluated in various practical applications, such as the game strategy system and the TV sports broadcasting system.

Contents

CONTENTS

List of Figures

List of Tables

LIST OF TABLES

Chapter 1

Introduction

1.1 Importance of Sports Analysis

The sports analysis technologies have attacked lots of attentions as big sports events going to be held in Tokyo in the near future, such as the 2020 Olympic games and 2019 Rugby World Cup. With rapidly development of computer vision technologies and widely utilization of high-quality videos productions, the vision cameras based sports analysis becomes key topic for supporting of sports industry.

By providing accurate and copious game data, the sports analysis mainly contributes in two aspects, the TV broadcasting system and game strategy analysis systems.

For the TV broadcasting system, the sports analysis technologies devote to the TV content of the game. With various kinds of game data extracted automatically from the vision sensor, the game contents and the player performance even the mental state are expected presented visually to the audience. The game contents include not only the game scoring but also the precise data that cannot be obtained manually such as the height of the jump and the speed of the ball. The player performance and mental state are represented by the quality of the action and the pulse rate. Based on these data, the implementation

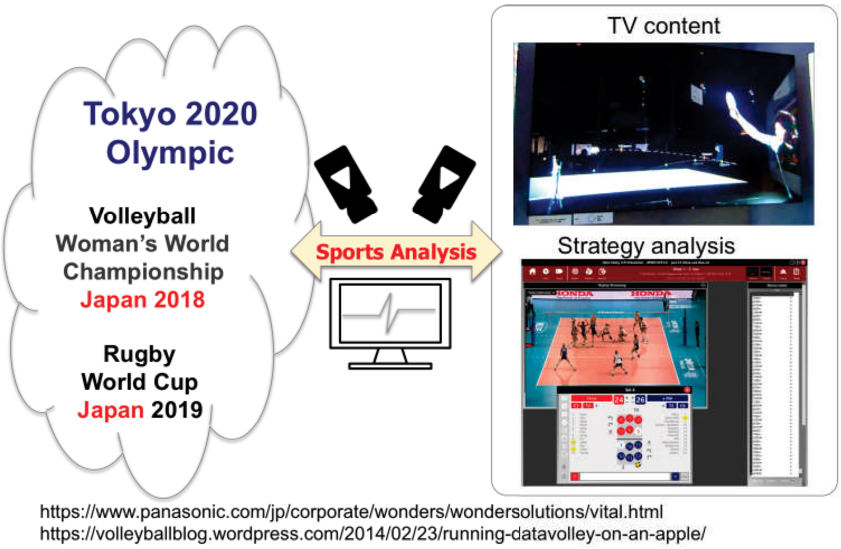


Figure 1.1: Importance of the vision camera based sports analysis.

of an attractive TV sports contents is expected to help the audience understand the game status deeply and comprehensively.

For the strategy analysis systems, the vision camera based technologies have great potential in providing accurate game data and saving human labor. The strategy analysis systems cover various aspects, such as the game data statistic, the tactics development and the coaching supporting. The effective strategy analysis systems play important role in the improvements of the team strength and the player skill by providing the tactics of teamwork, presenting the training progress and providing reasonable training plan. The performance and reliability of the strategy analysis systems are influences by the accuracy and amount of the game data acquired from the game.

1.2 Importance of 3D Sports Analysis

Currently, there have been some sports analysis system products being utilized in the sports event. For performing the swimming competition attractively, a video based broadcasting system shows the speed of each player in the TV image. For the development of the tactics in the soccer, the game videos based strategy analysis system evaluates the team performance from the players' positions and motions.

However, these products only work for acquisition of the 1D and 2D motion. In order to expand the application objects to a wide range of sports that require 3D data acquisition, the 3D sports analysis system is expected. The objects of 3D sports analysis include the volleyball, the pingpong and the tennis.

(a) Two-dimensional analysis (b) One-dimensional analysis

https://www.panasonic.com/jp/corporate/wonders/wondersolutions/vital.html

Figure 1.2: Examples of 1D and 2D sports analysis products.

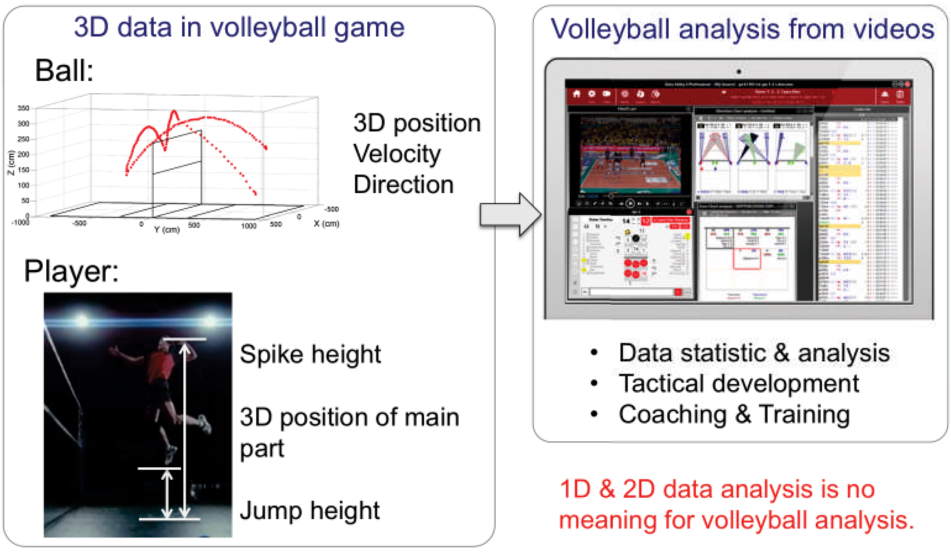
Particularly, in the volleyball game, motion of both the player and the ball are described in 3D concept. Firstly, the ball in volleyball game always flies in the air. To describe the ball information, especially the ball height, 3D motion is required. Furthermore, the players not only move left-to-right or front-to-back on the ground but also jump to play the ball. Such as in the serve, spike and block situation, since the ball flies in a

(a) Volleyball (b) Ping-Pong

https://www.panasonic.com/jp/corporate/wonders/wondersolutions/vital.html

Figure 1.3: For strategy analysis (e.g. volleyball, Ping-Pong and tennis) 3D data is necessary.

large height, the player should jump high to play the ball.



http://www.dataproject.com/Products/EN/en/Volleyball/DataVolley4

Figure 1.4: Target of 3D sports : volleyball.

1.3 Technologies in 3D Volleyball Analysis

In order to realize the 3D sports analysis system for volleyball, the technologies of game data acquisition is required. From the game data category point of view, the data acquisition technologies include three levels. First level is the acquisition of the position and speed of the ball and players, such as the 3D ball tracking and multiple players tracking technologies. Second level is the acquisition of the event and player motions such as the event detection based on the ball trajectory and the player action detection. The last level is the acquisition of the strategy data representing the play efficacy automatically. Furthermore, with the target of TV broadcasting system, the real-time implementation technology and CG display technology are also required. Combining all the 3D sports technologies mentioned above, various kinds of game data can be obtained accurately, automatically and timely from game videos. These 3D sports technologies not only can be used for volleyball game, but also has high potential to be extended to other ball sports such as the pingpong and the tennis game.

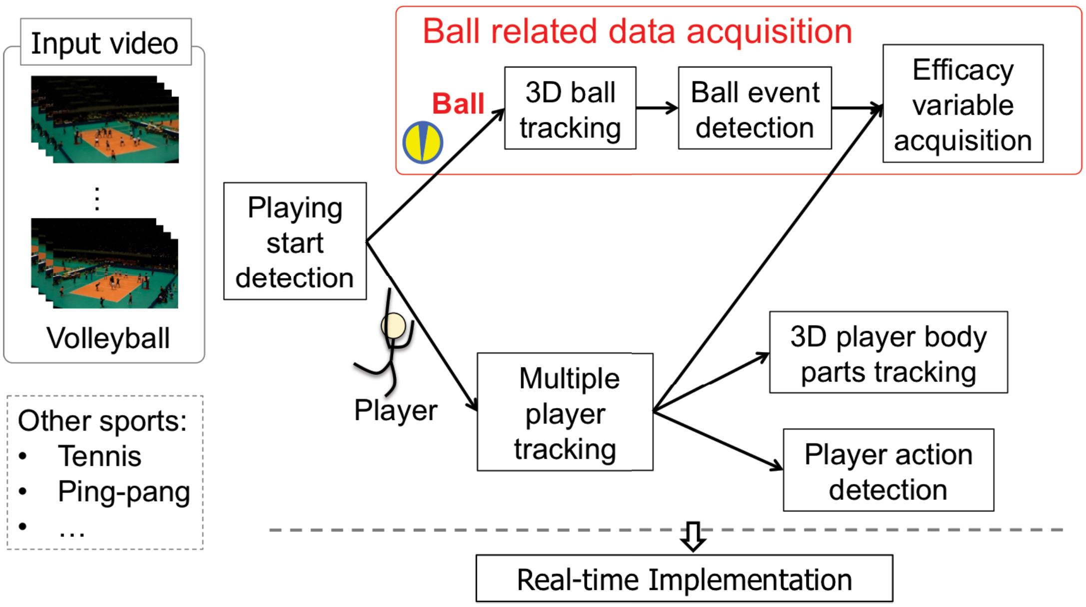


Figure 1.5: Technologies of 3D volleyball analysis.

Among all the 3D sports technologies of data acquisition, the acquisitions of the ball related data play the most important role and the ball data holds the key of further data acquisition and analysis. It is because that the movement of the ball is deeply involved in every situation during the play. The ball related data includes the physical data, the

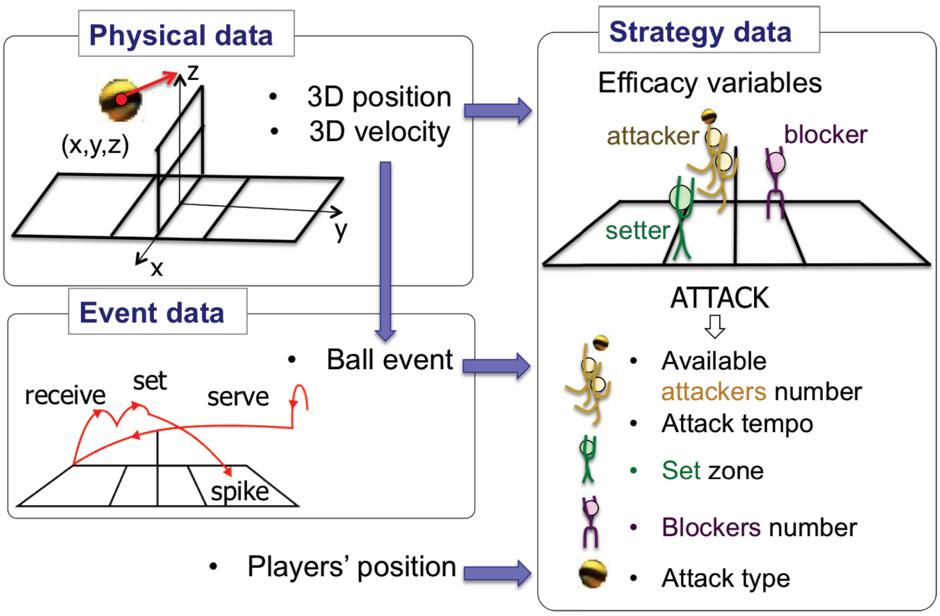


Figure 1.6: Ball related data acquisition.

event data and the strategy data. First, the physical data represents the 3D position and velocity of the ball. Second, the event data represents what kind of play has been done, such as receive, set and attack. Finally, the strategy data shows the play efficacy of the attack, such as the number of available attackers and blockers, the attack tempo, the attack type, and the setting zone. Therefore, acquisitions of these three kinds of ball related data accurately and automatically are key topics in 3D sports analysis.

1.4 Problems in Ball Related Data Acquisition

To achieve high performance for the acquisition of the ball related data, it is necessary to solve the problems in the 3D ball tracking for physical data acquisition, the ball trajectory based event detection for event data acquisition and the attack efficacy variables detection for strategy data acquisition.

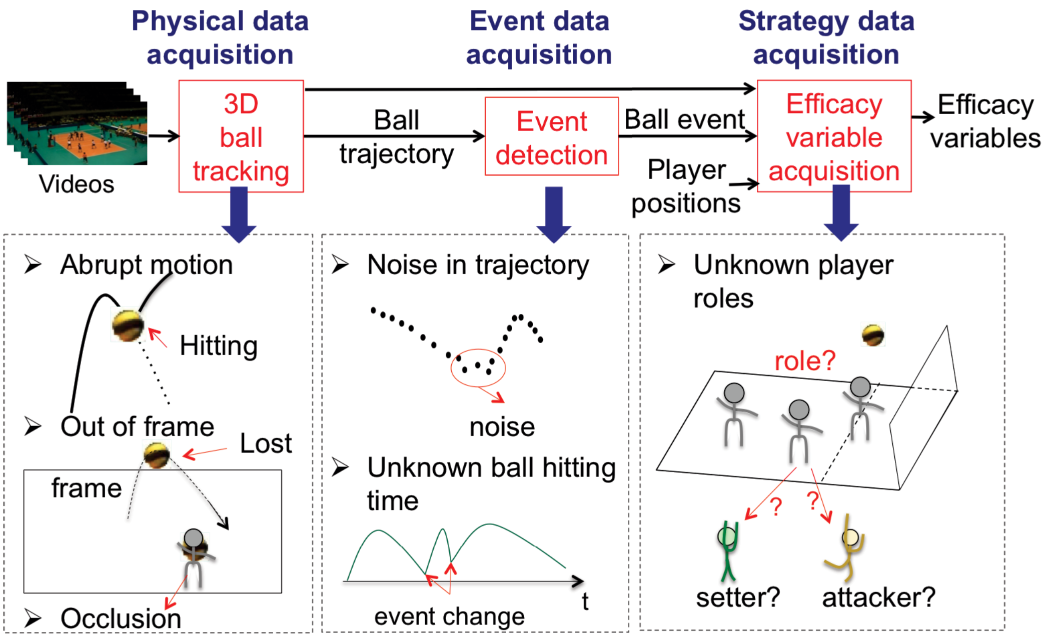


Figure 1.7: Problems in the automatic ball related data acquisition.

In the 3D ball tracking from videos for the physical data acquisition, the difficulties include the ball's small size, high speed and frequently occlusion by players [3] [4]. Part of these can be solved by developed high-resolution videos, but due to the special environment of real game and limitations of the shooting conditions in volleyball game, there are still three major problems remained causing the tracking lost: the abrupt ball motion, the ball moving out of the frame and the occlusion. Firstly, the abrupt motion of the ball is difficult to predict because of unknown external force. In volleyball games, there are copious events of abrupt motion including receive, set, attack and block. When these motions occur, speed and direction of the ball change abruptly and drastically by the hitting force of players. Different from ground bouncing, motions after hitting by players are unpredictable because of the unknown force. Secondly, the ball moves out of the frame due to the limitation of the camera position. In this situation, there is no target in the frame and the tracking fails. Since the time when the ball re-appears in the frame is unknown, the tracker is difficult to recovery the tracking. Thirdly, the ball is frequently occluded by players and other objects in game videos that make the tracking fail easily.

In the event detection, which is based on the ball trajectory and video for the event data acquisition, there are three factors affects the acquisition performance: the noise in the unstable ball trajectory, the unknown time of the player action that changes the event and the large intra-class difference of the same event type. Firstly, unstable and rough ball trajectories are bad for feature extraction of ball event. Not only the positions, but the speed and motion directions also reflect important features of a ball event as well. Especially, when the ball is hit by player, the velocity changes dramatically before the position trajectory turns. Without the integrate as well as accurate trajectory, the event change can only be estimated by the velocity variation since the turn of the trajectory is not obvious through local short trajectory segment. However, from real game videos the accurate 3D ball trajectory is difficult to track due to the complex background as well as the motion of the ball. The small flicker of position trajectory corresponding to large difference in velocity affects the reliability of event feature extraction from velocity information. Secondly, the event change when the player hits the ball. In volleyball game, without referring the information of players, the time when the player hits the ball is unknown and different event continues different time length in different situations. Therefore, it is difficult to predict the event change temporally based on the ball trajectory. Thirdly, the intra-class differences of some ball event are large, which lead to difficulties in event feature extraction. In volleyball game, the locating position and the continuing time of one event change a lot in different situations. Such as one pass (receive or set) event may occur in anywhere even the outside of the court. And one attack may continue until: 1) being blocked around the upside of the net; 2) being received after crossing the net; 3) hitting the ground. The continuing time of events under the above three conditions are different although all of them belong to ”attack” event.

In the attack efficacy variables detection for the automatic strategy data acquisition from the game videos, the unknown role information is the key factor that makes the automatic strategy data acquisition difficult. Without the manually labelling the role for each player (such as the attacker, blocker, setter), the role information cannot be obtained automatically from the physical data and event data of the ball and players. To acquire the attack efficacy variables, not only the player who attacks the ball, but also the setter, other available attacker, and the blockers need to be identified. In the volleyball game, even the physical and event data of the team is obtained, it is still difficult to analysis the contribution of each player to the play without any information of player roles.

1.5 Proposed Solutions

In the 3D ball tracking, the performance of the physical data acquisition suffers from the abrupt ball motion, the ball moving out of the frame and the occlusion. These cause

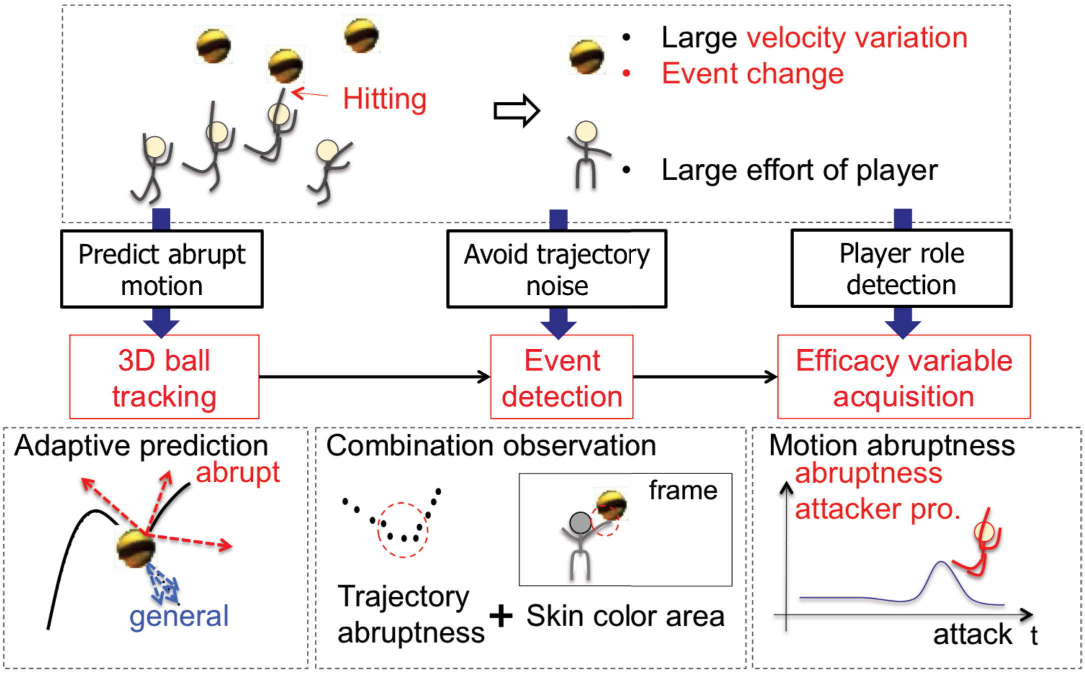


Figure 1.8: Proposed concepts: abrupt motion feature.

tracking lost. The proposed 3D tracking algorithm uses the abrupt motion feature defining by the velocity change. In the abrupt situation (like the ball being hit by players and the player rushing), the velocity change is large. In the general situation (the ball flying in the air and the player being standby), the velocity change is small. By combining two different system noises that cover the motions of the ball both in general situation and abrupt situation, the problem of the abrupt ball motion can be solved. Furthermore, based on the spatial distribution of existing probability for the ball candidates, the tracking can be recovered from the ball moving out of the frame by re-initializing the tracking at the area where the ball candidate's existing probability is large. In addition, referring the multi-view image information, the occlusion problem can be solved.

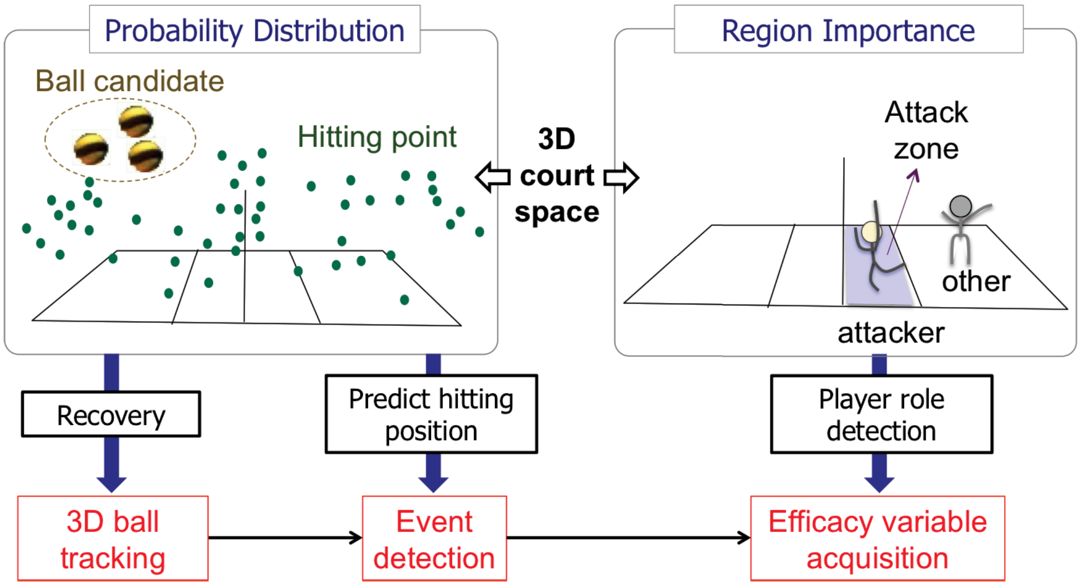


Figure 1.9: Proposed concepts: spatial importance.

In the event detection based on the ball trajectory, which are affected by the noise in the unstable trajectory and the event change that is difficult to predict. To deal with the noise in the unstable trajectory, the abrupt motion feature (the velocity of the ball changes a lot when being hit by the player) is used. Besides the trajectory abruptness that reflects the velocity change, the player skin area feature (that represents the probability of the ball being hit) based on the videos is also used in the observation of the event change so that the unidirectional dependency on the ball trajectory is avoided. At last, the feature of the distance between the ball and the specific court line are extracted to detect the event type.

In the attack efficacy variables detection for the strategy data acquisition based on the game videos, the main difficulty is the unknown player roles without the manual labeling. The proposed method overcomes this difficulty by using the relative motion abruptness (that is calculated from the ball motion and player motion) and the important court zone of each player role to filter out the role of each player. In addition, by summing up the ball relative motion density of all the players in specific court area, the influence of individual player information can be reduced for the team status detection.

1.6 Chapter Organization

This dissertation is targets on the automatic acquisition of the ball related data in volleyball for 3D sports analysis. Each chapter is organized below.

Chapter 1 explains the significance of 3D sports analysis technology and the necessity of automatic data acquisition based on vision sensors. To acquire the physical data, event data and strategy data from the game videos, the multi-view 3D ball tracking, the event detection based on the ball trajectory and videos, and the attack efficacy variables detection methods are proposed in chapter 2, 3 and 4, respectively. This paper consists of two full papers. Chapter 2 consists of one paper [5] whose content is the 3D ball tracking for physical data acquisition. Chapter 3 consists of one paper [6] targeting on ball event detection for event data acquisition.

Chapter 2 describes the proposed 3D ball tracking algorithm with abrupt motion adaptive system model and the ball candidate distribution based recovery. The proposed system model combines two different system noises that cover the motions of the ball both in general situation and abrupt situation. This system model solves the tracking lost from abrupt ball motion. Furthermore, the ball candidate distribution is used to recover the tracking from lost caused by the ball moving out of the frame.

Chapter 3 presents ball event detection with the multiple event change feature based observation and the spatial event change probability distribution based prediction. To solve the problem of noise in unstable trajectory, the proposed observation evaluates event change by using the feature of the trajectory abruptness and the area of players' skin as likelihood. The probability distribution of event change position is used for predicting event change accurately.

Chapter 4 proposes the relative motion abruptness and relative court zone division based player role detection to automatically acquire the strategy data. To acquire the player roles that is difficult to obtain automatically, the relative motion abruptness, which is calculated from the ball motion and player motion, and the court zone division by ball relative distance are proposed as filters of player roles.

In chapter 5, the overall dissertation is summarized and the future works are described. The proposed algorithms acquire the physical data, event data and strategy data automatically with high precision based on the game videos. For the game strategy analysis, the quality evaluation of play is the remained topic, which is one of future targets of research. And the real-time implementation is the other target for the TV broadcasting system.

1. INTRODUCTION

Chapter 2

Abrupt Motion Adaptive System Model and Ball Candidate Distribution based 3D Ball Tracking

2.1 Introduction

The physical ball data, including the 3D position and velocity, not only represents the ball motions but also is the key of event and strategy data acquisition. As Fig. 2.1 shows, in the acquisition of three kinds ball related data, the physical data acquisition is the first step whose input is the game videos and output is the ball trajectory consisting 3D position and velocity. Since the tracking rate and the accuracy of the 3D position of the ball affect the reliability of game analysis significantly, the goal of this chapter is to perform 3D ball tracking with a high rate of tracking success.

Difficulties in ball tracking from videos include the ball's small size, high speed and frequently occlusion by players [3] [4]. Part of these can be solved by developed high-resolution videos, but there are still three major problems remained due to the special environment of real game and limitations of the shooting conditions in ball sports. 1) Abrupt

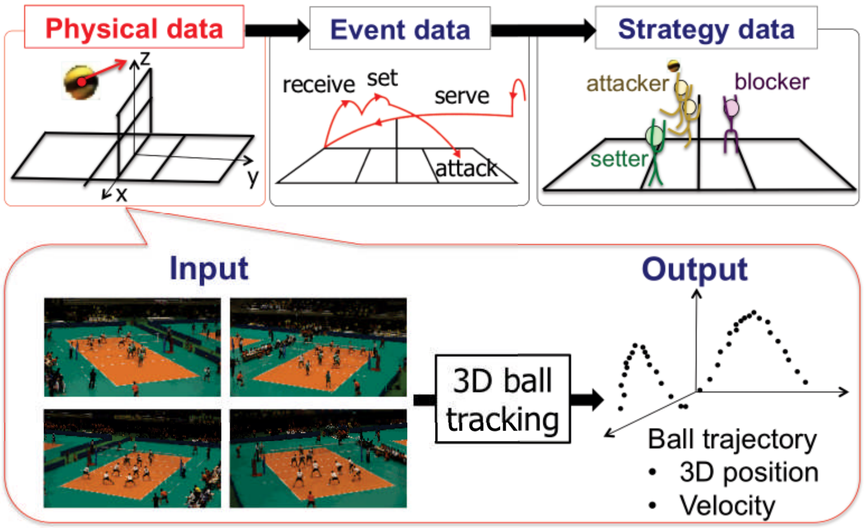


Figure 2.1: Introduction of the ball physical acquisition and whose position of in entire research.

motion of the ball is difficult to predict because of unknown external force. For instance, in volleyball games there are copious events of abrupt motion including spike, toss, receive and block. When these motions occur, velocity and direction of the ball change abruptly and drastically by the hitting force of players. Different from ground bouncing, motions after hitting by players are unpredictable because of the unknown force. 2) The ball is frequently occluded by players and other objects in game videos. Under this condition, the tracking target disappears from the image and the tracker easily fails. To solve these two problems, there are many approaches have been proposed. 3) Because of limitation in video recording, the ball moves out of the frame in some situations. In this way, there is no tracking target in the image so that the tracking failure occurs.

To solve the first problem, Huang [7] employed motion estimation into the state transition of the particle filter, thereby making the process noise variance adaptive to the motion. In Yan's work [8] involving ball tracking, the system equation is switched between two dynamic models based on the distance between the ball and the player. As the positions of the players are required, this method is adversely affected by games in which several players participate. In their subsequent work [9], Yan tracked the trajectory of the ball backward by using the posteriori probability data to accommodate the abrupt motion.

As for the occlusion problem, Scharcanski [10] aimed at solving this problem by switching to a different sampling model according to the detection result of the occlusion event, whereas Guo [11] enlarged the search region based on the algorithm of the particle filter. Similar to most occlusion handling methods [12] [13] [14], however, their methods are unable to track the real position of the ball due to the lack of multi-view information. With two-view videos, Rezaee [15] succeeded in tracking the target independently for each view and used the homographic relation to declare as well as cancel the occlusion in one view, but their work cannot obtain the 3D position of targets.

Besides the three problems mentioned above, there are still other problems in our research, such as the complex background in the game video. In an official game, there are a lot of audiences, staffs, judgers and some players being inside the video scene. These items not only make of a complex background with noise, some objects even share the similar feature with the ball. With the purpose of ball tracking in complex background, my one work [16] proposed the ball feature observation likelihood to distinguish the ball from other items in the background. In addition, the tracker sometimes fails to track the target or the ball moves out of the video frame because of the shooting condition limitation. In these situations, it is almost impossible for the tracker to recover tracking automatically.

There are two related works of 3D ball tracking. First, Chen [17] proposed an automated system for ball tracking and 3D trajectory approximation in volleyball games. A two-phase ball tracking algorithm is used to detect ball candidates and fit a trajectory based on information obtained from a single camera. However, the lack of multiple space information means that his approximated result is unreliable and the tracking accuracy is not high. Second, Deguchi's work [1] tracks reconstructed 3D positions of multiple objects from multi-view images by using mixed-state condensation algorithm (particle filter). The mixed-state model proposed in their work consists of a constant acceleration model and a bouncing model. In the bouncing model, which aims at solving the free bouncing motion of the ball, height component motion of the ball is predicted going to the opposite direction. This mixed-state model is unsuitable for ball motion caused by unknown external force of player's hitting. In addition, they track the ball in simple laboratory with little noise, so their observation model is too simple to be used in real game videos.

2.2 Multi-view 3D Tracking

The precise coordinate of 3D position is difficult to reconstruct from 2D video without in-depth information. Equation (2.1) and equation (2.2) is the projection theory between 2D and 3D space:

$$\Omega \mathbf{s} = \mathbb{A}\mathbf{S}\,, \tag{2.1}$$

$$\mathbf{s} = (\mu, \nu, 1)^T\,, \mathbf{S} = (x, y, z, 1)^T\ , \tag{2.2}$$

where, $\mathbf{s}$ and $\mathbf{S}$ are the homogeneous coordinate of the points in image space (μ, ν) and in real world (x, y, z) respectively; Ω is any nonzero constant; and matrix $\mathbb{A}$ is a 3×4 matrix derived by camera calibration [18]. This equation shows that it is impossible to directly reconstruct a 3D coordinate from a single 2D coordinate, whereas a 2D coordinate can be easily projected by one 3D position.

With the purpose of 3D tracking, many framework have been employed. Based on the same camera number, the merit and demerit of each framework are discussed below.

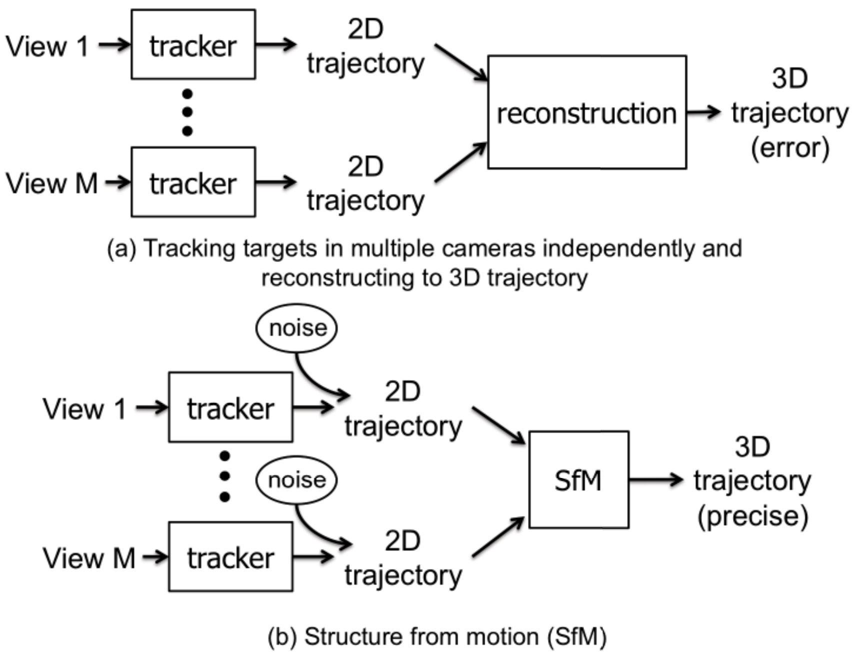


Figure 2.2: Conventional approach for 3D tracking.

The most general framework is tracking the target in multiple cameras independently and reconstructing the 2D trajectories into 3D position as Fig. 2.2 (a) shows. Through this method, Ren's work [19] and Takahashi's work [20] detect the 2D positions of the ball separately in each image plane and reconstructs the 3D coordinate through spatial geometrical relationship. For the tracking process of this framework, the tracking state vector is the 2D coordinate and the observation space is the single view image. Therefore I named this framework as 2D-single-view framework. The 2D-single-view framework uses multi-view information whose performance raises up with the increasing of camera number. Because of the 2D tracking processes in each view are independent with each other, there is no data dependence between views. However, error is easily produced in both 2D tracking process and coordinate reconstruction in this framework. The 2D tracking process with single camera is weak at occlusion and complex background. And the 3D coordinate reconstruction requires considerable compensation for noise distributions as Fig. 2.2 (b) shows [21].

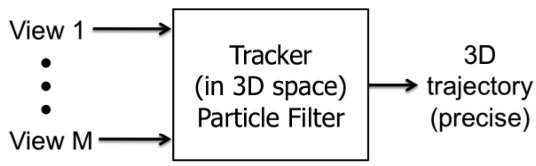


Figure 2.3: Structure of proposed 3D tracking framework.

The structure of the framework [1] shown in Fig. 2.3 avoids errors brought by reconstruction process, therefore our multi-view 3D tracking method is designed following this structure. State vector of this method is defined and initialized in 3D space. Input images are referred in observation step by projecting the state vector from 3D space to image space. So this framework is named as 3D-multi-view framework. Besides the initialization part, there is no reconstruction process. Without coordinate reconstruction, the error of reconstruction never occurs. In addition, by employing this tracking structure with particle filter, the randomly generated particles with a predicted system noise distribution can be compensated for some small errors produced by initialization. The demerit of the framework is that the 3D tracking space requires large number of particles to predict 3D motions comparing with 2D tracking. This paper tracks the 3D position of the ball with multi-view videos following this 3D-multi-view framework.

2.3 Overall Framework of 3D Ball Tracking

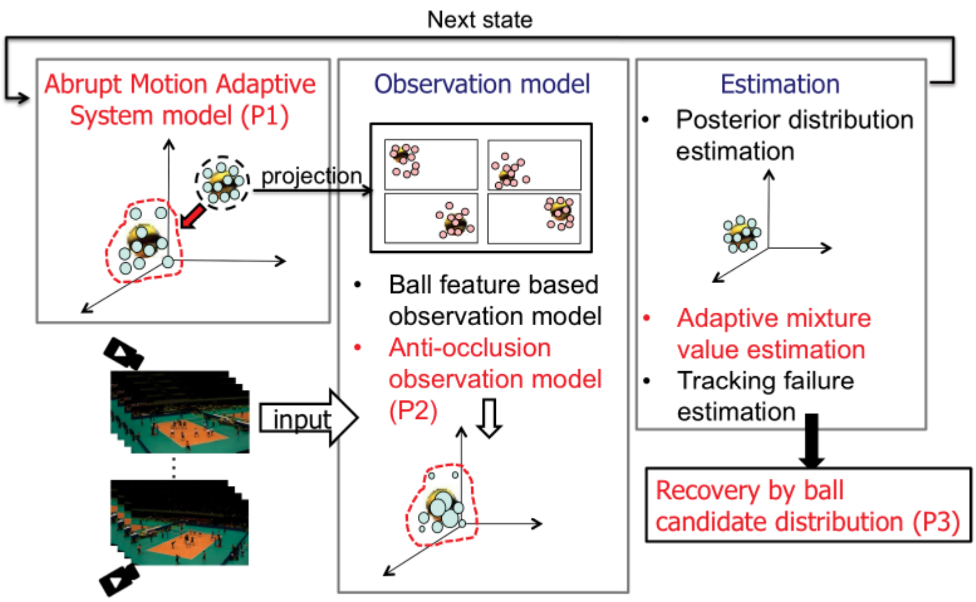


Figure 2.4: Overall structure and proposals of multi-view 3D ball tracking framework.

Fig. 2.4 shows the entire structure of the multi-view 3D ball tracking. The input of the system is a set of synchronization videos from different views, which provide the 3D information of the ball by combining 2D image data of the ball in each view. By equation (2.1) corresponding points are used to calibrate each camera. The output of the tracking is a sequence of 3D ball positions.

This multi-view 3D ball tracking algorithm has three advantages owing to our proposals. Firstly, system model is adaptive to the motion condition of the ball by estimating abrupt motion of ball so that it can be tracked even when being hit by players. Secondly, anti-occlusion observation model removes cameras with low confidence to solve the occlusion problem, which happens when the ball is occluded by players in more than one videos. Thirdly, when tracking fails, spatial ball candidates distribution is proposed to re-initialize the tracker based on the homographic relation between cameras.

2.4 State Space and Initial Distribution

Since the state space is the position of the ball in 3D space, I create a 3D coordinate system in the physical world whose origin is set at the center point of the court. Thus, the state $\mathbb{X}_k$ at discrete time k is defined as:

$$\mathbb{X}_k = \begin{bmatrix} \mathbb{S}_k \\ \mathbb{V}_k \end{bmatrix}, k \in \mathbf{N}, \tag{2.3}$$

$$\mathbb{S}_k = [x_k, y_k, z_k]^T, \mathbb{V}_k = [vx_k, vy_k, vz_k]^T , \tag{2.4}$$

where $\mathbb{S}_k$ represents the position and vector $\mathbb{V}_k$ is the velocity of the center position on the ball, which can also be represented in the polar coordinate system as:

$$[\rho_k, \theta_k, \phi_k], \theta \in [0, \pi], \phi \in [0, 2\pi]. \tag{2.5}$$

In these coordinates, ρ_k is the length of the velocity, θ_k and ϕ_k is the angle from the velocity direction to the z-axis and x-axis, respectively.

Then, initial distribution is defined in equation (2.6) and (2.7):

$$\mathbb{S}_0 \sim N(\overline{\mathbb{S}}, \Sigma_S), \Sigma_S = diag\left(\tau_S^2, \tau_S^2, \tau_S^2\right), \tag{2.6}$$

$$\mathbb{V}_0 \sim N(\overline{\mathbb{V}}, \Sigma_V), \Sigma = diag\left(\tau_V^2, \tau_V^2, \tau_V^2,\right), \tag{2.7}$$

where, $\overline{\mathbb{S}}$ and $\overline{\mathbb{V}}$ are the mean value of the distribution, and Σ_S and Σ_V are the noise variance.

2.5 Proposed Abrupt Motion Adaptive System Model

Time evolution of $\mathbb{X}_k$ in the tracking algorithm is shown in equation (2.8):

$$\begin{pmatrix} \mathbb{S}_k \\ \mathbb{V}_k \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbb{S}_{k-1} \\ \mathbb{V}_{k-1} \end{pmatrix} + \begin{pmatrix} \frac{T^2}{2} \\ T \end{pmatrix} W_k, \tag{2.8}$$

where $\{W_k, k \in N\}$ is is the noise term, which is a mixture of two components as Fig. 2.5 shows. One is the general noise component W_k^G, the other is the abrupt noise component W_k^A. They are integrated with an adaptive ratio α_k:

$$W_k = \alpha_k W_k^G + (1 - \alpha_k) W_k^A \,, \tag{2.9}$$

$$W_k^G = \begin{bmatrix} W_{k,x}^G \\ W_{k,y}^G \\ W_{k,z}^G \end{bmatrix}, W_k^A = \begin{bmatrix} W_{k,x}^A \\ W_{k,y}^A \\ W_{k,z}^A \end{bmatrix}. \tag{2.10}$$

By converging the equation of cartesian coordinates and polar coordinates, noise terms W_k^G and W_k^A can be presented by ω_k^G and ω_k^A in polar coordinate system:

$$W_k^G \Leftrightarrow \omega_k^G, W_k^A \Leftrightarrow \omega_k^A \,, \tag{2.11}$$

$$\omega_k^G = \begin{bmatrix} \omega_{k,\rho}^G \\ \omega_{k,\theta}^G \\ \omega_{k,\phi}^G \end{bmatrix}, \omega_k^A = \begin{bmatrix} \omega_{k,\rho}^A \\ \omega_{k,\theta}^A \\ \omega_{k,\phi}^A \end{bmatrix}. \tag{2.12}$$

In general situation, $\mathbb{X}_k$ is predicted by physical dynamic law taking gravity into consideration. The general system noise ω_k^G is described as

$$\begin{cases} \omega_{k,\rho}^G \sim N(0, \tau_{g,\rho}^2) \\ \omega_{k,\theta}^G \sim \mathfrak{F}(0, \kappa_{g,\theta}) \\ \omega_{k,\phi}^G \sim \mathfrak{F}(0, \kappa_{g,\phi}) \end{cases}, \tag{2.13}$$

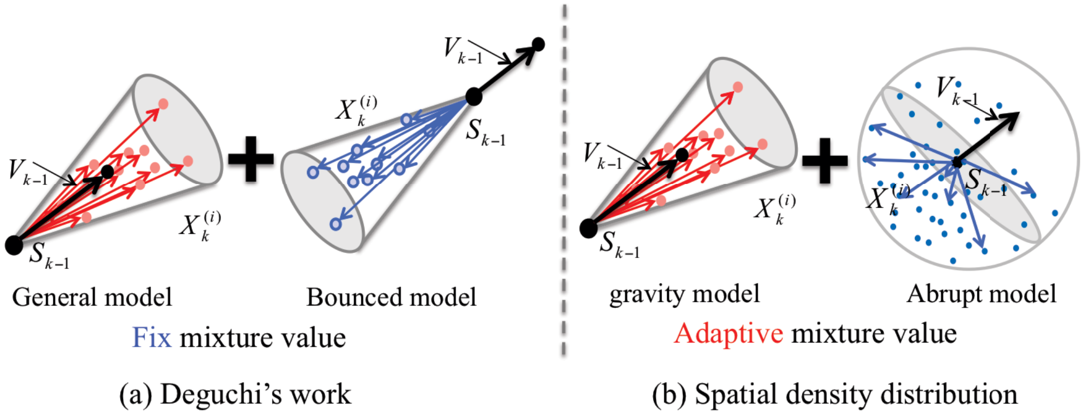


Figure 2.5: Comparison of abrupt mix system models.

where $\mathfrak{F}(\mu_\pi, \kappa)$ is the von Mises distribution [22] which is a continuous probability distribution on circle. μ_π is a measure of location and κ is the measure of concentration. G is the gravity prediction parameter. ω_k^A is designed by taking abrupt motion of the ball into consideration. Since the direction and speed value of the ball changes sharply and abruptly after being hit by unknown external force, the abrupt system noise is designed as

$$\begin{cases} \omega_{k,\rho}^A = A \cdot f(\sigma), \sigma \sim U(0, \pi) \\ \omega_{k,\theta}^A = 0.5 \cdot \omega_f, \omega_f \sim \mathfrak{F}(\pi, \kappa_{a,\theta}) \\ \omega_{k,\phi}^A \sim \mathfrak{F}(\pi, \kappa_{a,\phi}) \end{cases}, \tag{2.14}$$

where A is the maximum value of the changing speed of the ball. And the function of $f(\sigma)$ is

$$f(\sigma) = \begin{cases} sin(\sigma), & 0 \leq \sigma = \frac{1}{2}\pi \\ 1, & \frac{1}{2}\pi \leq \sigma = \frac{5}{6}\pi \\ sin(3\sigma), & \frac{5}{6}\pi \leq \sigma \leq \pi \end{cases}, \tag{2.15}$$

whose curve is represented as Fig. 2.6, which includes three different curves. First, when abrupt motion occurs, velocity of the ball changes a lot, so the value of $f(\sigma)$ is small when

σ is small and it increases with the velocity variance. Then, when σ is large enough, value of $f(\sigma)$ keeps high. At last, since the velocity cannot change without limitation, it falls down quickly as σ approaches to the upper limit.

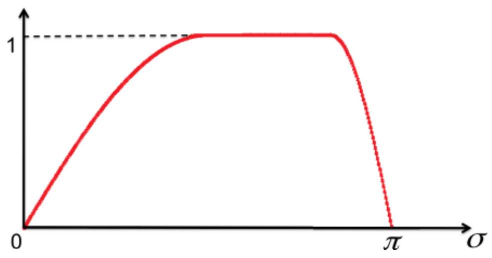


Figure 2.6: Function curve of $f(\sigma)$.

The ratio α_k adapts the system model by adjusting the particle numbers distributed to each component model ($I_k^{general}$ and I_k^{abrupt}) at every discrete time k:

$$\begin{cases} I_k^{general} = \alpha_k \cdot I_k \\ I_k^{abrup} = (1 - \alpha_k) \cdot I_k \end{cases}. \tag{2.16}$$

Determination of α_k and I_k is the key and we will describe them in section 3.4.2.

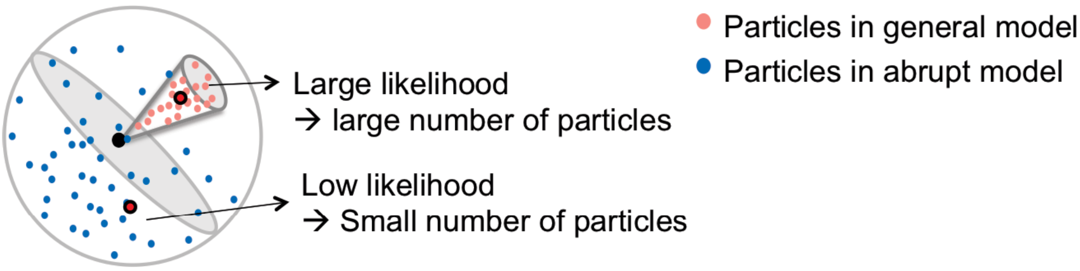


Figure 2.7: Adaptive mixture value estimation in abrupt motion adaptive system model.

2.6 Ball Feature based Anti-occlusion Observation

The observation likelihood is evaluated by image information extracted from different cameras. The observation $\mathbb{I}_k$ are defined as a collection of image frames at discrete time k:

$$\mathbb{I}_k = \left\{\mathbb{I}_k^1, \mathbb{I}_k^2, ..., \mathbb{I}_k^m, ..., \mathbb{I}_k^M\right\} . \tag{2.17}$$

where M is the total number of cameras. Thus, the likelihood $L(\mathbb{X}_k^{(i)})$ of each particle (sampling) is indicated as:

$$L(\mathbb{X}_k^{(i)}) = g\left[L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^1), ..., L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m), ..., L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^M)\right] . \tag{2.18}$$

Here, the $L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$, which is defined as "image likelihood" is the likelihood value of the i_{th} particle estimated from the observation region in the frame of the m_{th} camera at the state $\mathbb{X}_k$. The shape of the observation region is a circle whose size is calculated by projecting the physical size of ball from 3D space. The set containing all the image likelihoods is represented as $L_{img}^{(i)}$. $g(x)$ is a function to combine each image likelihood obtained from each camera. We propose an anti-occlusion observation model, which switches the function $g(x)$ based on the confidence of cameras, to solve the occlusion problem.

2.6.1 Anti-occlusion Observation Model

Image likelihood is used to evaluate confidence of each camera. The confidence of one camera represents the observation probability of a ball in the video of this camera. If the image likelihood value $L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$ satisfies

$$L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) > thr_{img} \tag{2.19}$$

with a given threshold thr_{img}, the ball is assumed being visible to the m_{th} camera since the confidence of the m_{th} camera is high.

By this definition, I judge the occlusion type for each particle and switch the combination function $g(x)$ based on value #, the number of cameras with high confidence by equation (2.20).

$$g(\cdot) = \begin{cases} \sqrt[\#]{\prod_{\alpha=1}^{\#} L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^{\alpha})} & 2 < \# \leqslant M \\ \sqrt[M-1]{\prod_{\beta=1}^{M-1} L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^{\beta})} & \# \leqslant 2 \end{cases} . \tag{2.20}$$

1) When $\# = M$, all of the cameras are visible cameras. In this case, there is no occlusion since the confidence of every camera is high; hence, all of them can be used to calculate the likelihood of the particle.

2) When $2 < \# < M$, it is assumed that the ball is occluded in some cameras while being visible in others. In this case, since the number of visible cameras is enough to obtain 3D position of the ball, the low image likelihood can be removed directly and the image likelihood of visible cameras can be combined by $g(x)$. Here, $L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^{\alpha}) \in L_{img,v}^{(i)}$ and $L_{img,v}^{(i)}$ is the set of visible image likelihood:

$$L_{img,v}^{(i)} \subseteq L_{img}^{(i)} . \tag{2.21}$$

3) When $\# \leqslant 2$, there are two possible situations. One is that the ball is severely occluded in most cameras; the other is that this particle is not located at the position of the ball. In this case, the camera with the lowest confidence is detected in the combination step. Here, $L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^{\beta}) \in L_{img}^{(i)} - \left\{L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^{m})\right\}_{lowest}$.

2.6.2 Ball Feature based Likelihood Model

There are three kinds of likelihood defined in this tracking algorithm: circle likelihood, HSV color likelihood and foreground likelihood, which are based on the shape, color, and foreground feature of the ball, respectively. Thus, the image likelihood $L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$ is represented as:

$$\begin{aligned} &L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) = \\ &L_{circle}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) \cdot L_{color}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) \cdot L_{foreground}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) \,. \end{aligned} \tag{2.22}$$

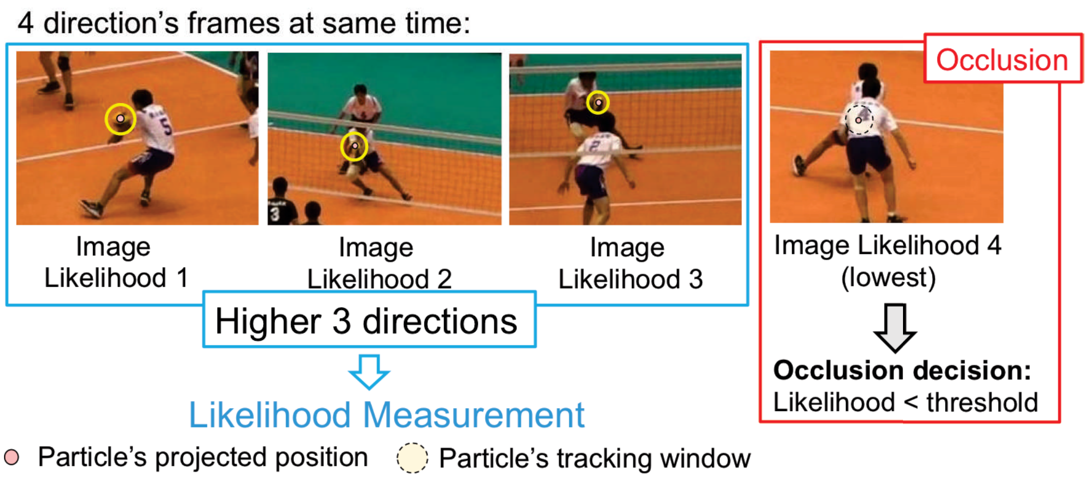


Figure 2.8: Anti-occlusion observation model.

2.6.2.1 Circle likelihood

The circle likelihood is evaluated by the circumference gradient of the observation region that is effective for distinguishing circle objects from other items. For each pixel on the circumference of the observation region, the component of their gradient pointing to the center of the region is defined as the circle gradient g_c:

$$g_c = g_y \cdot sin\theta + g_x \cdot cos\theta \, . \tag{2.23}$$

Here g_y and g_x are the vertical and horizontal gradient values of the pixel which is calculated in gray scale images. θ is the included angle between the line from center of the sampling region to current target pixel and the horizontal line. I choose N representative pixels on the circumference with the same internal to calculate the circle likelihood $L_{circle}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$:

$$L_{circle}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) = 1 - e^{-\frac{1}{N}\sum_{n=1}^{N}|g_c|} \, . \tag{2.24}$$

2.6.2.2 Color Likelihood

To evaluate the color likelihood, some typical template images are prepared, which cover the different color appearance of the tracking target from different perspectives while under different lighting environment, and extract the HSV color histogram of these images as the template histogram. After obtaining the HSV color histogram from the observation region, the color likelihood is evaluated by comparing the distance between sampling region and the prepared template images:

$$L_{color}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) = e^{-\frac{1}{N}\sum_{n=1}^{N} d_n} \, , \tag{2.25}$$

where d_n is the Bhattacharyya distance between the color histogram of the observation region and the n_{th} template histogram and N is the number of templates.

2.6.2.3 Foreground Likelihood

The foreground feature of the ball can be used to distinguish moving objects from static ones and still background. Here, background subtraction is performed such that a binary mask image is obtained in which the black pixels represent static objects. The foreground likelihood $L_{foreground}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$ of the particles is calculated by using equation (2.26):

$$L_{foreground}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) = \frac{1}{N} \sum_{n=1}^{N} \frac{value_n \cdot d_n}{R \cdot 255}, \tag{2.26}$$

where d_n is the distance between the n_{th} pixel and the center pixel of the sampling region, N is the number of pixels in the sampling region whose radius is R, and $value_n$ is the pixel value of the n_{th} pixel.

2.7 Estimation

2.7.1 State Estimation

The state is estimated according to the posterior distribution $p(\mathbb{X}_k; \mathbb{I}_k)$ by the Monte Carlo method and a Bayesian equation [23].

Particles are reordered from the largest to the smallest according to their weight, and a sequence of important particles is chosen to estimate the required state. I^* is the number selected important particles, which is decided when the value $|\sum_{i=1}^{I^*} \mathbf{w}^i - \mathbf{w}^*|$ is smallest. $\mathbf{w}^*$ is a given weight threshold and $\mathbf{w}_k^i$ is the normalized importance weight, which can be computed by

$$\mathbf{w}_k^i \sim p(\mathbb{X}_k | \mathbb{I}_k) = L_i(\mathbb{X}_k). \tag{2.27}$$

Then this sequence of important particles devotes their associated weights to approximating the required state probability as:

$$p(\mathbb{X}_{1:k}|\mathbb{I}_k) \approx \sum_{i=1}^{I^*} \mathbf{w}_k^i \delta(\mathbb{X}_k - \mathbb{X}_k^i)\,, \tag{2.28}$$

where $\delta(\cdot)$ is the Dirac delta function.

2.7.2 Adaptation to Abrupt Motion

α_k and I_k are two important items in equation (2.16) for adaption the system model to the motion condition of the ball, which is estimated depending on weight distribution of particles transited by different system noise. The particles are assumed following the general and abrupt models as P_g and P_a, respectively.

The velocity and the moving direction of the ball change dramatically when being hit by players. Since the strength and direction of players' force is unknown and unpredictable, the particles transited by general model cannot track the target thus the weight of general particles is low. This leads the value of $\frac{E_k[w_a^{(i)}]}{E_k[w_g^{(i)}]}$ to be small and $Var_k[X_g^{(i)}]$ to be large. $E_k[w_g^{(i)}]$ and $E_k[w_a^{(i)}]$ are the weight expectation of P_g and P_a, and $Var_k[X_g^{(i)}]$ is the position variance of P_g. Abrupt motion requires a large number of particles I_{k+1} and small value of the model ratio α_{k+1} in next time step. Therefore, the values of α_{k+1} and I_{k+1} update at every time step as:

$$\alpha_{k+1} = \alpha_0 + a \cdot e^{-\frac{E_k[w_a^{(i)}]}{E_k[w_g^{(i)}]}}\,, \tag{2.29}$$

$$I_{k+1} = I_{init} \cdot (1 + b \cdot Var_k[X_g^{(i)}])\,, \tag{2.30}$$

where I_{init} is the initial number of particles, a and b are two constants.

With this proposal, the proposed approach not only enables the system model to be adaptive according to the moving situation of the ball, but also rectify a small bias of the tracking result to some extent by increasing the number of particles.

2.7.3 Judgement of Tracking Failure

In particle filter algorithm, tracking failure occurs under two situations. First, the tracker miss-tracks on other objects or it loses the target completely. Second, the ball moves out of the image frame. In this situation, there is no corresponding observation space, so that the tracker cannot track the ball. Therefore, there are two corresponding criterions to judge the tracking failure situation as equation (2.31) and equation (2.32).

$$Var[\mathbb{X}^{(i)}] > tr_{fail}\,, \tag{2.31}$$

$$\#^{*} > M - 2\,. \tag{2.32}$$

1) Equation (2.31) is the criterion for the first situation. In this case, the weight distribution of particles $\mathbf{w}_k^i$ will be smoothed without obvious peaks; thus, the position variance of all the particles is expected to be large.

2) Equation (2.32) is the criterion for detection of ball moving out of image. In this equation, $\#^{*}$ is the number of cameras from which the ball disappears, which is calculated by coordinate projection.

In my tracking approach, if any one of the above judgment criterions is satisfied, an automatic recovery module will be employed to re-initialize the tracker by proposed spatial density which is described in following subsection.

2.8 Automatic Recovery from Tracking Failure

When tracking failure is detected, the tracker will be re-initialized by two steps: filtration of 3D ball candidates from each image frame, and generation of particles according to spatial density.

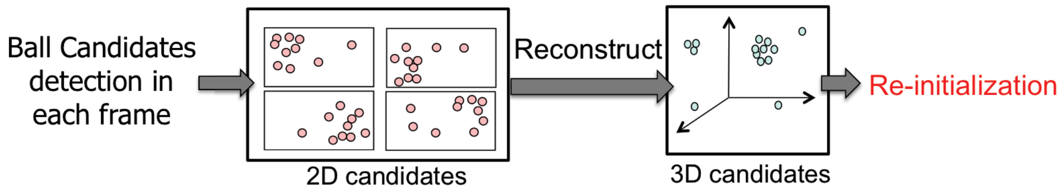


Figure 2.9: Ball candidate distribution based automatic recovery from tracking failure.

2.8.1 Ball Candidates Filtration

All the pixels in the video frame are raster scanned with a certain interval. For each scanning pixel p, a likelihood value ξ_p is assigned to it, which represents the probability of pixel p being the center of the ball:

$$\xi_p = \sum_{d_{p,q}<R} L_q \frac{d_{p,q}}{R}, \tag{2.33}$$

where L_q is the likelihood value of pixel q, which is calculated by the ball feature-based likelihood model in a circle region whose radius is R and $d_{p,q}$ is the distance between pixel q and the scanned pixel p.

There are two reasons to use equation (2.33) to calculate ξ_p. First, for one target pixel, the likelihood value not only represents this pixel, but also represents a circle whose center is this pixel. Second, since the frame is canned with an interval, for each un-calculated pixel, it must have a likelihood value. Considering the overlap part of different circles and the un-calculated pixel, the equation (2.33) is employed to calculate the likelihood value.

Then, the pixels with a high likelihood value are chosen and concentrated into some 2D ball coordinates according to their distance relationship. After reconstructing each combination of these 2D coordinates into 3D space by equation (2.1), a sequence of 3D coordinates $\mathbf{S}_k^n, n = 1, 2, ..., N$ is obtained. Here N is the total number of reconstructed 3D coordinates.

At last, the 3D ball candidates are filtered out by eliminating the obvious noise that is evaluated as:

$$min\{|\mathbf{S}_k^n - \mathbf{S}_k^{n^*}|\} > tr_{dis1} , \tag{2.34}$$

where tr_{dis1} is a threshold of isolate coordinate judgment. Isolated coordinates are removed because they are erroneous matches of 2D positions.

2.8.2 Re-initialization by Ball Candidates Distribution

The objective new state $\mathbb{X}_k^{(i)}$ of particles is a set of states, which is derived according to the proposed ball candidate density.

In 3D space, the importance of each point is described by the density of filtered ball candidates $\mathbf{S}_k^n$. For the n_{th}^* candidate, the evaluation of the density $\mathbf{D}_k^{n^*}$ is

$$\mathbf{D}_k^{n^*} = 1 - \frac{1}{\lambda} \sum_{n=1}^{N} (\mathbf{S}_k^n - \mathbf{S}_k^{n^*})^2 , \tag{2.35}$$

where, λ is a coefficient to ensure

$$\sum_{n=1}^{N} \mathbf{D}_k^n = 1 . \tag{2.36}$$

Therefore, the set of $\mathbf{D}_k^{(i)} = \left\{\mathbf{D}_k^1, ..., \mathbf{D}_k^n, ..., \mathbf{D}_k^N\right\}$ can be used to describe the importance of 3D positions in the whole 3D space.

Among all the 3D ball candidates, if density $\mathbf{D}_k^n$ is high, it is assumed that the 3D position $\mathbf{S}_k^n$ is more important and more particles will be generated to the corresponding state $\mathbb{X}_k^n$. Hence, number of particles distributed to each $\mathbb{X}_k^n$ is decided as:

$$I_n = I_{init} \cdot \mathbf{D}_k^n \,. \tag{2.37}$$

For each 3D candidate $\mathbf{S}_k^n$, $\mathbb{X}_k^n$ is generated as

$$\mathbb{X}_k^n = \begin{pmatrix} \mathbf{S}_k^n + \mathbf{V}^n \\ \mathbf{V}^n \end{pmatrix}. \tag{2.38}$$

$$\mathbf{V}^n \sim N(0, \mathbf{Q}), \qquad \mathbf{Q} = diag\left(\tau^2, \tau^2, \tau^2\right). \tag{2.39}$$

In some special situations, there is no tracking target in the videos or the target is visible in only one camera. By spatial geographic relation, all the reconstructed 3D points are removed as isolate coordinates in ball candidate filtration step. In this case, without 3D ball candidates, this frame will be skipped whereupon automatic recovery is executed in the next frame.

2.9 Experimental Results

2.9.1 Introduction of Experimental Videos

The experiment is based on multi-view videos recorded during the final game of an official volleyball match (2014 Inter High school Men's Volleyball Games held in the Tokyo Metropolitan Gymnasium in Aug. 2014), which contains three sets. These experimental

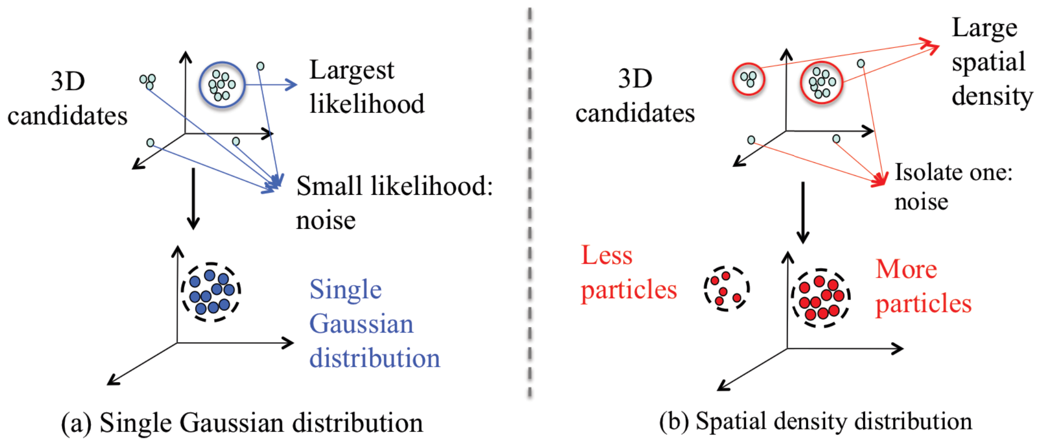


Figure 2.10: Comparison of the ball candidate distribution based initialization.

game videos were captured by four cameras located at each corner of the court. View angles of the four input videos are shown in Fig. 2.11. Video resolution is 1920×1080, frame rate is 60 frames per second, and shutter speed of the cameras is 1000 frames per second. The latter parameter prevents motion blur in the video sequence. Under these view angles and camera parameters, the radius of the ball in image changes from about 7 to 15 pixels.

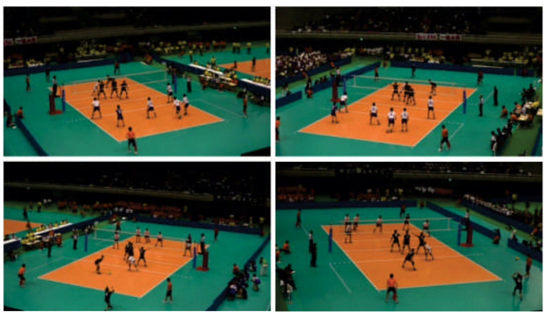


Figure 2.11: View angles of four input videos for the experiment.

The recorded match includes all kinds of volleyball scenes, such as start serve, receive, set, attack, and block. In these scenes, motion of the ball changes to different states, especially in serve and spike, where velocity and direction of the ball change abruptly and dramatically. In addition, sometimes the ball is occluded by players or other objects in certain views.

2.9.2 Implementation Details

The entire multi-view tracking algorithm of particle filter is implemented with OpenCV and GSL library. For coordinate projection and reconstruction between 2D and 3D space, all the cameras are calibrated by manually chosen key points. The implementation of particle filter algorithm follows the description in section 2.3 except the three proposals. All the parameters are set by physical laws and adjusted based on experiments. Initial particle number I_{init} is assigned as 1000 while the variance value τ_S in equation (2.6) is $0.5m$ and τ_V in equation (2.7) is $100m/s$. $\mathbf{w}^*$ in estimation step is selected as 0.382.

As for the implementation of proposals, in order to appeal the contribution of each one distinctly, the three proposals are implemented step by step.

The first one is the system model. Two system noises are employed in system model and adaption of the system model parameter to the estimated motion is added in estimation part. In general system noise, $\tau_{g,\rho}$, $\kappa_{g,\theta}$, G and $\kappa_{g,\phi}$ in equation (2.13) are $500m/s^2$, $0.1\pi rad/s$, $0.002\pi rad/s$ and $0.2\pi rad/s$, respectively. As for the abrupt system noise in equation (2.14), $\kappa_{a,\theta}$, G and $\kappa_{a,\phi}$ are $0.1\pi rad/s$ and $0.2\pi rad/s$. And in the abrupt motion adaption part, value of a in equation (2.29) is 0.01 and b in equation (2.30) refers the size of the ball radius and in the experiment it is 0.1 times ball radius.

Meanwhile, two conventional system models [1] [24] are also implemented for comparisons. Including my abrupt motion adaptive system model, these three system models are implemented in totally same tracking framework.

System model used in [24] is based on typical Gaussian distribution whose state vector only represents the position state. Although this work focus on 2D objects tracking, I implemente Hess's system noise in the 3D particle filter system model to compare the performance between the Gaussian distribution and the abrupt adaptive system model. In [1], the mixed-state system model introduced in their paper is difficult to employ in the framework (since there is no sampling time in their equations). So I modify their equations and parameters without changing the their concept.

The second one is the anti-occlusion observation model. This part is implemented based on the particle filter framework with abrupt motion adaptive system model. At last the ball candidate distribution based recovery is employed in addition to the former two proposals.

2.9.3 Experiment of Tracking Success Rate

2.9.3.1 Evaluation Method

Since the real 3D position of the ball is almost impossible to obtain, the success rate is difficult to evaluate in frame level. Instead, a definition of HIT is given as an evaluation unit to calculate the tracking rate. HIT is a class of 3D ball coordinates between two consecutive hit time:

$$HIT = \{S_t, S_{t+1}, ..., S_T\}, \tag{2.40}$$

where, S_t and S_T are two hit points and there is no other hit point in class HIT. For example, there are five HITs in Fig. 2.12. By this definition, the experimental game videos

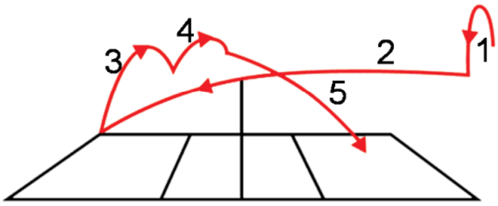


Figure 2.12: Example to explain the definition of HIT.

contain a total of 684 HITs. The successful tracking of a HIT are judged depending on following conditions:

1) In this HIT, for each tracked coordinate $\mathbb{S}_{1:k}$, which is the position of estimated state $p(\mathbb{X}_{1:k}|\mathbb{I}_k)$, the criteria below is satisfied:

$$|\mathbb{A}_m \cdot \mathbb{S}_{1:k} - S_k|\mathbb{I}_k^m| < R_m, m = 1, 2, 3, 4\,, \tag{2.41}$$

where $S_k|\mathbb{I}_k^m$ is the ball's true coordinate and R_m is the radius of the ball in the m_{th} camera; $\mathbb{A}_m$ is the calibration matrix of the m_{th} camera.

2) The plotted trajectory during the HIT is a continuous curve without sharp gap.

If a HIT is satisfied by the two conditions mentioned above, this HIT is a successful HIT. Therefore, the tracking success rate is described as:

$$success\ rate = \frac{\sum successful\ HIT}{HIT} \times 100\%\,. \tag{2.42}$$

2.9.3.2 Experimental Result

Data in Table 2.1 shows the tracking success rate of all the experiments.

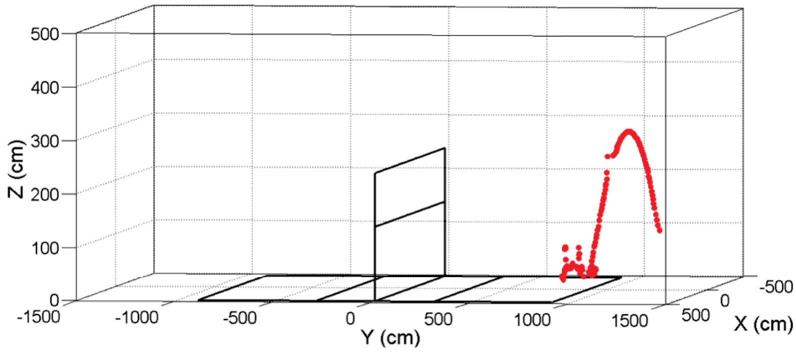
Rows 2, 3 and 4 in Table 2.1 is the comparison result of different system models. Success rate of proposed abrupt motion adaptive system model (80.12%) is higher than that of the conventional method (48.25% and 53.65%). It is because when motion of the ball changes abruptly with high speed, the adaptive system model can estimate the abrupt change and adapt the model that addresses to the motion. In Deguchi's work [1], which also mentions the similar solution by bouncing model. However this model aims at ground bouncing situations in which the ball goes to opposite direction after hitting, which cannot cope with the motion caused by the unknown external force of players. Fig. 2.13 shows an example of this visually. Fig. 2.13 (a) is the result trajectory tracked by Deguchi's work. At the start of the serve, direction and speed of the ball changes abruptly and the tracker loses the target. Meanwhile, as trajectory (b) shows, the proposed adaptive system model is capable of tracking the ball even in situations such like this.

Table 2.1: Experimental results of the proposals.

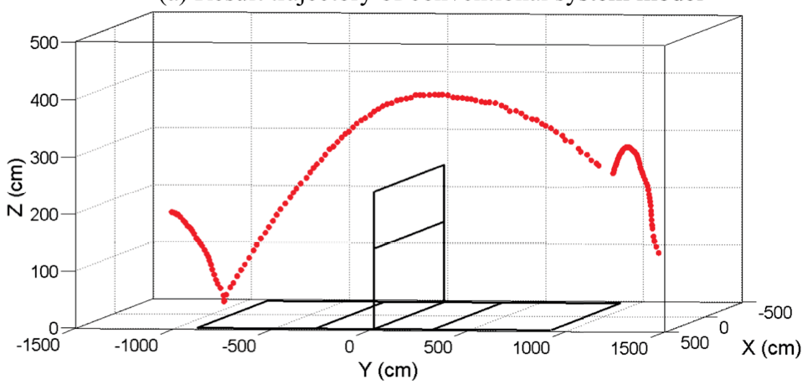
Proposal	**Successful HIT**	**Success rate**	**Time (per frame)**
Conventional system model 1 [24]	330	48.3%	138 ~155ms
Conventional system model 2 [1]	367	53.7%	139~155ms
P1	548	80.1%	154~411ms
P1+P2	671	98.1%	154~411ms
P1+P2+P3	680	99.4%	20~30s

P1: Abrupt motion adaptive system model
P2: Anti-occlusion observation model
P3: Ball candidate distribution based automatic recovery

As the data shown in row 4, 5 and 6 in table 1, the contribution of each proposal



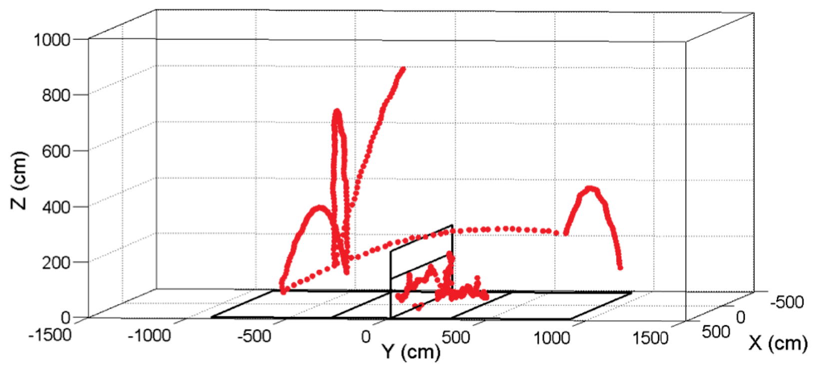
(a) Result trajectory of conventional system model



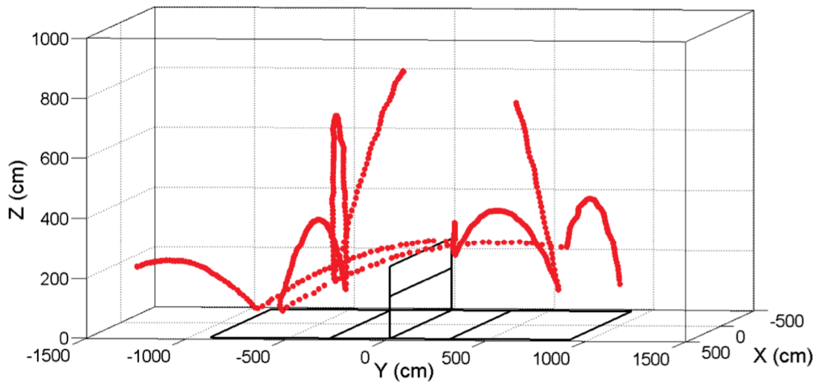
(b) Result trajectory of abrupt motion adaptive system model

Figure 2.13: Tracking trajectories comparison of system model. (a) Result trajectory of conventional system model [1] that fails tracking at serve of the ball; (b) Result trajectory of proposed abrupt motion adaptive system model.

in this paper is shown. By solving the occlusion problems through eliminating cameras with low confidence, the anti-occlusion observation model raised the tracking rate from 80.12% to 98.10%. At last, with additional help from the ball candidate distribution based recovery part, the tracking success rate achieves 99.42%. As for the time computation, the additional calculation for the particles adaption to the ball motion increases about 10 300ms for each frame. The large difference of time computation is because the particle number will increase a lot when the ball is in abrupt motion. The implementation of my



(a) Result trajectory without automatic recovery



(b) Result trajectory with employing automatic recovery

Figure 2.14: Tracking trajectories comparison for automatic recovery. (a) Result trajectory without automatic recovery; (b) Result trajectory with employing automatic recovery, which can re-tracking the ball when it moves back to image.

algorithm is sequential processing for each particle, so that the time computation becomes large when particle number increases. In the last row, it can be known that the automatic recovery method takes long time. This method only works when tracking failure occurs. In total sequence, there are more 114 round games including more than 2,052,000 frames (assuming there are 300 frames in each round) and the recovery method only works 5 times. Since the proportion of recovery method is very small, the long time computation doesn't matter so much in the whole tracking. Fig. 2.14 shows the performance in terms of the recovery part. Fig. 2.14 (a) is an trajectory of tracking failure caused by the ball leaving the video frame. Fig. 2.14 (b) shows the trajectory obtained by the tracker to which a recovery method has been added. The ball can be tracked again at once when it moves back into the scene. From the view of 2D frame, the tracking trajectories are also plotted in the frame as Fig. 2.15 shows.

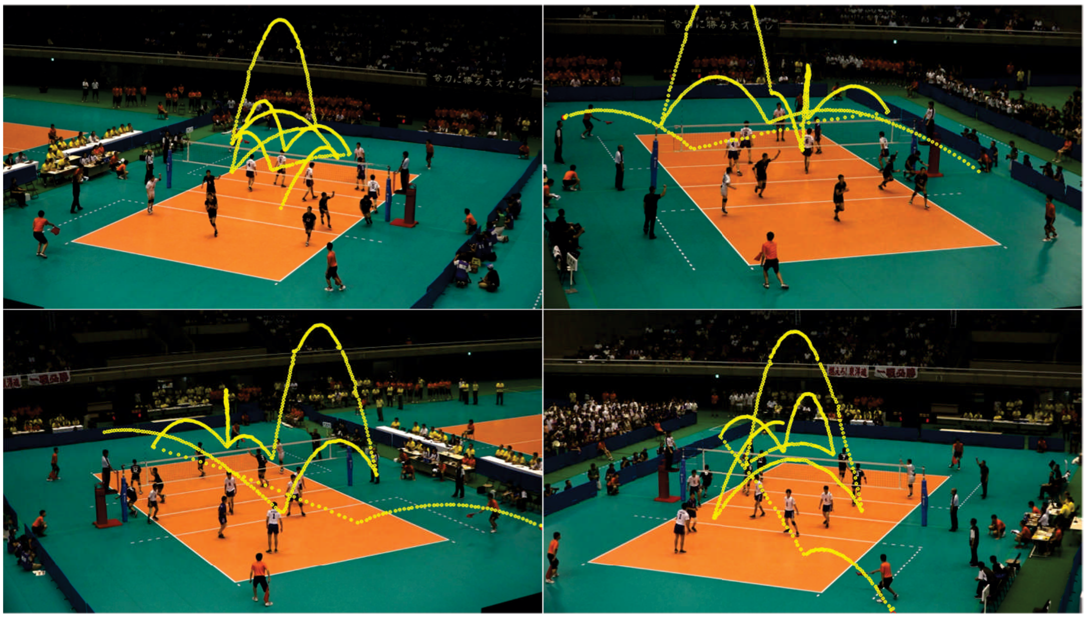


Figure 2.15: The tracking trajectory on each view.

2.9.4 Evaluation of Result Precision

The precision is evaluated from two aspects: the 2D position precision and the 3D position precision.

I choose one round from the game video sequence that contains 200 frames to test the accuracy of the tracking result. Trajectory of this round is shown in Fig. 2.16. There are one serve (frame 1 to frame 75), one receive (frame 74 to frame 131), one set (frame 132 to frame 179) and one attack (frame 179 to frame 200) in this round.

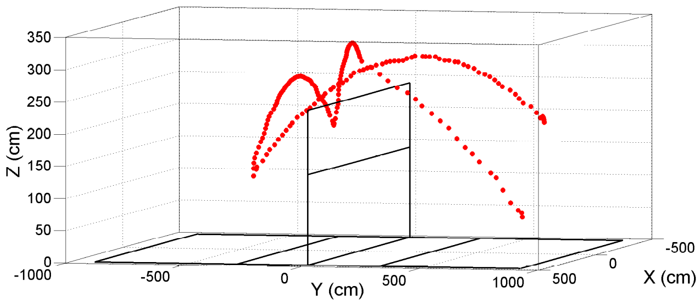


Figure 2.16: Trajectory of the sequence in position precision evaluation.

2.9.4.1 2D Position Precision

To evaluate the 2D position precision, all the tracked 3D coordinates are projected at each sampling time to four image spaces. Then the 2D precision error is defined as the distance between the projected tracking result and the center position of the ball in the image. Center position of the ball in each image is picked up manually.

The result graph is shown in Fig. 2.17. In our video image, the diameter of the ball changes from 15 pixels to 30 pixels. In the graph, even the largest distance error is less than 10 pixels and the average 2D distance error of each camera is 1.52, 3.06, 2.24 and 2.67 pixels.

2.9.4.2 3D Position Precision

As the statement in the former subsection, in our research it is impossible to obtain the ”true” 3D position of the ball from just four game videos, because of the absolute error in the measurement process. Without a ”true” position, it is difficult to evaluate the 3D

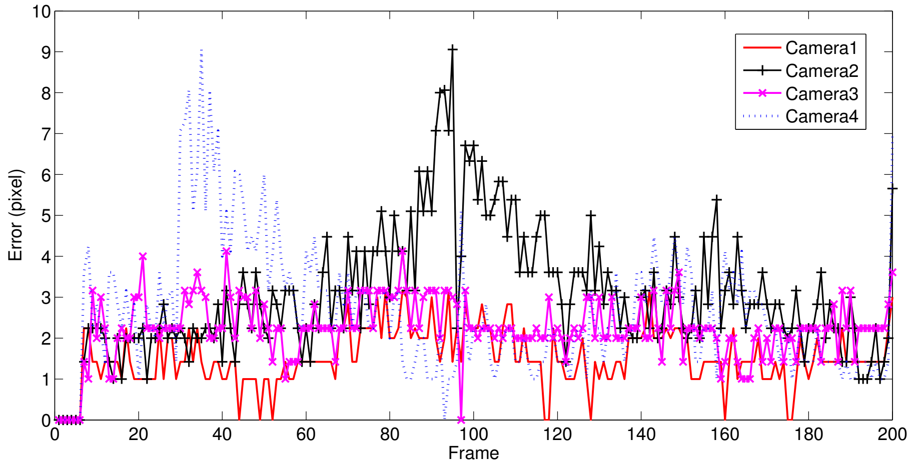


Figure 2.17: The 2D precision error.

precision. In our experiment part, to evaluate the 3D precision error, I create the 3D positions by the multiple images information and assume these positions as ”true” value to evaluate precision.

To obtain a 3D reference position, I pick up the center pixel of the seperate frame and reconstructed these 2D coordinates from different views to 3D coordinate in real space. The 3D precision error is the distance between the tracked 3D position and the reference position. Fig. 2.18 shows the 3D position error of the test sequence. As the graph shows, the largest 3D position error is less than 19 centimetres whereas the diameter of the ball is about 21 centimetres. And the average 3D precision error is 5.32 centimetres. Because in the process of 2D coordinate picking and the 3D coordinate reconstruction, there are absolute error produced which cannot be eliminated, both the error of 3D precision and 2D precision is not the ”true error” in our experiment. The result is just a rough precision evaluation.

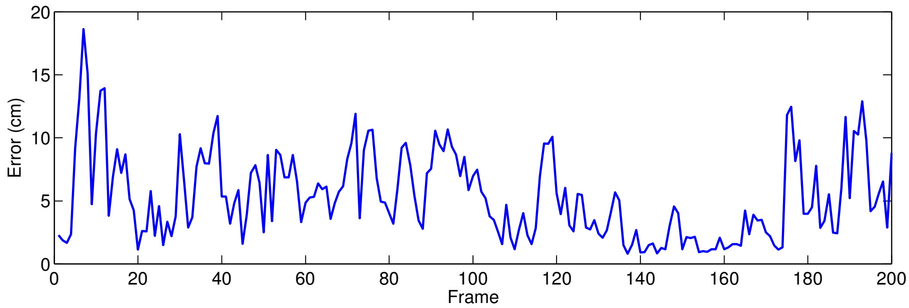


Figure 2.18: The 3D precision error.

2.10 Conclusions

For the acquisition of physical data in 3D sports analysis, this chapter proposes a multi-view 3D ball tracking with three proposals to achieve high tracking success rate.

Firstly, the abrupt motion adaptive system model reduces tracking loss by abrupt motion of the ball caused by unknown external force. System noise is adapted to different motion of the ball depending on the weight distribution of particles that follows each system model. Secondly, the anti-occlusion observation model eliminates interference by the occlusion problem that occurs in certain directions. The confidence of each camera is evaluated by the corresponding image likelihood. Thirdly, ball candidate distribution is proposed in the automatic tracking recovery part. When tracking failure is detected, the tracker will re-generate new particles according to the ball candidate distribution, which is evaluated by distance between 3D ball candidates. The proposed tracking approach incorporating three methods was implemented on real volleyball game videos acquired from four different views. The tracking success rate is 99.4%.

Due to the filtering scheme, the system model adaption after abrupt motion occurrence causes tracking delay. For experiments, various game videos, which include different matches with different players in different places, are needed for comprehensive evaluation of our work. In addition, in order to dive into higher-level analysis of the sports analysis, a smooth method is necessary for the acquisition a more stable ball trajectory.

Chapter 3

Multiple Event Change Feature and Event Change Probability Distribution based Event Detection

3.1 Introduction

For the acquisition of the ball event data in the 3D sports analysis, the ball event detection is a key technology. As Fig. 3.1 shows, the ball event data acquisition is the process based on the physical data results. In the volleyball game, the motion of the ball not only helps a lot in score judgment but also plays a great role in representing players' behaviors and team performances. Targets of this chapter is the acquisition of the ball event, like the serve, pass (receive and toss) and spike event based on the ball trajectory and the game videos.

Due to the limitations of the shooting conditions in real game and the complex game rules of the volleyball game, there are three factors affect the performance of ball event detection: the noise in the unstable ball trajectory, the unknown time of the player action that changes the event and the large intra-class difference of the same event type. Firstly,

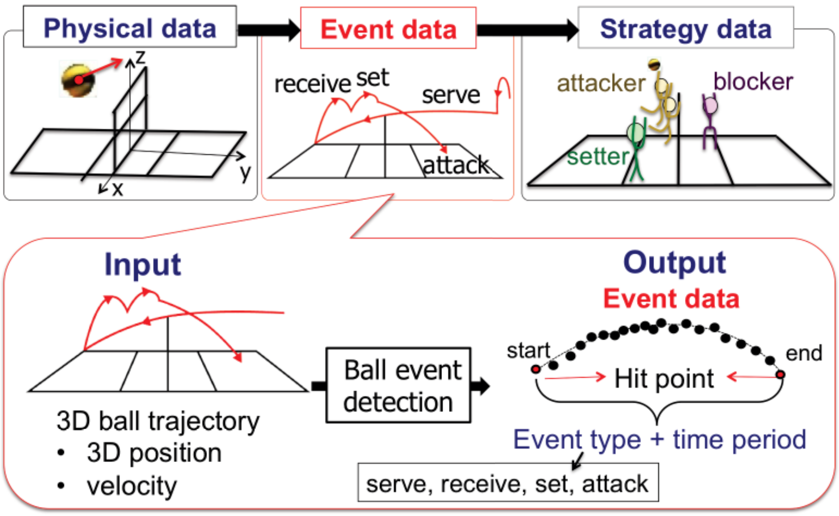


Figure 3.1: Introduction of the ball event acquisition and whose position of in entire research.

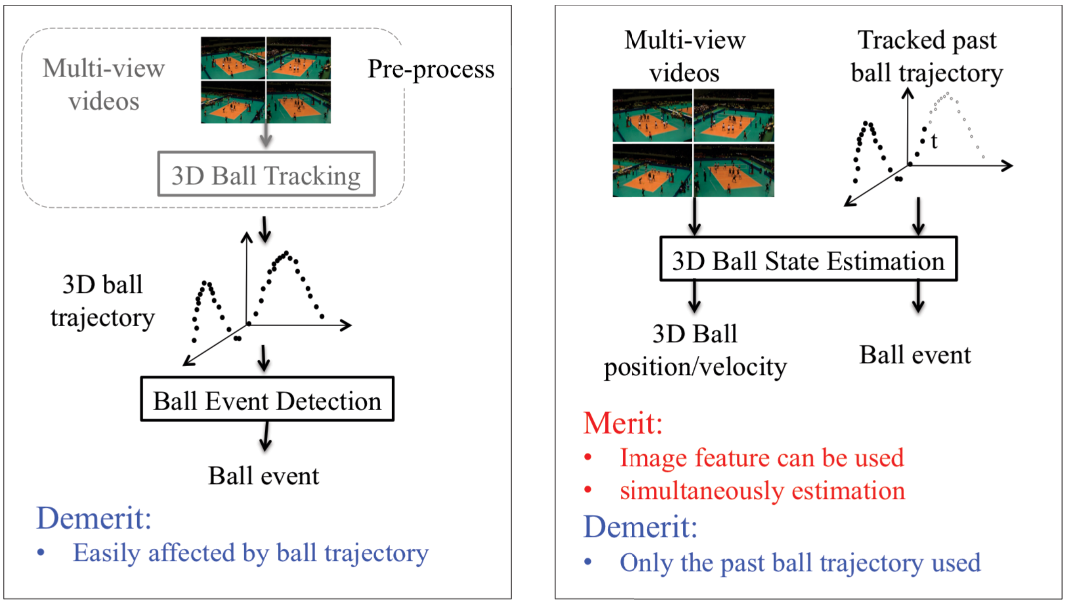
noise in the unstable and rough ball trajectories is bad for feature extraction of ball event. Not only the positions, but the speed and motion directions also reflect important features of a ball event as well. Especially, when the ball is hit by player, the velocity changes dramatically before the position trajectory turns. Without the integrate as well as accurate trajectory, the event change can only be estimated by the velocity variation since the turn of the trajectory is not obvious through local short trajectory segment. However, from real game videos the accurate 3D ball trajectory is difficult to track due to the complex background as well as the motion of the ball. The small flicker of position trajectory corresponding to large difference in velocity affects the reliability of event feature extraction from velocity information. Secondly, the time of the event change when the player hits the ball is unknown. In volleyball game, without referring the information of player, the time when the player the ball is unknown and different event continues different time length in different situations. Therefore, it is difficult to predict the event change temporally based on the ball trajectory. Thirdly, the intra-class differences of some ball event are large, which bring difficulties in event feature extraction. In volleyball game, the located position and the continuing time of one event change a lot in different situations. Such as one pass event may occur in anywhere even the outside of the court. And one spike may continue until: 1) being blocked around the upside of the net; 2) being received after crossing the net; 3) hitting the ground. The continuing time of events under the above three conditions are different although all of them belong to ”spike” event.

With the similar target in tennis ball event detection, whose event type is much simpler than that of volleyball, works of [25] classifies ball events according to the distance between the ball and the player or the court line. Because of the requirement of players' positions, this work is not suitable for volleyball game in which there are twelve players on the court. To avoid using the relationship between the ball and players, Almajai [26] has employed a hidden Markov model event detection method based on the work [25]. Both these two works are implemented on 2D game videos, event occurs at some critical cases, such as bouncing near the court bounds, are difficult to detect due to the lack of 3D information. By using 3D ball trajectory information and physical court locations, I have proposed a 3D ball trajectory base particle filter to detect ball event accurately in volleyball game analysis [27]. However, being same with work [25] and [26], this work detects the ball event after obtaining the whole trajectory of the ball so that they are unfriendly to be implemented in a real-time system.

In addition, all the works mentioned above for the event detection follows the sequential processing order as Fig. 3.2: tracking the ball trajectory first, then detecting the ball event based on the entire ball trajectory. There are two demerits of this sequential method: one is the performance of event detection is easily affected by imprecise ball trajectory;

the other is the large time delay between the ball tracking process and the ball event detection process, which makes it unfriendly to be implemented in a real-time system.

Figure 3.2: The sequential structure of ball tracking and event detection.



(a) Sequential structure

(b) Simultaneous structure

Figure 3.3: Comparison of sequential and simultaneous structure for ball tracking and event detection.

3.2 Framework of Event Detection

One volleyball game video can be compartmented by three kinds of temporal units: ”round-unit”, ”event-unit” and ”sampling time-unit”. The round-unit and event-unit are defined according to the volleyball game rules. One round-unit starts from the serve time and ends at scoring, and it is denoted by a series of event-units with disjoint time-steps. For each single round-unit, let n be the time index of event-unit beginning with 1 and N is total number of the event-units. The sampling time-unit whose time index is k, is defined by the frame rate of input videos. The overall structure of the physical and conceptual ball state tracking method over sampling time k is shown in Fig. 3.4. Compared with the sequential ball tracking and event detection method, this structure can obtain the physical ball trajectory and event simultaneously at each time step so that it has high potential for real-time implementation and the image feature based method which are dependent from using ball trajectory achieves a higher detection performance.

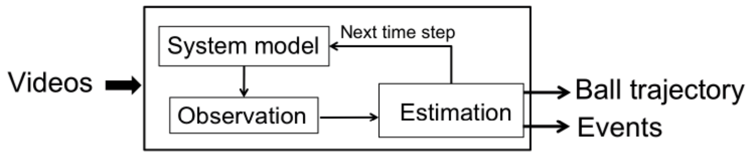


Figure 3.4: The proposed structure of the physical and conceptual ball state tracking based on particle filter.

3.3 State Vector and Initialization

The ball state $\mathbf{X}_k$ at discrete time k is denoted as

$$\mathbf{X}_k = (\mathbf{z}_k\ ,\ d_k\ ,\ e_k)\ , \tag{3.1}$$

where $\mathbf{z}_k$ is physical state of the ball consisting of its position and velocity in 3D space. d_k is a flag variable of external force (except the gravity), and e_k is a integer label representing the event type shown in Table 3.1.

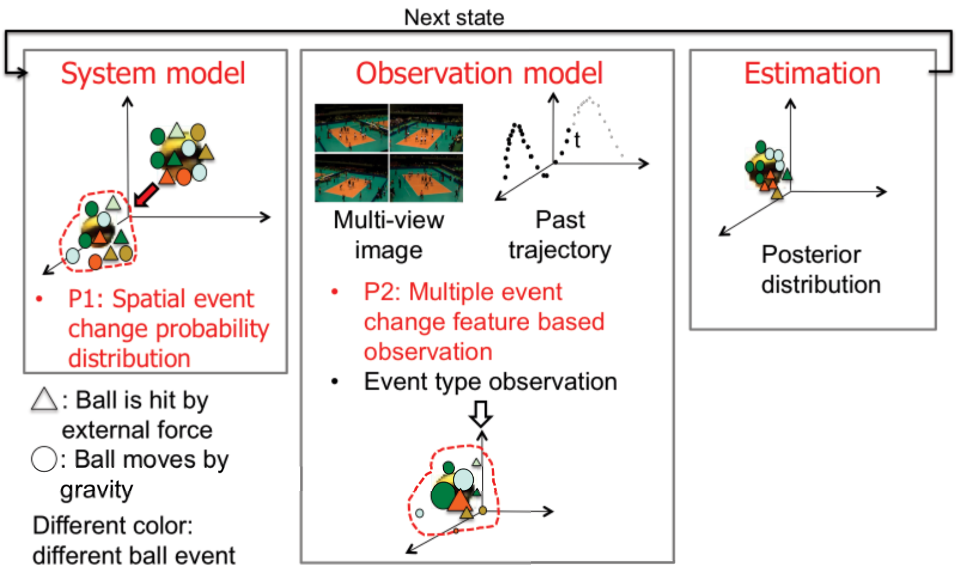


Figure 3.5: Overall structure and proposals of ball trajectory based event detection.

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

$$\mathbf{z}_0 \sim N(\bar{\mathbf{z}}, \Sigma_0), \Sigma_0 = diag(\tau_p^2, \tau_p^2, \tau_p^2, \tau_v^2, \tau_v^2, \tau_v^2)\,, \tag{3.2}$$

where $\bar{\mathbf{z}}$ is the manually given mean value of the distribution, τ_p and τ_v are the Gaussian noise variance for position term and velocity term. Initial distributions of flag variable of external force is

$$d_0 \sim \mathfrak{B}(p)\,, \tag{3.3}$$

where $\mathfrak{B}(p)$ is Bernoulli distribution with p being probability for value 1, and integer label of event type has a deterministic initial value

$$e_0 = -1\,. \tag{3.4}$$

Table 3.1: Event ID of a ball in volleyball game.

Event		
ID	Type	Description
-1	OTHERS	Others event besides the regular ones
0	SERVE+	Serve by team in positive half court
1	SERVE-	Serve by team in negative half court
2	PASS+	Pass ball by team in positive half court
3	PASS-	Pass ball by team in negative half court
4	SPIKE+	Spike by team in positive half court
5	SPIKE-	Spike ball by team in negative half court

3.4 Event Change Distribution based System Model

Time evolution of $\mathbf{X}_k$ is modelled by a Markov conditional probability distribution

$$\begin{aligned} p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1}) &= p(\mathbf{z}_k, e_k \,|\, d_k, \mathbf{X}_{k-1}) \cdot p(d_k \,|\, \mathbf{X}_{k-1}) \\ &= p(\mathbf{z}_k, e_k \,|\, d_k, \mathbf{X}_{k-1}) \cdot p(d_k \,|\, d_{k-1}) \,, \end{aligned} \tag{3.5}$$

where the spatial hitting points dense distribution are proposed for transition of d_k.

3.4.1 Spatial Hitting Points Dense Distribution

To express the transition probabilities of ball motion d_k, a transition matrix $\mathbb{Q}$ is used,

$$\mathbb{Q} = [\mathbf{q}_{ij}] = \begin{bmatrix} 1 - Pr_{hitting} & Pr_{hitting} \\ 1 & 0 \end{bmatrix}. \tag{3.6}$$

In matrix $\mathbb{Q}$, since each elements

$$\mathbf{q}_{ij} = Pr\{e_k = j \,|\, e_{k-1} = i\}\,, \tag{3.7}$$

$Pr_{hitting}$ means the probability the ball being hit by external force. Value of $Pr_{hitting}$ is decided by referring the corresponding probability of the position $\mathbf{z}_{k-1}$ in proposed spatial hitting point dense distribution.

In volleyball game, the ball hitting positions of each event always located at a certain area because of the game rule and team strategy, such as the spike and block hitting points are always near the net. To represent the spatial probability of event change, a group of hitting positions that can cover the entire event hitting situations can be used. In this paper, a Gaussian distributions H_m is utilized to approximate the probability distribution of one specific kind of hitting points. As preparation, some hitting point coordinates are prepared manually and clustered by K-Means depending on the relation of space position. The coordinate of each cluster center and the variance of the m_{th} cluster is defined as the mean value $\mathbb{C}_m$ and variance $[\sigma_x, \sigma_y, \sigma_z]^T$ of distribution H_m. Therefor, the spatial hitting point dense distribution H is a superposition of M independent Gaussian distributions.

$$H = \sum_{m=0}^{M} \frac{1}{M} H_m\,, \tag{3.8}$$

$$H_m \sim N(\mathbb{C}_m, \Gamma_m)\,,\, \Gamma_m = diag(\sigma_x^2, \sigma_y^2, \sigma_z^2)\,, \tag{3.9}$$

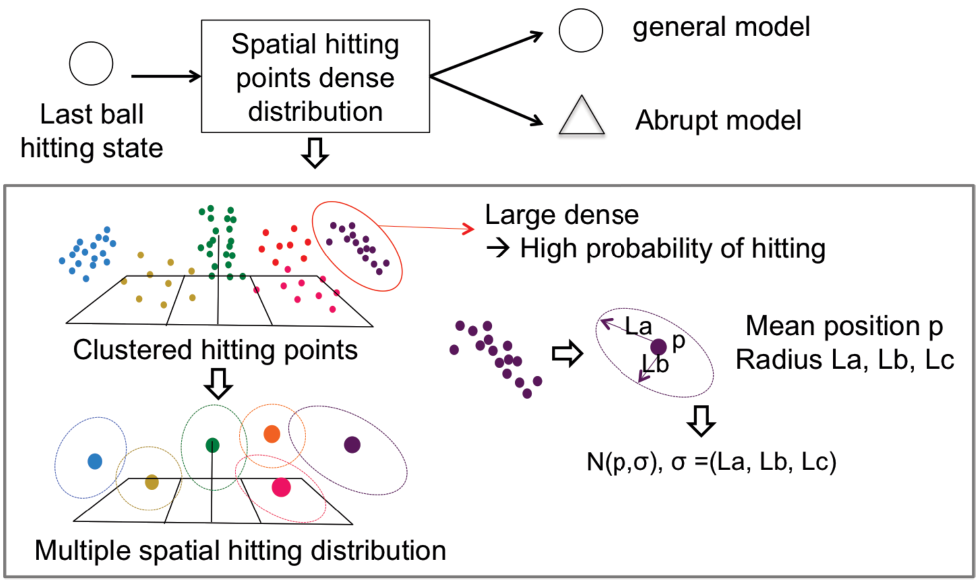


Figure 3.6: The spatial event change probability distribution based prediction.

3.4.2 External Force Adaptive System Model

After the prediction of d_k, the system model equation (3.5) of the ball event state can be rewritten as

$$p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1}) = \begin{cases} p_g(\mathbf{z}_k, e_k \,|\, \mathbf{X}_{k-1}), & d_k = 0 \\ p_a(\mathbf{z}_k, e_k \,|\, \mathbf{X}_{k-1}), & d_k = 1 \end{cases} . \tag{3.10}$$

When $d_k = 0$, the ball is moving without acting by other external force except the gravity, so the ball event also keep the same as previous ball state. The ball trajectory state $\mathbf{z}_k$ is transited as the general system model proposed in chapter 2 and

$$e_k = e_{k-1} . \tag{3.11}$$

On the other hand, when the ball is hit by external force, the motion speed, direction and event will change abruptly. Transition of $\mathbf{z}_k$ are defined as the same with the abrupt system model in chapter 3. Time evolution of event type e_k in abrupt motion situation is governed by state transition matrix $\mathbb{E}$ with its elements,

$$\mathbf{E}_{ji} = Pr((e_k = i) \,|\, (e_{k-1} = j)) \tag{3.12}$$

represents the probability of event type transition from e_{k-1} to e_k. To calculate the value of each elements in $\mathbb{E}$, referring some off-line data, such as the tracked event type and event result at the previous event-unit time e_{n-1}. In each event, there are at least event result. The values of all event result are summarised in the Table 3.2.

Table 3.2: Event result of a ball in volleyball game.

Event Result	**Description**
CROSS	The ball crosses the net
RECEIVE	The ball is received by play
NET	The ball touches the net
GROUND	The ball touches the ground
OUT	The ball moves out of the court

3.5 Observation

In this section, the likelihood of each term in the ball state vector are evaluated based on the synchronized multi-view image frames $\mathbb{I}_k$ and the past tracked ball trajectories $\hat{\mathbf{z}}_{k-L:k-1}$. $\hat{\mathbf{z}}_{k-L:k-1}$ is updated at each time step, whose length is L. Let y as the observation, then the full likelihood of the ball state $p(y_k \,|\, \mathbb{X}_k; \mathbb{I}_k, \hat{\mathbf{z}}_{k-L:k-1})$ consists of three terms:

$$\begin{aligned} p(y_k \,|\, \mathbb{X}_k; \mathbb{I}_k, \hat{\mathbf{z}}_{k-L:k-1}) =& p(y_k \,|\, \mathbf{z}_k; \mathbb{I}_k) * p(y_k \,|\, e_k, \mathbf{z}_k; \hat{\mathbf{z}}_{k-L:k-1}) \\ & * p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}_k, \hat{\mathbf{z}}_{k-L:k-1}) \end{aligned}. \tag{3.13}$$

Firstly, $p(y_k \mid \mathbf{z}_k; \mathbb{I}_k)$ is the ball feature likelihood evaluated by using the HSV color feature, the circle shape feature and the foreground feature. Secondly, to calculated the ball event likelihood $p(y_k \mid e_k, \mathbf{z}_k; \hat{\mathbf{z}}_{k-L:k-1})$, the ball's velocity state and related distance of the ball position and the court lines are extracted as the event feature. Based on the event feature, the detection result by SVM classifier [28] is used as the likelihood value. Thirdly, $p(y_k \mid d_k, \mathbf{z}_k; \mathbb{I}_k, \hat{\mathbf{z}}_{k-L:k-1})$ is the proposed image noise feature based past trajectory referred hitting point likelihood, which is introduced in the following subsection. The observation for ball event likelihood calculation is shown as Fig. 3.7.

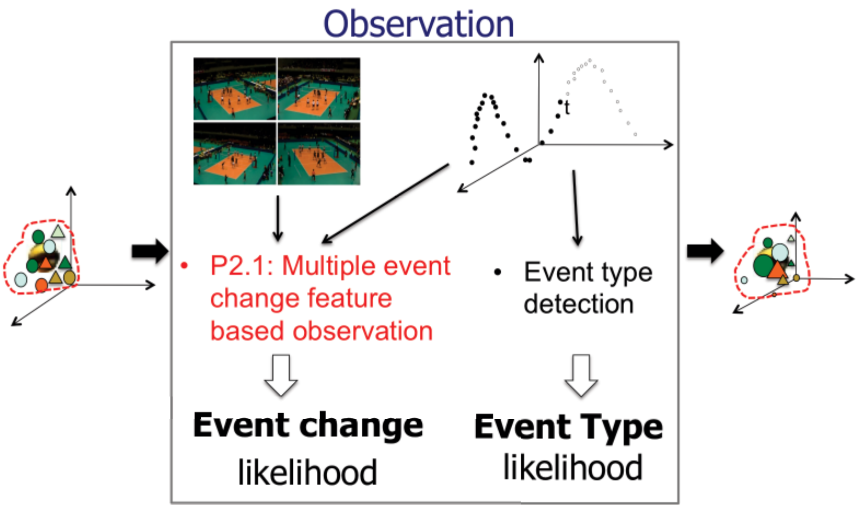


Figure 3.7: Observation of ball event.

3.5.1 Trajectory Abruptness and Player Skin Color Area based Event Change Observation

The event change observation evaluates the probability of external force added to the ball from two aspects: the image skin color area of the player body and the abruptness of the ball trajectory.

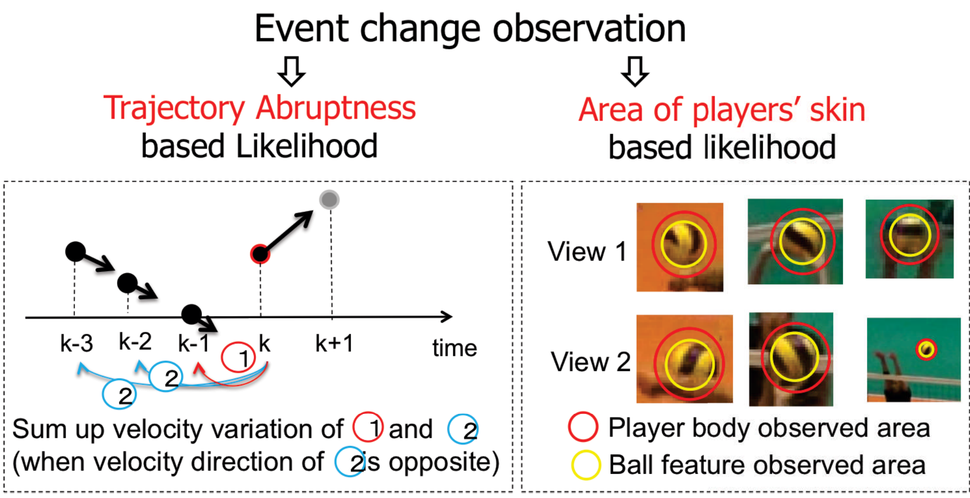


Figure 3.8: Trajectory abruptness and player skin color area based event change observation.

3.5.1.1 Player Skin Color Area based Event Change Observation

$$p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}_k, \hat{\mathbf{z}}_{k-L:k-1}) = p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}_k) \cdot p(y_k \,|\, d_k, \mathbf{z}_k; \hat{\mathbf{z}}_{k-L:k-1})\,, \tag{3.14}$$

where, $p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}_k)$ is the likeness calculated by skin color area feature in image:

$$p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}_k) = \prod_{i=0}^{I} p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}^i_{k,noise})\,, \tag{3.15}$$

where $p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}^i_{k,noise})$ is the likelihood calculated based on the filtered mask image of the player noise $\mathbb{I}_{k,noise}$ from the i_{th} view's image frame. The mask image is obtained by employing skin color filter and background subtraction on the original image frame. Since the skin color shares the similar feature with the ball color, this mask is assumed as image of player noise.

In volleyball game, when the ball moves close to the human body parts, probability of the ball being hit is high. Therefore the noise feature based hitting point likelihood is:

$$p(y_k \,|\, d_k, \mathbf{z}_k; \mathbb{I}^i_k) = \sum_{s \in \mathbb{O}} \frac{value_s}{dis_s}, \; \mathbb{O} = \mathbb{O}_{noise} \ominus \mathbb{O}_{ball}\,, \tag{3.16}$$

where $value_s$ is the pixel value of the pixel s in observation region $\mathbb{O}$ and dis_s is the distance from s to the center of $\mathbb{O}$. $\mathbb{O}_{noise}$ is the observed noise region in the image, whose diameter is two times of the projected ball region $\mathbb{O}_{ball}$.

3.5.1.2 Trajectory Abruptness based Event Change Observation

The other term $p(y_k \,|\, d_k, \mathbf{z}_k; \hat{\mathbf{z}}_{k-L:k-1})$ in equation (3.14) is the hitting likelihood calculated through the ball trajectory abruptness consisting of the variation of the velocity value and the moving direction variation on the ball's trajectory.

$$p(y_k \,|\, d_k, \mathbf{z}_k; \hat{\mathbf{z}}_{k-L:k-1}) = 1 - e^{(-\lambda \cdot var_k)}\,, \tag{3.17}$$

where var_k is the past trajectory referred velocity variation, which calculated both the value variation and the average difference summation of the velocity value at time k and those at the past time. λ is a constant value.

When the ball is hit by players or bounced from touching the ground, the velocity and moving direction of the ball must change abruptly and greatly. Therefore the hit point likelihood of the point at time s_k is calculated by velocity variation term $\mathbf{y}_p(s_k; \mathbf{z}_{0:k-1})$ and motion direction variation term $\mathbf{y}_d(s_k; \mathbf{z}_{0:k-1})$:

$$\mathbf{y}_{hit}(s_k; \mathbf{z}_{0:k-1}) = \mathbf{y}_p(s_k; \mathbf{z}_{0:k-1}) \cdot \mathbf{y}_d(s_k; \mathbf{z}_{0:k-1})\,. \tag{3.18}$$

where $\mathbf{y}_d(s_k; \mathbf{z}_{0:k-1})$ is the changed angle of ball's moving direction after time s_k;

$$\mathbf{y}_d(s_k) = \angle \widetilde{l}_{s_k-L:s_k}, \widetilde{l}_{s_k:s_k+L}\,, \tag{3.19}$$

where, $\widetilde{l}_{k_1:k_2}$ is the fitting curve from discrete positions $\mathbf{x}_{k_1}...\mathbf{x}_{k_2}$ by the least square fitting method. To evaluate the likelihood term $\mathbf{y}_d(s_k)$ which requires the trajectories of the future L time steps, this integrate hit point likelihood $\mathbf{y}_{hit}(s_k; \mathbf{z}_{0:k-1})$ can be obtained only when

$s_k < k - L$. In volleyball game, taking the consideration of the shutter speed and the ball velocity, one event continues longer than 3 time steps. Therefore, as long as $L \leqslant 3$ the hit point likelihood can be calculated immediately after obtaining ball tracking result.

In equation (3.18), the other term $\mathbf{y}_p(s_k; \mathbf{z}_{0:k-1})$ is evaluated by proposed past trajectory referred hit point likelihood, which only uses tracked trajectories.

$$\mathbf{y}_p(s_k; \mathbf{z}_{0:k-1}) = 1 - e^{(-\lambda_1 \cdot var_{s_k})}, \tag{3.20}$$

where var_{s_k} is the past trajectory referred velocity variation, which calculated both the value variation and the average difference summation of the velocity value at time s_k and those at the past time; Since the motion direction always changes at hitting time, the velocity difference contributes to the summation only when the motion directions of the two velocity are opposite. Through this algorithm, the variation feature of the velocity can be obtained distinctly.

$$var_{s_k} = |\dot{\mathbf{x}}_{s_k} - \dot{\mathbf{x}}_{s_k-1}| \sum_{l=1}^{l=L} \frac{1}{l} |\dot{\mathbf{x}}_{s_k} - \dot{\mathbf{x}}_{s_k-l}| \cdot \Delta_{s_k, s_k-l}, \tag{3.21}$$

$$\Delta_{s_k, s_k-l} = \begin{cases} 0, & \dot{\mathbf{x}}_{s_k} \cdot \dot{\mathbf{x}}_{s_k-l} > 0 \\ 1, & \dot{\mathbf{x}}_{s_k} \cdot \dot{\mathbf{x}}_{s_k-l} < 0 \end{cases}, \tag{3.22}$$

where L is the length of referred previous trajectories and λ_1 is a constant value.

3.5.2 Court-line Distance Feature based Event Type Detection

In order to distinguish each event type from others, this paper uses the position and the velocity of the ball at the starting time of an event as the feature of ball event type detection. According to the volleyball game rules, each event occurs at certain area in the court which is divided by court-lines. Thus I use distances between each court-line and the position of the ball at the start time to build the feature vector. In addition, since features of difference event types are mainly reflected in the long side orientation and the height orientation, the court-line distance feature is defined as:

$$\Lambda = [Y_0, ..., Y_A, Z_0, ...Z_B, V_{y+}, V_{y-}, V_{z+}, V_{z-}]\,, \tag{3.23}$$

where $Y_a\,(a = 0, ...A)$ represents abstracted features in long side orientation and $Z_b\,(b = 0, ...B)$ is features in height orientation. For event state $\mathbf{d}_k$ at event time k, values of Y_a and Z_b are calculated as:

$$Y_a(\mathbf{d}_k) = dis(\mathbf{x}_{s_k}, l_a)\,, \tag{3.24}$$

$$Z_b(\mathbf{d}_k) = dis(\mathbf{x}_{s_k}, l_b)\,, \tag{3.25}$$

where, $dis(\mathbf{x}, l)$ is a function for calculating the distance from position $\mathbf{x}$ to line l. l_a represents the a_{th} court-line in the long side and l_b is the b_{th} court-line in the height orientation.

In equation (3.23), V_{y+}, V_{y-}, V_{z+} and V_{z-} are the values of the velocity $\dot{\mathbf{x}}_{s_k}$ in the forward direction of the long side, the backward direction of the long side, the forward direction of the height and the backward direction of the height of the court respectively.

Based on the proposed court-line distance feature, this paper using the SVM classification result as the likelihood of the event type. For each event type, some training samples are randomly selected to obtain the court-line distance feature vector and the training result is saved as preparation of the observation. In observation model of each particle $\mathbf{d}_k^{(i)}$, the court-line feature vector $\Lambda_k^{(i)}$ is extracted depending on $\mathbf{z}_k^{(i)}$ and compute the feature vector $\Lambda_k^{(i)}$ to the SVM [28] classifier of corresponding event type $e_k^{(i)}$. In this paper, the value of the event type likelihood $\mathbf{y}_{et}(\mathbf{d}_k^{(i)}; \mathbf{z}_{0:k-1})$ is assigned as the detection result of SVM.

3.6 Estimation

The target posterior distribution $p(\mathbb{X}_k \,|\, y_{1:k})$ of the ball state is approximated by particle filter algorithm according to the Monte Carlo method and a Bayesian equation [23]. Since the physical and conceptual state is difficult to estimate as an integrate one, the posterior distribution is a joint probability:

$$\begin{aligned} p(\mathbb{X}_k \,|\, y_{1:k}) =& p(e_k \,|\, \mathbf{z}_k, d_k, y_{1:k}) \\ & \times p(d_k \,|\, \mathbf{z}_k, y_{1:k}) \times p(\mathbf{z}_k \,|\, y_{1:k}) \,. \end{aligned} \tag{3.26}$$

The particle group is defined as $\mathbb{J}$.

First, for the physical state estimation, the normalized weight $\mathbf{W}_k^j$ of each particle can be computed according to their ball feature likelihood $p(y_k^{(j)} \,|\, \mathbf{z}_k^{(j)}; \mathbb{I}_k)$. After selecting the important particles with large weight, all the new particles $\mathbb{J}'$ can be obtained and devote their associated weights to approximating the ball trajectory state as

$$p(\mathbf{z}_k \,|\, y_{1:k}) \approx \sum_{j=1}^{J} \mathbf{W}_k^j \delta(\mathbf{z}_k - \mathbf{z}_k^j) \,, \tag{3.27}$$

where $\delta(\cdot)$ is the Dirac delta function and J is the particle number.

Second, flag variable of external force is estimated based on the resampled particle group $\mathbb{J}'$. The hitting likelihood $p(y_k^{(j)} \,|\, d_k^{(j)}, \mathbf{z}_k^{(j)}; \mathbb{I}, \hat{\mathbf{z}}_{k-L:k-1})$ of each particle is compared with an threshold TH. The number of particles whose hitting likelihood is larger than TH is defined as J_{abrupt}. If equation

$$J_{abrupt} > \alpha \cdot J \tag{3.28}$$

is satisfied (α is a constant parameter), $p(d_k \,|\, \mathbf{z}_k, y_{1:k})$ is estimated as external force, otherwise it is estimated as no-external force.

Third, when the external force is estimated, the ball event is estimated through choosing the event ID whose particle amount is the largest in the abrupt motion particles. After deciding the event ID, the event type including receive, set and pass can be detected as following. First, the first PASS for one team is the receive. Second, the second PASS before ATTACK is the set. At last if there is no ATTACK in one offense round, then all the detected PASS is assumed as pass. Attack is all the event passing the ball over net.

Depending on the theory that the starting time of $\mathbf{d}_n$ is the ending time of the event $\mathbf{d}_{n-1}$, the event result $\mathbf{r}_{n-1}$ of the $(n-1)_{th}$ event can be estimated once the event starting time $p(s^{(i)}_{1:k,n} \,|\, \mathbf{z}_{0:k-1})$ is decided.

The event result $\mathbf{r}_{n-1}$ can be judged basing on the trajectory during this event period $\mathbf{z}_{s_{n-1}:s_n}$. The evaluation criteria of each event result is described as follows:

(a) CROSS: Positions of the ball at time s_{n-1} and time s_n locate at each side of the net;

(b) RECEIVE: The position of the ball at time s_n is within the receive region defined manually through game rules;

(c) NET: During the Estimated event period, there are several successive positions of the ball are near the net area;

(d) GROUND: Distance from the ground to the position of the ball at time s_n is less than a given threshold;

(e) OUT: The position of the ball at time s_n is out of the court area.

3.7 Experimental Result

3.7.1 Introduction of Experimental Videos

The experiment is based on multi-view videos recorded during the final game of an official volleyball match (2014 Inter High school Men's Volleyball Games held in the Tokyo

Metropolitan Gymnasium in Aug. 2014), which contains three sets. These experimental game videos are captured by four cameras located at each corner of the court. View angles of the four input videos are shown in Fig. 3.9. Video resolution is 1920×1080, frame rate is 60 frames per second, and shutter speed of the cameras is 1000 frames per second. The latter parameter prevents motion blur in the video sequence. The recorded match contains 114 rounds and 606 ball events totally, which covers all the basic event scenes including serve, pass (receive, toss) and spike in difference situations and player by different players.

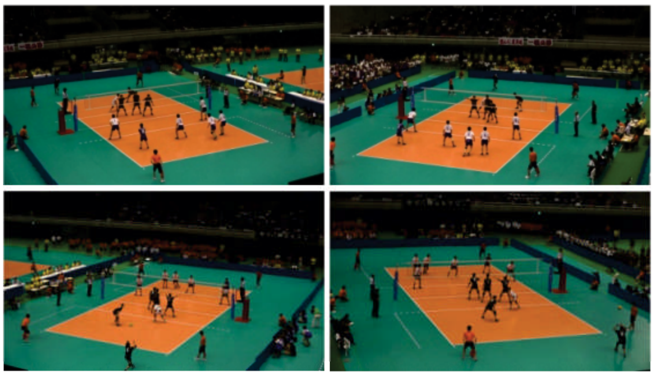


Figure 3.9: View angles of four input videos for the experiment.

3.7.2 Implementation Details

The entire algorithm of the ball state based ball tracking and event detection method is implemented with OpenCV and GSL library.

The particle number M utilized for event detection is assigned as 5000; In the observation model, λ_1 in equation (3.20) are 5×10^{-6}, and the length of refereed future trajectory L is 3; In the estimation part, the court-lines I used to obtain the feature by equation (3.24) and (3.25) are the two 9 meters sides line, two 3 meters lines and the center line of the court, and the ground line and the up side line of the net in the height orientation.

3.7.3 Evaluation Result

With the same sequence and same evaluation method, the tracking success rate keep at the same 99.42% with that shown in chapter 1. The output of the tracking process is a sequence of 3D positions and velocities. At each time step, the approximated 3D ball position and velocity will be sent to the event detection process as observation.

The performance of the event type detection is the basement of the event likelihood observation. Before evaluating the performance of ball event detection, I implement the court-line feature based event type detection independently to evaluate the performance of this idea.

3.7.3.1 Evaluation of the court-line distance feature

To evaluate the proposed court-line feature, depending on the tracked 3D ball trajectory data, I prepared some samples belonging to each event type to test the detection accuracy. In the prepared event type samples, the amount of the ”SERVE+”, ”SERVE-”, ”PASS+”,”PASS-”,”SPIKE+” and ”SPIKE-” were 59, 48, 152, 146, 34 and 61, respectively.

Table 3.3: Evaluation criteria [2] for event type detection.

Input \ Output	**Positive**	**Negative**
Positive	TP	FN
Negative	FP	TN

Three standard evaluation criteria used here are defined in [2], the accuracy, recall and precision, to evaluate the proposed court-line distance feature. Definitions of used criteria are:

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN}, \tag{3.29}$$

$$recall = \frac{TP}{TP + FN}, \tag{3.30}$$

$$precision = \frac{TP}{TP + FP}. \tag{3.31}$$

Because of the amount limitation of test samples and unbalance amounts of positive and negative samples for each event type, the experimental is obtained by the following process.

For each test event type, I randomly chose 10 samples as positive templates and choose 30 samples from other event type samples to make negative samples for the SVM classifier to do data training. Then, 20 samples of the test event type and 20 samples of other event types are chosen randomly from the left samples as the test data for SVM classifier. Through this way, 40 experimental results can be obtained. In the experiment, this process is repeated 10 times and total 400 experimental results are obtained, which are shown in the Table 3.3. From the results shown in Table 3.3, it can be known that the detection accuracy for event SPIKE and SERVE are close to 100% and the accuracy of PASS is also larger than 97%.

To evaluate the performance of the proposed event type detection for real-time process, another group of experiment was implemented. In this experiment, the SVM classifier was trained with event type samples randomly selected from another volleyball game

Table 3.4: Detection results of the court-line distance feature based event type detection.

	Recall	Precision	Accuracy
SERVE+	100.0%	99.0%	99.5%
SERVE-	100.0%	99.5%	99.8%
PASS+	96.5%	98.5%	97.5%
PASS-	98.5%	97.5%	98.0%
SPIKE+	100.0%	99.0%	99.5%
SPIKE-	100.0%	99.0%	99.5%

(the semifinals of 2014 Inter High School Men's Volleyball Games), which were played by different teams. With the same test data used in Table 3.4, the new experimental results of SVM classifier are shown in Table 3.5. By comparing the data in Table 3.4 and Table 3.5, it can be known that there is almost no influence on event type detection performance by using different game data samples.

3.7.4 Evaluate Result of the Ball Event Detection

Both the physical and conceptual ball state is estimated by event-unit because the true value of physical state is impossible to obtain.

During one event-unit, if the ball trajectory is a continuous curve without sharp gap and all the projected 2D image coordinates are located within the ball's region in image frames, the ball trajectory of this event-unit tracking is judged as successful one.

Based on the definition of the evaluation method, the experimental results of the proposed tracking algorithm employed on test videos are shown in Table 3.6.

Table 3.5: Detection results of the court-line distance feature based event type detection using training data from other games.

	Recall	**Precision**	**Accuracy**
SERVE+	97.5%	99.5%	98.5%
SERVE-	100.0%	100.0%	100.0%
PASS+	97.0%	96.5%	96.8%
PASS-	100.0%	95.2%	97.5%
SPIKE+	100.0%	100.0%	100.0%
SPIKE-	100.0%	100.0%	100.0%

Table 3.6: Experimental results of the proposals.

	Physical state	**Flag variable of external force**	**Conceptual state**
Success event-units	577	543	537
Total event-units	581	581	581
Success rate	99.3%	93.5%	92.4%

In order to show the tracking result visually, I chose one test sequence and plotted the tracked 3D trajectory in Fig. 3.10. The trajectory is plotted by different markers and colors once the conceptual state changes. Also from the video frame view, the trajectory of each event is plotted on the frame. Fig. 3.11 is an example of serve, Fig. 3.12 and 3.13 is an example of pass, and Fig. 3.14 is the spike.

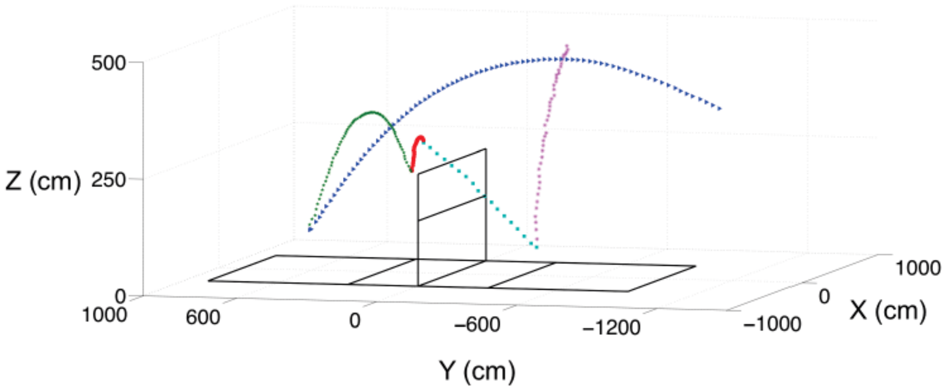


Figure 3.10: The 3D trajectory of the tracking result.

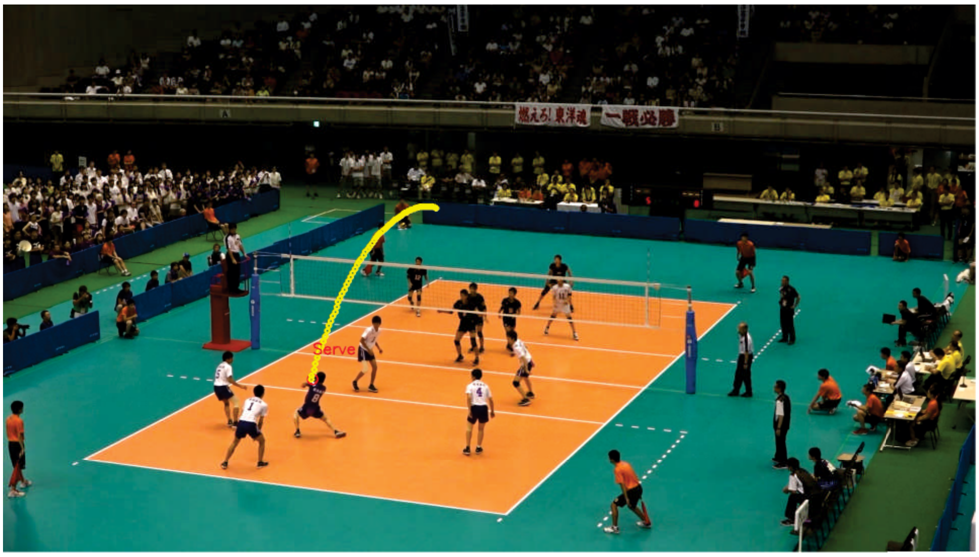


Figure 3.11: The trajectory of the SERVE+ in image.

3.8 Conclusions

With the target of the event data acquisition in volleyball game for the 3D sports analysis, this chapter proposes a ball trajectory based event detection method based on particle filter framework, which using the game videos and physical ball trajectory as the input. For

Figure 3.12: The trajectory of the 1st PASS+ in image.

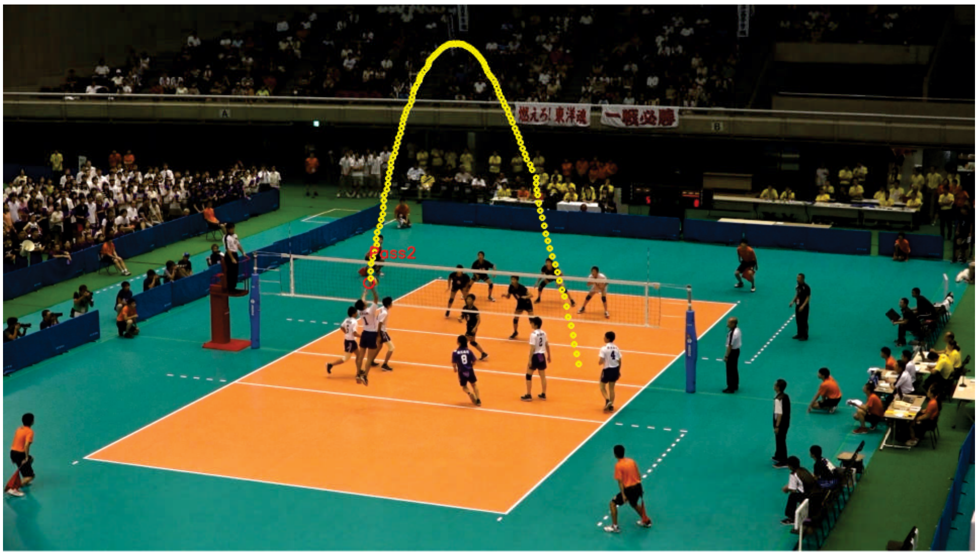


Figure 3.13: The trajectory of the 2nd PASS+ in image.

the ball event detection, this chapter proposes an event change distribution based system model, a trajectory abruptness and skin color area based event change likelihood and a

Figure 3.14: The trajectory of the SPIKE- in image.

court-line distance feature based event type detection. Firstly, event states of all the particles consisting of the event change flag variable and the event type are transited through the external force adaptive system model. The probability distribution of event change is approximated according to the fact that the probabilities of event change in different spatial locations are different, such as in the spike and block areas the probability becomes large. Based on this, event changes can be accurately predicted. Secondly, in the observation, the trajectory abruptness and skin color area based event change likelihood is proposed to deal with the noise in ball trajectory for evaluation of the event change. The influence of unstable tracking ball trajectory can be reduced by using information of neighbouring trajectory segments. When players' hands are close to the ball, the ball is more likely to be hit. So the area of players' skin in the region around the ball is calculated as the likelihood of event change. With the image feature, the event change feature which are calculated dependent from the ball trajectory contributes a lot in accurately evaluation for observation. Thirdly, for the event type detection in observation, the court-line distance feature is proposed including the relative position feature and the ball motion information. Experimental results show that the event type detection accuracy of court-line distance feature is more that 97.5% and the integrate event detection accuracy by proposed particle filter framework achieves 92.4%.

The future target will dive into two aspects. First, combining current work with other sports analysis technologies such as player tracking and motion analysis, I achieve at a comprehensive volleyball game analysis and tactics development system. Secondly, my final goal is a real-time product for supporting TV content broadcasting in Tokyo Olympic Games. Based on the achievements of the GPU real-time acceleration for 3D ball tracking [29], the GPU implementation of the whole ball tracking and event detection is the next important target.

Chapter 4

Relative Motion Abruptness and Court Zone Division based Player Role Detection for Efficacy Variables Acquisition

4.1 Introduction

Automatic game strategy data acquisition is important for the realization of the professional strategy analysis systems by providing the evaluation value such as the team status and the efficacy of plays. As Fig. 4.1 shows, the ball strategy data acquisition is based on the ball physical data obtained by the ball physical data acquisition, the ball event data obtained by the event data acquisition and the players' positions. Outputs of the strategy data acquisition are values that decides the attack efficacy in attack play, such as the setting zone, number of available attackers, attack tempo, number of blockers and attack type. The key factor that influences the performance in the volleyball game strategy data acquisition is the unknown player roles, which loses the contribution of each player to the play and makes the information of individual player unreliable for the player performance evaluation.

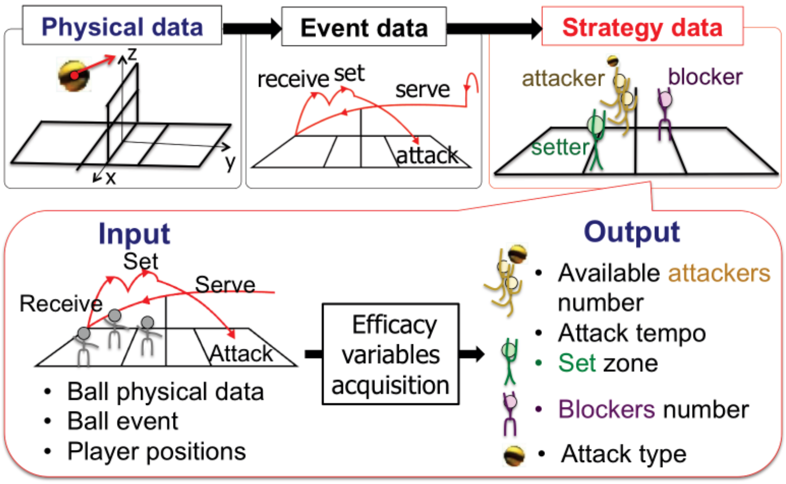


Figure 4.1: Introduction of the ball strategy acquisition and whose position of in entire research.

Among all the volleyball game plays including the serve, reception, set, attack, block, dig, and pursuit, the attack is the one presenting the higher correlation with success, independent of the game phase. [30] Besides the physical and event information, there are two sufficient elements for the attack play evaluation: team tactical status and attack efficacy variables.

Firstly, the attack efficacy variables help to understand how the point is achieved in the attack. Taking the previous play (set) and the following ones (block or receive) into consideration, the basic attack efficacy variables are defined as the setting zone, available attacker number, attack tempo, attack type and blockers number. Descriptions of these variables are in Table 4.1 and Fig. 4.2 . All the variables not only make effort to the evaluation of attack efficacy but also reflect the teamwork performance in attack. For example, when the attack play starts that means the setter has played, the teamwork of attack team can be evaluated through the value of setting zone, available attacker number and attack tempo. And the blockers number can be used for evaluate the reaction ability of defence.

Table 4.1: Description of attack efficacy variables.

Variables	Description
Setting zone	Zone where the previous touch before attack is executed.
Available attackers number	Number of attackers available to attack in the moment of setting.
Attack tempo	Moment when the attacker starts the approach to attack.
Attack type	Speed or power of the attack.
Blockers numbers	Number of blockers opposing the attack.

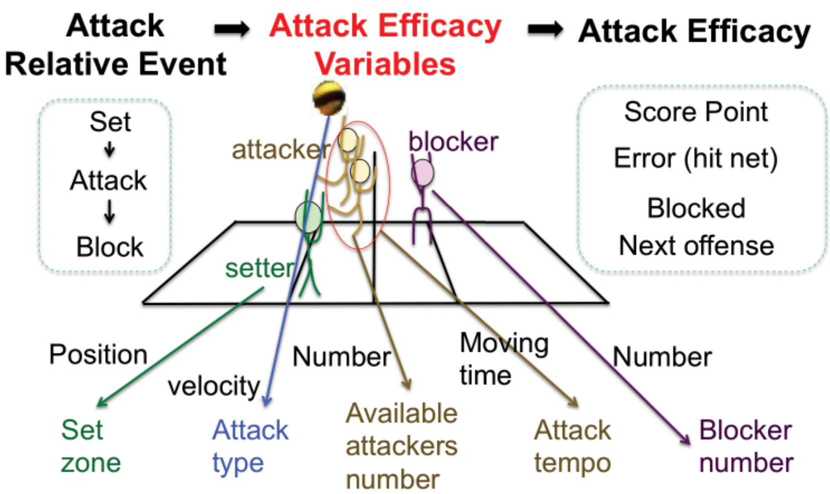


Figure 4.2: Concepts of attack efficacy variables.

Secondly, the team tactical status plays important roles in the teamwork evaluation, such as the tactical status of the attacking team reflects the attack quality and the tactical status of the rival team represents the ability of defense reaction. According to the volleyball game rules and tactics, the tactical status is classified into four types: the defensive ready, the defensive, the offensive ready and the attack. Descriptions of each tactical status are listed in Table 4.2 and the transition relation between all the team status is shown in Fig. 4.3. Among these team tactical statuses, the attack status is the most effective status for the attacking team and the defensive status is the most effective one for the rival team. For example, when the attack play starts, the teamwork of attacking team can be evaluated through whether the team status has transited from offensive ready to the attack status.

Table 4.2: Description of team tactical status.

Status	Description
Defensive Ready	Players change positions from other status according to the rival's attack.
Defensive	The players move to receive or block the ball.
Offensive Ready	Players change the formation to standing by for attack.
Attack	The team spike or pass the ball to the rival.

For the strategy data analysis, the Data Volley system is widely used. However, its input is performed manually by visual observation, which is inaccurate and requires much manually labelling. Therefore, there is a demand for a technology that automatically and accurately acquires the strategy data. It is necessary to obtain the player role only from the data acquired from the video, such as the event data and the physical data of the player and the ball. Without the manually labelling of the role for each player (such as the attacker,

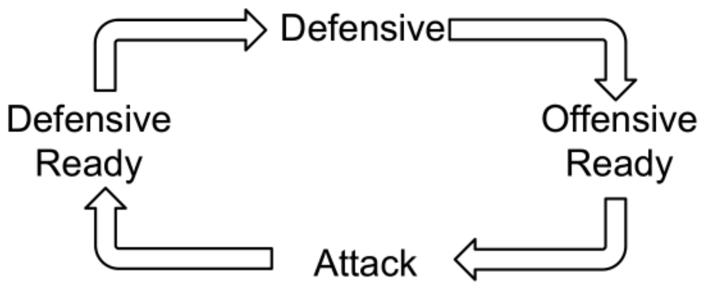


Figure 4.3: Transition relationship between four team tactical statuses.

blocker, setter), even the physical and event data of the team is obtained, it is still difficult to analysis the contribution of each player to the play without any information of player roles. With out the player roles, the local position and motion information of individual

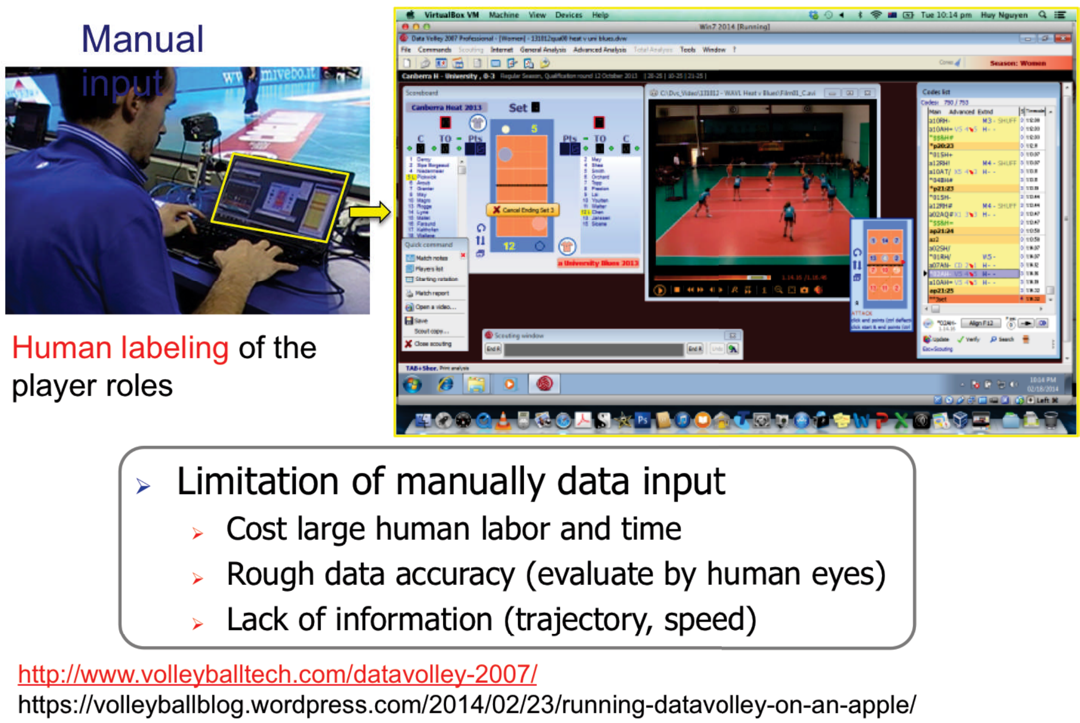


Figure 4.4: Data Volley based player roles identification in strategy data acquisition.

player is meanness in volleyball strategy data analysis. Once a team gains the serve, all the players (no matter what roles they play) must rotate in a clockwise direction, so during one same game tactical status, the motions and positions of each player are different in a different rotations. In addition, different team formations, which are made by evaluation the players' ability and the cooperation of teamwork tactically, also results in different player performances and features in same team tactical status.

As for the acquisition of the attack efficacy variables, not only the player who attacks the ball but also the setter, other available attacker, and the blockers need to be identified. Bertini [31] has proposed a jersey number detection and face recognition based player identification system. However, this work is still not suitable for the attack efficacy variables acquisition in the volleyball game. For the team tactics status detection, Atmosukarto [32] has proposed an automatic recognition of offense team formation in American football plays, which uses the estimated court line of scrimmage to extract the team formation for detection. Since this work is implemented on the 2D simple view videos, there is a large limitation on the view angle change.

4.2 Framework of Attack Efficacy Variables Acquisition

The overall framework of the automatic attack efficacy variables acquisition is shown as Fig. 4.5. Input of the framework is multi-view game videos, which record the games from different views synchronously. With the multi-view videos, not only the image features but also the 3D spatial information can be used. There are two steps of data acquisition from the game videos. First is the acquisition of the 3D physical and event data, such as the 3D ball position, ball event and the players' positions. This paper uses my past work of 3D ball state estimation method [33] and multiple players tracking algorithm [34] to obtain physical and event data. Second is the strategy data acquisition based on the obtained data. Here the player role detection is generated to acquired the attack efficacy variables and the area team motion density feature is proposed for detection of the team status.

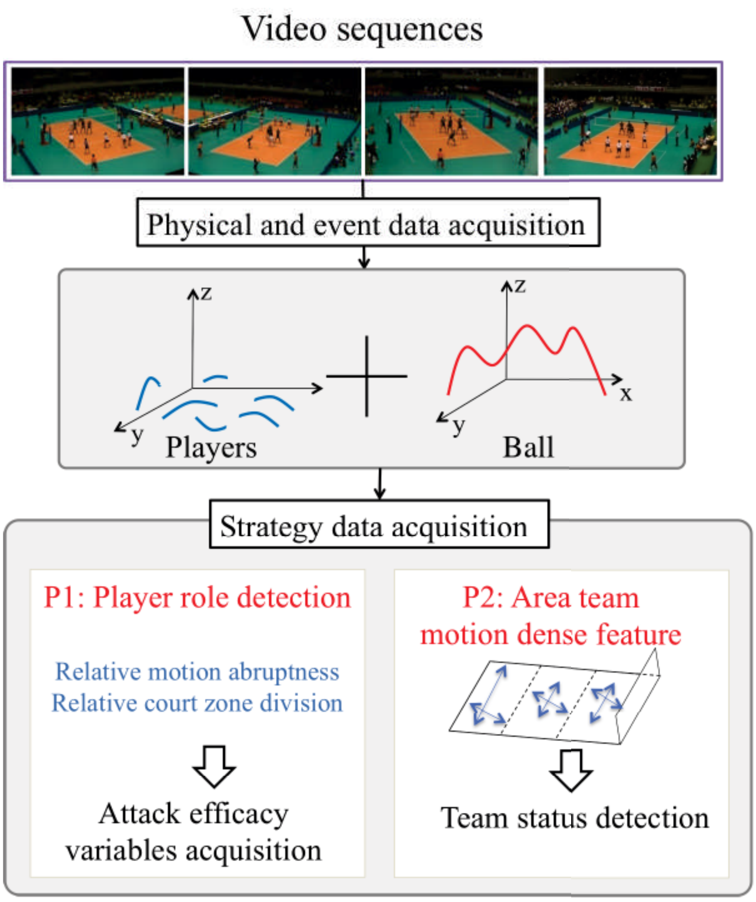


Figure 4.5: The overall framework of the automatic attack efficacy variables acquisition.

Since the 3D ball trajectory, ball event and players' positions are the basement of further data acquisition for team play evaluation. Here, the particle filter [23] based 3D ball physical and event state estimation is introduce in chapter 2 and chapter 3. At each sampling time, the estimated state $\hat{\mathbf{X}}_{0:k}$ by particle filter is the tracked 3D ball trajectory and event. And the particle filter based multiple players tracking method [34] is introduced is introduced below.

As for the acquisition of the player position, the jersey number $c \in C$ is used to manage multiple targets in one team. C is the set of all jersey numbers. Then each player's state is defined as $\mathbf{y}_k^c$ at a discrete time k, which represents the 3D position in real 3D space. And the state is transited as

$$\mathbf{y}_k^c = \mathbf{y}_{k-1}^c + W_k \cdot \Delta T \,, \tag{4.1}$$

where W_k is the prediction model combined with a Gaussian model and a least square fitting prediction model. Observation of each player includes jersey number likelihood, color likelihood and sobel gradient likelihood. The estimated player positions are indicated as $\hat{\mathbf{y}}_{0:k}^c$.

Based on the obtained physical and event data of the ball and player, the relative motion abruptness and court area division based player role detection method is proposed for the acquisition of attack efficacy variables. And the spatial team motion feature is proposed for the team status detection.

4.2.1 Relative Motion Abruptness and Court Area Division Feature

As the definition of attack efficacy variables, the players whose motion affect the attack play include not only the player attacking the ball but also the setter, the other available attackers, and the blockers. In order to acquire these attack efficacy variables, the role of each player related to the attack play is required. There are three features of player performance being proposed for distinguishing different players' roles. First is the relative court area division feature. For each player, this feature are presented by the ball relative distance value $f_1(k)$ during the time period from the setting and blocking, which is for the detection of the players who directly touch the ball, such as the setter and the attacker. Second is the relative motion abruptness feature, which are presented by the value of ball approach motion abruptness $f_2(k)$ during the time period from the setting and blocking. This feature is proposed for the detection of the players who don't touch the ball, such as the blockers and other available attackers. In addition, the attack motion feature $f_3(k)$ is also proposed for the detection of players who don't touch the ball. Detail definitions of these player roles features are demonstrated as the following subsection.

4.2.1.1 Relative Court Area Division Feature

The definition of $f_1(k)$ is the distance to the ball of each player during the setting-attack period,

$$f_1(k) = |\hat{\mathbf{y}}^c_{0:k} - \hat{\mathbf{x}}_k|, k \in [k_{set}, k_{attack}] \,, \tag{4.2}$$

where k_{set} and k_{attack} is the starting time of event set and attack, which are obtained from the ball event data $\hat{d}_{0:k}$ and $\hat{e}_{0:k}$. Based on the distance from the ball at the time of each event happening, the court area can be divided into different role area. At the setting time k_{set}, the 3D court area near the ball (that means the distance to the ball is small) is the setter area. In a similar way, the attacker area and the blocker area are defined by the distance to the ball at the attack time and the block time.

4.2.1.2 Relative Motion Abruptness feature

The relative motion abruptness feature shows the motivation of the player moving to hit the ball, which is defined by the possible minimum distance between the player and the

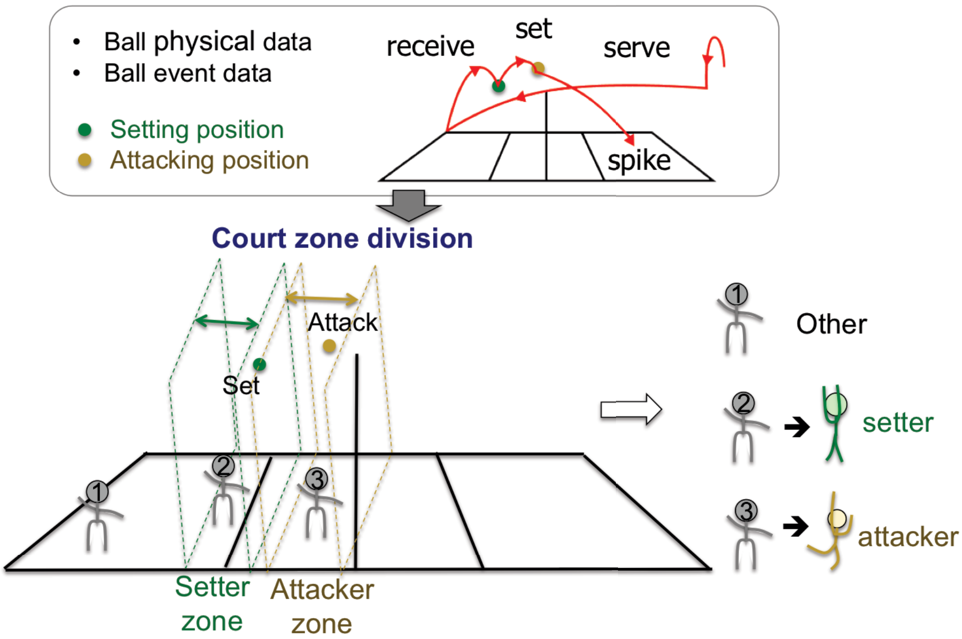


Figure 4.6: The relative court area division feature.

ball around the attack play dis_k,

$$dis_k = min(dis^q_{k,p}) \,, \tag{4.3}$$

The $dis^q_{k,p}$ is a predicted distance between the ball and the player in the attack play, which is calculate by:

$$dis^q_{k,p} = |\hat{\mathbf{y}}^c_{0:k} + \dot{\mathbf{y}}^c_{p:k} \cdot (q-k) \cdot \Delta T, \hat{\mathbf{x}}_q| \,. \tag{4.4}$$

q is the time when the ball is hit. For attack play, $q = k_{attack}$. $\dot{\mathbf{y}}^c_{p:k}$ is the velocity of the player estimated by the positions at time p and time k:

$$\dot{\mathbf{y}}^c_{p:k} = \frac{(\hat{\mathbf{y}}^c_{0:k} - \hat{\mathbf{y}}^c_{0:p})}{(k-p) \cdot \Delta T} \,, \; p \in [k-3, k-1] \,. \tag{4.5}$$

With the distance between the player and the ball, the probability of the player moving to attack the ball can be evaluated so that the available attackers are detected. When the distance dis_k becomes small, that means the player moves towards to ball. The motivation of this player playing the ball is large can be judged. Therefore, the ball approach player curve $f_2(k)$ is defined as

$$f_2(k) = e^{dis_k \cdot \lambda}, p \in [k_{set}, k_{attack}]. \tag{4.6}$$

With this relative motion abruptness feature, the available attackers can be detected.

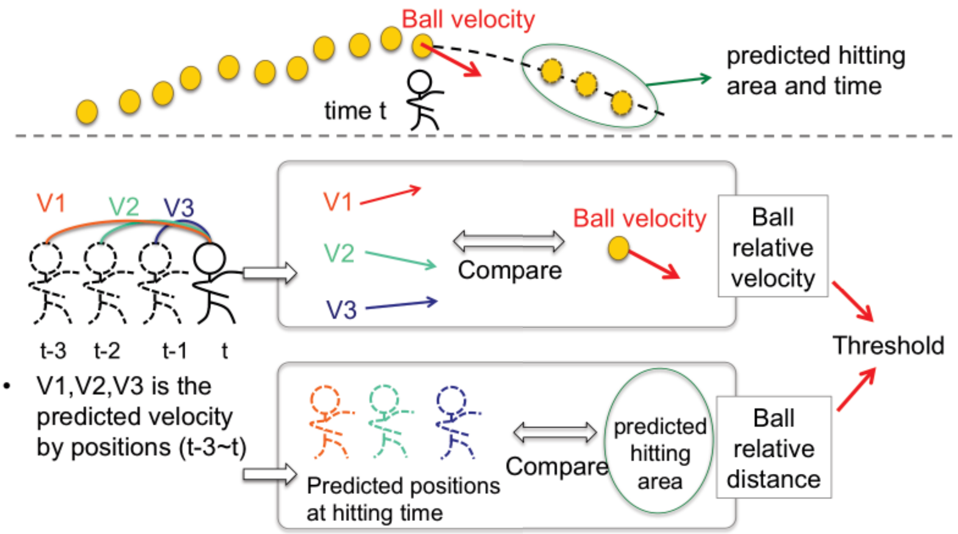


Figure 4.7: The relative motion abruptness feature.

4.2.1.3 Attack Motion Feature Curve

The attack motion feature represents the motion feature of the available attackers and blockers during the set and attack event. When the setters play the ball, the available attackers and the blockers move to the ball then jump to gaining of defencing score based on the ball movement. Then based on the ball movement, the available attackers move and jump to spike the ball for gaining score. Movements of the attackers are mainly performed in the velocity direction to the ball and the height variance. Therefore, this two information are used for obtaining the attack motion feature value $f_3(k)$ in the time period from setting to blocking. Definition (x_k, y_k, z_k) is the 3D coordinate of $\hat{\mathbf{y}}^c_{0:k}$ in 3D coordinate space, so that the feature value $f_3(k)$ is defined as

$$f_3(k) = z^c_k \cdot v^c_k \,, p \in [k_{set}, k_{attack}] \,, \tag{4.7}$$

where v_k is the relative velocity between the player to the ball attacking position. During the time period from setting to blocking, the peak value of the attack motion feature denotes the attack motion of the player. Based on the time and the position of the attack motion, the available attackers and blockers can be detected.

4.2.2 Attack Efficacy Variables Acquisition

Based on the attack efficacy curve of each player, the acquisition of each attack efficacy variables is introduced in this section. Here, the value of each attack efficacy variable including setting zone, available attacker number, attack tempo, attack type, blockers number is denoted SZ, AN, A_{tempo}, A_{type} and BN [30]. Values of each attack efficacy variables are defined in Table 4.3.

4.2.2.1 Acquisition of Setting Zone

The setting zone is defined as *excellent setting zone* (ESZ), *acceptable setting zone* (ASZ) and *not acceptable setting zone* (PSZ) [35], whose court area is shown in Fig. 4.8. The ESZ is an area of $8m^2$, $2m$ deep from the net and $4m$ wide, at a distance of 2m from the right sideline and $3m$ from the left side. The ASZ is an area of $6m^2$, 2m deep from the

ESZ zone and $3m$ wide, at a distance of $2m$ from the right sideline and $4m$ from the left sideline. And the PSZ is court area excluding the area of the ESZ and the ASZ.

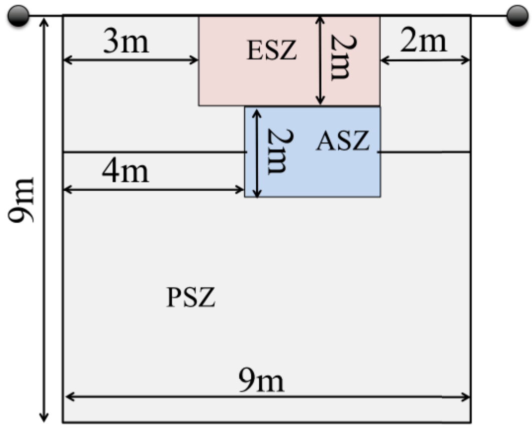


Figure 4.8: Definition of the setting zone. ESZ: *excellent setting zone*, ASZ: *acceptable setting zone*, PSZ: *not acceptable setting zone*

The position $\mathbf{y}_{set}$ of the setter is

$$\mathbf{y}_{set} = \hat{\mathbf{y}}^{c_{set}}_{0:k_{set}}, \tag{4.8}$$

where, $\hat{\mathbf{y}}^{c_{set}}_{0:k_{set}}$ is the position of the player c_{set} at the set time k_{set}. The player number c_{set} makes the value $f^c_1(k_{attack})$ smallest when $c = c_{set}$.

Then the value can be judged by comparing the position $\mathbf{y}_{set} = (x_{set}, y_{set}, z_{set})$ with the border line of the defined setting zones:

$$\begin{cases} SZ = ESZ, & (Line_1 < x_{set} < Line_2) \cap |(y_{set}| < 2m) \\ SZ = ASZ, & (Line_3 < x_{set} < Line_1) \cap (2m \leq |y_{set}| < 4m) \\ SZ = PSZ, & other\ situation \end{cases}, \tag{4.9}$$

where $Line_1$, $Line_2$ and $Line_3$ is the threshold of the setting zone.

4.2.2.2 Acquisition of Available Attackers Number

In the attack play, the available attacker can be detected by three conditions, which are defined by the role of attackers in the game. Firstly, at the attacking time, the attacker should have prepared at the attack zone which is defined as the area from the 3-meters line of the court to the net. That means:

$$0 < |y^c_{attack}| < 3\,. \tag{4.10}$$

Secondly, in order to cooperate with the setting, the available attacker should move approach to the ball. This feature is reflected on the relative motion abruptness feature $f_2(k)$. In the basic volleyball tactics, there are at least one, at most four available attackers in one team. For each player c, if the relative motion abruptness feature value $f_2^c(k)$ satisfies the following condition:

$$max(f_2^c(k)) > \Lambda_1\,, \tag{4.11}$$

then this player c can is decided as one candidate of available attackers. Λ_1 is the threshold.

At last, the available attacker must have attack motion before the attack time, whose feature is evaluated by the attack motion feature curve $f_3(k)$. Therefore, for the candidate available attacker c, the third condition is that there must be a peak in the spike motion curve $f_3^c(k)$. The mathematic expression of this condition is with threshold Λ_2:

$$max(f_3^c(k)) > \Lambda_2\,. \tag{4.12}$$

If the data of player c satisfy these three conditions, this player is judged as the available attacker. The value of AN is the total number of the detected available attackers.

4.2.2.3 Acquisition of Attack Tempo

As the definition in Table 4.3, the value of attack tempo is judged by comparing the time of setting k_{set} and the time k_s when the attacker moves to attack the ball as:

$$\begin{cases} A_{tempo} = 0, & k_{set} - k_s > L_{tempo} \\ A_{tempo} = 1, & |k_{set} - k_s| \leq L_{tempo} \\ A_{tempo} = 2, & k_s - k_{set} > L_{tempo} \end{cases}, \tag{4.13}$$

where the L_{tempo} is a value denoting the average length of setting play. k_s is decided when the time k first time satisfies

$$f_3^c(k) < min(f_3^c(k+1), f_3^c(k+2), f_3^c(k+3)) \tag{4.14}$$

before the attack time. The value of k_{set} is obtained by the relative motion abruptness feature $f_2(k)$ and the attack motion feature $f_3^c(k)$. When the attacker moves to the ball, the value of $f_2(k)$ and $f_3^c(k)$ increase dramatically.

4.2.2.4 Acquisition of Attack Type

The attack type is evaluated be the ball velocity after attack time directly. Taking the tracking noise of the ball into consideration, the average velocity of k_L time after attack is used:

$$V_{ball} = \frac{1}{M} \sum_{k_L=0}^{M-1} \hat{\mathbf{x}}_{k+k_L}, \tag{4.15}$$

where M is the number of past trajectories. Therefore, the value of A_{type} is decided with the threshold TH:

$$\begin{cases} A_{type} = 1, & V_{ball} \leq TH \\ A_{type} = 2, & V_{ball} > TH \end{cases}, \tag{4.16}$$

4.2.2.5 Acquisition of Blocker Number

The block is the play of the rival team after attack. The blockers move near the net and jump to block the ball. There are three criteria for judging whether one player is the blocker. First, position of the player $\mathbf{y}_k^c = (x_k^c, y_k^c, z_k^c)$ should be at the rival team area when attack start.

$$y^c_{k_{attack}} \cdot y^{k_{attack}}_{c_{attack}} < 0\,, \tag{4.17}$$

where, k_{attack} is the time of attack play starts and c_{attack} is any jersey number of the attack team. Second, in the attack play, the distance between the ball and the blocker should be small enough so that it is possible for them to block the attack. Therefore, the second criterion is based on the relative court area division feature $f_1(k)$ with the threshold Λ_3.

$$f_1^c(k_{attack}) < \Lambda_3\,. \tag{4.18}$$

At last, since the blockers always jump up to block the ball, the attack motion feature $f_3(k)$ is used as the third criterion:

$$f_3^c(k_{attack}) > \Lambda_4\,. \tag{4.19}$$

When one player's data satisfy the above three criteria, the player is judged as blocker of this attack play. The value of BN is the total number of the players who are judged as blockers.

4.2.3 Area Team Motion Density based Team Tactical Status Detection

The feature of the team tactical status utilizes two information, the ball relative team motion and the team motion density on the specific court area. The spatial team features

F_i is the sum of each players' spatial ball relative motion F_i^c:

$$F_i = \sum_{c=0}^{C-1} F_i^c, \quad i \in [0, I), \tag{4.20}$$

where c is the jersey number of each player and the integrate value i is the index of the spatial team feature vector whose length is I. The feature vector length I is defined by the product of the amount of ball relative motion directions A, the amount of the specific court areas B and the three dimension of physical 3D space:

$$I = A \times B \times 3\,. \tag{4.21}$$

First, the ball relative motion at each physical direction is defined as a 4 dimensional feature. As equation (4.22) shows, the four dimensions are defined when the position of the ball and the player is in the same sides of the court, in the different side of the court, the moving direction of the ball is same with the player and the moving direction is opposite with the player.

$$\begin{cases} \hat{\mathbf{y}}^c_{0:k} \cdot \hat{\mathbf{x}}_k < 0 \cap \hat{\dot{\mathbf{y}}}^c_{0:k} \cdot \hat{\dot{\mathbf{x}}}_k < 0, & \alpha = 0 \\ \hat{\mathbf{y}}^c_{0:k} \cdot \hat{\mathbf{x}}_k < 0 \cap \hat{\dot{\mathbf{y}}}^c_{0:k} \cdot \hat{\dot{\mathbf{x}}}_k > 0, & \alpha = 1 \\ \hat{\mathbf{y}}^c_{0:k} \cdot \hat{\mathbf{x}}_k > 0 \cap \hat{\dot{\mathbf{y}}}^c_{0:k} \cdot \hat{\dot{\mathbf{x}}}_k < 0, & \alpha = 2 \\ \hat{\mathbf{y}}^c_{0:k} \cdot \hat{\mathbf{x}}_k > 0 \cap \hat{\dot{\mathbf{y}}}^c_{0:k} \cdot \hat{\dot{\mathbf{x}}}_k > 0, & \alpha = 3 \end{cases}, \tag{4.22}$$

where α is the serial number of ball relative direction.

Then, value of specific area amount B is 3 based on the segmentation of the court area, which is decided by the distance from the player to the net. The number of each area noted with β.

$$\begin{cases} |\hat{\mathbf{y}}^c_{0:k}| < 3, & \beta = 0 \\ 3 \le |\hat{\mathbf{y}}^c_{0:k}| \le 6, & \beta = 1 \\ 6 < |\hat{\mathbf{y}}^c_{0:k}|, & \beta = 2 \end{cases}, \tag{4.23}$$

Therefore, the ball relative spatial motion feature of each player F_i^c is calculated as:

$$F_i^c = \Delta \hat{\mathbf{y}}^c_{k-L:k}\,. \tag{4.24}$$

In equation (4.24), L is the length of trajectory used for extracting the feature and $\Delta\hat{\mathbf{y}}^c_{k-L:k}$ is the moving distance during frame k and frame $k-L$.

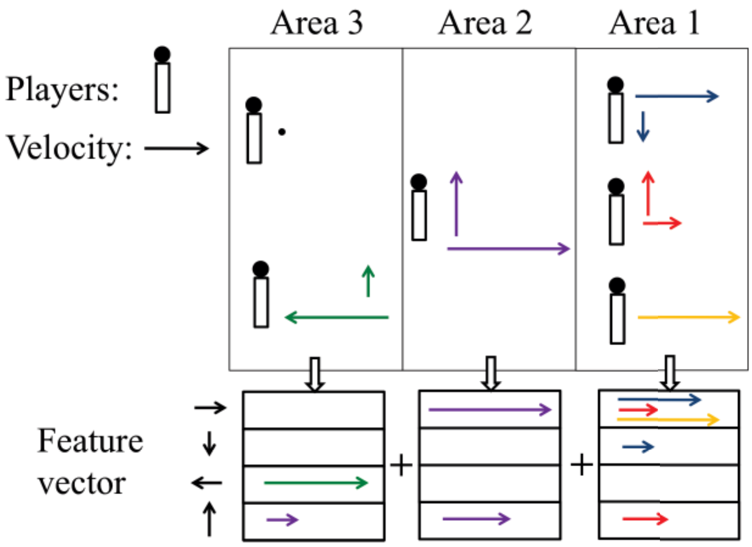


Figure 4.9: A conceptual example of the generation process for area team motion density feature.

Fig. 4.9 gives a conceptual example of the generation process of the proposed area team motion density feature. Here are six players with known positions and velocities in the court. These six players can be classified into three groups according to the court area they locate. In area 1 of the court, there are three players and their ball related velocity can be decomposed into four directions (in 3D space, it should be decomposed into six directions). The team motion density feature of this area is a four dimensional vector, each demotes different motion direction. Value of each direction is the sum of all the players' motion in this direction as shown in the figure. The features of all the three areas are the proposed area team motion density feature, which is used for the detection of the team status detection.

4.3 Experimental Results and Discussions

4.3.1 Experimental Datasets and Environment Settings

Performance of the automatic game strategy data acquisition is evaluated on multi-view videos of an official volleyball match (2014 Inter High school Men's Volleyball Games held in the Tokyo Metropolitan Gymnasium in Aug. 2014). These experimental game videos are captured by four cameras located at each corner of the court. View angles of the four input videos and the projection relationship with the physical coordinate system are shown in Fig. 4.10. The match records all kinds of team tactical status including copious volleyball events, such as serve, set, attack and block. After employing out previous methods of the ball and players state estimation, whose success rate is 99.4% for 3D ball tracking, 92.4% in ball event estimation and 99.87% for 3D multiple player tracking. These 3D physical and event data are used as the input of the strategy data acquisition system. Environment of the implementation of the proposed algorithms is C++ language and OpenCV 2.4.11 on a machine of 3.40GHz CPU and 8 GM RAM.

4.3.2 Experiments of Attack Efficacy Variables Acquisition

The experiments choose the attack plays of one team as the test sequence whose ball and players can be tracked successfully. To evaluate the performance of the tested team, there is 25 effective attack plays being prepared for evaluating the attack ability, and 27 effective block plays prepared for evaluation of the team defense ability. As for the implementation details, The comparison value in equation (4.9) which are defined according to the setting zone definition and which half court the team locates:

$$\begin{cases} Line_1 = -1, Line_2 = 2, Line_3 = -2, & y < 0 \\ Line_1 = -2, Line_2 = 1, Line_3 = 2, & y > 0 \end{cases}, \tag{4.25}$$

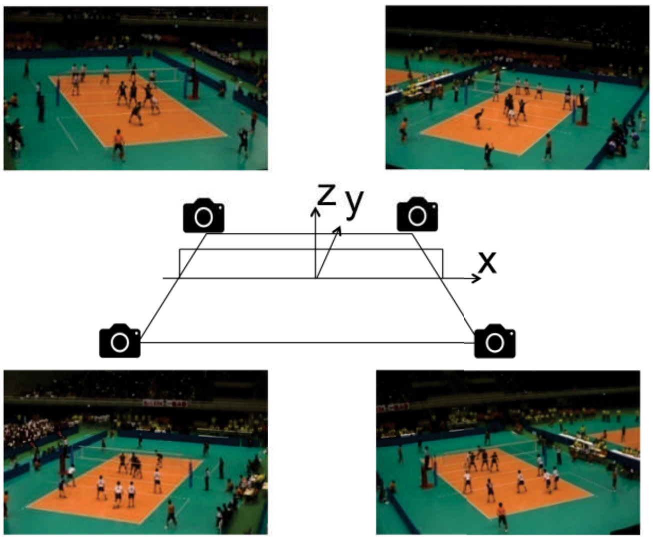


Figure 4.10: View angles of input volleyball game videos.

where y is the team location in y direction, whose unit of the value is meter. Values of Λ_1, Λ_2, Λ_3 and Λ_4 in equation (4.11), (4.12), (4.18) and (4.19) are 0.5, 1.8, 1 and 1.7. The value of λ in equation (4.6) is -0.02, L_{tempo} in equation (4.13) and M in equation (4.15) is 3. The value of TH in equation (4.16) for attack type acquisition is a special case that is discussed in the next section.

Three kinds of data are shown for the comparison of the proposed player performance curve. First one is the ground truth of the request data, which are acquired by one person who one person watches the game (four views) any times he needs with referring the physical data and judge the data carefully. The second method is the conventional method, which are acquired by one person who watches each view (four views) of the game and real-time labels manually before the video finished. The last one is the acquired data of proposed automatic attack efficacy variables acquisition method, which obtains the data from the video directly without human judgment. Table 4.4 shows the evaluation data for the acquisition of the setting zone, the number of available attackers, the attack tempo and the number of blockers. The proposed automatic data acquisition method has a robust performance, which achieves 100% precision in the acquisition of the setting zone, the number of available attackers and the number. The compared acquisition method by human eyes is weak at judgment of the physical items. Such as the acquisition of the setting zone, it is difficult to evaluate the player position when the setter locates at the border of two setting area. Thanks to the accurate physical data of the 3D player position, the setting zone can be corrected detected with proposed player performance curve. The only one error acquisition is the acquisition of the attack tempo, there are two reasons affect the acquisition performance: one is the unstable player trajectory and the other is the ambiguous player motion.

As an exception in the evaluation, the criterion of attack type is a special case since there is no definite explanation of the ”strong and powerful attack” and the ”slower attack”. Therefore, the only method to obtain the ground truth is referring the conventional manual method. However, the attack type is the variable that evaluates the velocity of the ball and the power of the hitting, which is difficult to be observed correctly by human eyes. In the proposed automatic attack type acquisition method, the ”strong and powerful attack” and the ”slower attack” are identified by a velocity threshold in equation (4.16). In order to show how the value of the threshold TH affects the acquisition precision, seven different values are evaluated comparing with the conventional method. The result is shown in Fig 4.11. Since the velocity measurement ability of human eye is weak. From the theory, the precision of the proposed automatic variables acquisition method is higher than that of the manual acquisition as long as the physical data of the ball and player is precise. This evaluation is for comparing the difference between the automatic acquisition and manually acquisition.

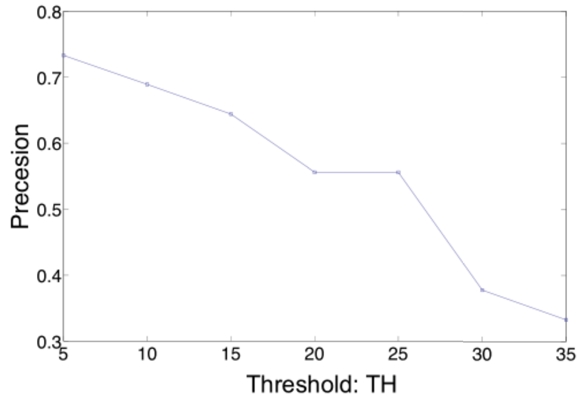


Figure 4.11: Acquisition precision for the attack type of different values of threshold TH compared with the human eyes judgment.

4.3.3 Experiments of Team Tactical Status Detection

As for the ground truth of team status detection, considering that the team status detection requires all the players and balls being recorded in the videos and the valid events, there are in total 29 rounds games are selected as the test data. With the decided parameter L whose value equals 30, there are 182 defensive ready statuses, 99 defensive statuses, 103 offensive ready statuses and 90 attack statuses in the selected test games.

In order to show the contribution of detailed part of the proposal, which uses both court region information and ball related information. Three kinds of features are trained and predicted by Random Forest classifier to show the performance of the proposed 3D space motion density feature.

The first experimental feature is generated by listing all the players' average positions and velocities. The second one is proposed area team motion density feature without using the ball information. The third experimental feature vector is the area team motion density feature vector using the ball related information.

In the experiments, for each team tactical status, I randomly select 30 samples as the training data and test the rest data. I use three standard evaluation criteria, the accuracy, recall and precision, to evaluate the detection performance. Definitions of used criteria are:

$$\begin{aligned}
&accuracy = \frac{TP + TN}{TP + FN + FP + TN} \\
&recall = \frac{TP}{TP + FN} \\
&precision = \frac{TP}{TP + FP}.
\end{aligned} \tag{4.26}$$

Table 4.5 shows the data of experimental result, in which the detection result of the trained three kinds of feature is listed. From the table it can be known, the test feature, which uses the individual player position, cannot detect the team status. Compared with the feature of individual player features, the spatial court region information and the ball related motions improve the detection rate a lot. Also, there is still large room for us to increase the detection rate. After analysis of the detection failure in the experimental result, there are two main reasons. One is the unstable players' position that leads to inaccurate feature vector. The other one is the videos samples in the intersection of two continuous team statuses, which has both features of the two statuses and becomes severe noise in training and prediction.

4.4 Conclusions

With the target of the automatic game strategy data acquisition from vision cameras, this section proposes the player roles detection for the acquisition of attack efficacy variables with the relative motion abruptness and the court area division, and the area team motion density feature is proposed for the detection of team tactical status. Firstly, the relative motion abruptness, which is calculated from the ball motion and player motion, represents the effort of each player to the play. This feature can be used to filter out the candidate role for each player. By dividing the court zone into setter, attacker and blocker zones based on the distance between the ball and the player, a court zone based filter is modeled to further detect the player role. the detection rate of the setter position, the number of available attackers, the attack time and the number of blockers is 100%, 100%, 97.8%, 100%. This is an average improvement of 8.3% compared with the detection rate of manual acquisition. The proposed method appeals higher performance on the evaluation of aspects including velocity information, play time and court border judgment. Secondly, the area team motion density feature is for the detection of team tactical status, which focuses on the team motion in different court zone that is corresponding to different plays. Thanks to using the spatial court information, the proposed feature detects the team status from the view of the integrate team on the court and the relationship to the ball, so that the influence caused by the team rotation, the player changing and different team formation can be reduced. The experimental detection rate of each team status (attack, defense-ready, offense-ready and offense status) are 75.2%, 84.2%, 79.7% and 81.6%.

My future target will dive into two aspects. First, extension to acquisition the strategy data for other volleyball play is necessary to achieve a comprehensive volleyball game analysis and tactics development system. Furthermore, based on the achievements of the GPU real-time acceleration [29], an real-time and low-delay volleyball analysis system is expected for the supporting of TV content broadcasting in world big games.

Table 4.3: Category of the value definition for attack efficacy variables.

Variables	Category	Definition
Setting zone (SZ)	Excellent setting zone	A rectangle zone located near the net
	Acceptable setting zone	A rectangle zone next to the excellent zone
	Not acceptable setting zone	All the remaining area
Number of available attacker (AN)	1	When only one attacker is available to attack
	2 or 3	When two or three attackers are available to attack
	4 or 5	When four or five attackers are available to attack
Attack tempo (A_{tempo})	1	Before the ball arrives to the setter's hands
	2	When the ball arrives to the setter's hands
	3	When the ball reaches the highest point of trajectory after leaving the setter's hands
Attack type (A_{type})	1	Power: A powerful attack
	2	Off-speed: A slower attack
Number of blockers (BN)	1 ×3	One attacker against three blockers
	1×2	One attacker against two blockers
	1×1	One attacker against one blocker
	1×0	One attacker against zero blocker

Table 4.4: Experimental result of attack efficacy variables acquisition.

Efficacy variables	**Metric**	**Conventional method**	**Proposed method**	**Total number**
SZ	correct number	38	45	45
	precision	84.4%	100%	\
AN	correct number	41	45	45
	precision	91.1%	100%	\
A_{tempo}	correct number	39	44	45
	precision	88.6%	97.8%	\
BN	correct number	47	47	47
	precision	100%	100%	\

Table 4.5: Detection results of the area team motion density based team tactical status detection.

Team status		Precision	Recall	Accuracy
	Feature1	0%	0%	0%
Defensive ready	**Feature2**	83.8%	50.00%	60.4%
	Feature3	74.6%	86.8%	81.6%
	Feature1	0%	0%	0%
Defence	**Feature2**	31.8%	39.1%	56.5%
	Feature3	48.7%	79.7%	79.7%
	Feature1	0%	0%	0%
Offensive ready	**Feature2**	28.7%	34.3%	68.9%
	Feature3	64.6%	87.7%	84.2%
	Feature1	0%	0%	0%
Attack	**Feature2**	31.3%	51.7%	73.4%
	Feature3	39.9%	88.3%	75.4%

Feature 1. Feature vector are generated by list all the players' average positions and velocities.
Feature 2. Area team motion density without using ball relative information.
Feature 3. Proposed Area team motion density feature.

Chapter 5

Conclusion and Future work

In this dissertation, the vision sensor based automatic ball related data acquisition algorithms are proposed for the realization of volleyball analysis systems. The proposals are mainly based on the abrupt motion feature and spatial importance for the acquisition of ball physical data, the event data and the strategy data. The proposed automatic acquisition algorithms achieve high detection rate (92%~100%) on various kinds of game data, which has a large potential in the contribution of the automatic Data Volley system for strategy analysis and providing new data for TV contents.

The proposed data acquisition system makes significant contribution to the further applications of the strategy analysis systems such as the automatic Data Volley system and provides new data for TV broadcasting contents. The high potential on further application of proposed data acquisition system is achieved by for features. Firstly, the data acquired by proposed algorithm has various types. All the physical data, event data and even strategy data are possible to be obtained. These data not only contribute to the game analysis from the low-level (game statistics) to high-level (couching supporting), but also provide copious new data for TV contents, such as the 3D ball trajectory. Secondly, the obtained data has centimeter-level precision. Since the data obtained manually such as using Data Volley are always analysed in court zone unit (3m×3m), the game data with centimeter level precision has higher potential in advanced strategy analysis. The high precision of game data also guarantees the reliability in TV contents. Thirdly, the proposed algorithm also acquires data that is difficult to be observed by human eyes, such as the velocity and direction information. By adding these new data, the applications of strategy analysis and TV contents are expanded. Finally, the data is automatically required from the game videos. That saves human labor and processing time a lot.

The future work dive into two aspects: the application of the further strategy analysis, and the real-time implementation of automatic data acquisition. Firstly, besides the achieved acquisition work for physical data, event data and strategy data, the automatic quality evaluation and new strategy creation are still remained for volleyball strategy analysis. These two topics are parts of my future works. For this future work, the challenges includes the limitation of professional-level game data and lack of supporting of volleyball specialists. By collaborating with the Sports Science School and Company, expanding the database, evaluation of the play quality and creating new game strategy is expected to use for the achievement the quality evaluation and strategy creation. Secondly, for the application of the TV broadcasting and acquisition game data timely for the effective strategy analysis, the real-time implementation of automatic data acquisition is necessary. By fully utilization of the GPU, there is a high potential to achieve a real-time system for current system. As the publications paper [29] shows, the GPU based real-time implementation for 3D ball tracking has been achieved. For the target of real-time automatic data acquisition, my future works on real-time system are the GPU implementation of the event data acquisition and strategy data acquisition. In order to achieve the algorithm acceleration, the GPU platform based algorithm parallelism and complexity reduction are expected.

References

[1] K. Deguchi, K. Hamasaki, T. Nakajima, and T. Okatani, "3-dimensional tracking of multiple object motions from multi-view images by mixed-state condensation algorithm." in *International Conference on Machine Vision Applications*, 2002, pp. 408–411.

[2] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[3] X. Wang, V. Ablavsky, H. B. Shitrit, and P. Fua, "Take your eyes off the ball: Improving ball-tracking by focusing on team play," *Computer Vision and Image Understanding*, vol. 119, pp. 102–115, 2014.

[4] A. Maksai, X. Wang, and P. Fua, "What players do with the ball: a physically constrained interaction modeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 972–981.

[5] X. Cheng, N. Ikoma, M. Honda, and T. Ikenaga, "Multi-view 3d ball tracking with abrupt motion adaptive system model, anti-occlusion observation and spatial density based recovery in sports analysis," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 100, no. 5, pp. 1215–1225, 2017.

[6] X. Cheng, N. Ikoma, M. Honda, and T. Ikenaga, "Ball state based parallel ball tracking and event detection for volleyball game analysis," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 100, no. 11, pp. 2285–2294, 2017.

[7] Y. Huang and J. Llach, "Tracking the small object through clutter with adaptive particle filter," in *International Conference on Audio, Language and Image Processing, 2008*. IEEE, 2008, pp. 357–362.

[8] F. Yan, W. Christmas, and J. Kittler, "A tennis ball tracking algorithm for automatic annotation of tennis match," in *British machine vision conference*, vol. 2, 2005, pp. 619–628.

[9] F. Yan, W. Christmas, and J. Kittler, "A maximum a posteriori probability viterbi data association algorithm for ball tracking in sports video," in *18th International Conference on Pattern Recognition, 2006.*, vol. 1. IEEE, 2006, pp. 279–282.

[10] J. Scharcanski, A. B. de Oliveira, P. G. Cavalcanti, and Y. Yari, "A particle-filtering approach for vehicular tracking adaptive to occlusions," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, pp. 381–389, 2011.

[11] C. GUO, Y. Lu, and T. Ikenaga, "Multiple likelihood models based particle filter for long-term full occlusion," *The Journal of the Institute of Image Electronics Engineers of Japan*, vol. 39, no. 5, pp. 580–589, 2010.

[12] Z. Duan, Z. Cai, and J. Yu, "Occlusion detection and recovery in video object tracking based on adaptive particle filters," in *Control and Decision Conference, 2009*. IEEE, 2009, pp. 466–469.

[13] G. Di Caterina and J. Soraghan, "Robust complete occlusion handling in adaptive template matching target tracking," *Electronics letters*, vol. 48, no. 14, pp. 831–832, 2012.

[14] I. Iraei and K. Faez, "Object tracking with occlusion handling using mean shift, kalman filter and edge histogram," in *2nd International Conference on Pattern Recognition and Image Analysis (IPRIA), 2015*. IEEE, 2015, pp. 1–6.

[15] H. Rezaee, A. Aghagolzadeh, M. H. Seyedarabi, and S. Al Zu'bi, "Tracking and occlusion handling in multi-sensor networks by particle filter," in *GCC Conference and Exhibition (GCC), 2011*. IEEE, 2011, pp. 397–400.

[16] X. Cheng, X. Zhuang, Y. Wang, M. Honda, and T. Ikenaga, "Particle filter with ball size adaptive tracking window and ball feature likelihood model for ball's 3d position tracking in volleyball analysis," in *Pacific Rim Conference on Multimedia*. Springer, 2015, pp. 203–211.

[17] H.-T. Chen, W.-J. Tsai, S.-Y. Lee, and J.-Y. Yu, "Ball tracking and 3d trajectory approximation with applications to tactics analysis from single-camera volleyball sequences," *Multimedia Tools and Applications*, vol. 60, no. 3, pp. 641–667, 2012.

[18] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[19] J. Ren, J. Orwell, G. A. Jones, and M. Xu, "Real-time modeling of 3-d soccer ball trajectories from multiple fixed cameras," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 3, pp. 350–362, 2008.

[20] M. Takahashi, K. Ikeya, M. Kano, H. Ookubo, and T. Mishina, "Robust volleyball tracking system using multi-view cameras", *23rd International Conference on Pattern Recognition (ICPR), IEEE*, pp. 2740–2745, 2016.

[21] D.P. Robertson, and R. Cipolla, "Structure from motion", *WileyBlackwell*, 2008.

[22] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009, vol. 494.

[23] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

[24] R. Hess and A. Fern, "Discriminatively trained particle filters for complex multi-object tracking," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 240–247.

[25] I. Kolonias, *Cognitive vision systems for video understanding and retrieval*. Citeseer, 2007.

[26] I. Almajai, J. Kittler, T. de Campos, W. Christmas, F. Yan, D. Windridge, and A. Khan, "Ball event recognition using hmm for automatic tennis annotation," in *2010 IEEE International Conference on Image Processing*. 2010, pp. 1509–1512.

[27] X. Cheng, N. Ikoma, M. Honda, and T. Ikenaga, "Event state based particle filter for ball event detection in volleyball game analysis," in *20th International Conference on Information Fusion (Fusion), 2017*. IEEE, 2017, pp. 1–6.

[28] C. Campbell and Y. Ying, "Learning with support vector machines," *Synthesis lectures on artificial intelligence and machine learning*, vol. 5, no. 1, pp. 1–95, 2011.

[29] Y. Hou, X. Cheng, and T. Ikenaga, "Real-time 3d ball tracking with cpu-gpu acceleration using particle filter with multi-command queues and stepped parallelism iteration," in *2nd International Conference on Multimedia and Image Processing (ICMIP)*. IEEE, 2017, pp. 235–239.

[30] J. Castro, A. Souza, and I. Mesquita, "Attack efficacy in volleyball: elite male teams," *Perceptual and motor skills*, vol. 113, no. 2, pp. 395–408, 2011.

[31] M. Bertini, A. Del Bimbo, and W. Nunziati, "Player identification in soccer videos," in *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, 2005, pp. 25–32.

[32] I. Atmosukarto, B. Ghanem, S. Ahuja, K. Muthuswamy, and N. Ahuja, "Automatic recognition of offensive team formation in american football plays," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 991–998.

[33] X. Cheng, N. Ikoma, M. Honda, and T. Ikenaga, "Simultaneous physical and conceptual ball state estimation in volleyball game analysis," in *Visual Communications and Image Processing (VCIP), 2017*. IEEE, 2017.

[34] S. Huang, X. Zhuang, X. Cheng, N. Ikoma, M. Honda, and T. Ikenaga, "Player feature based multi-likelihood and spatial relationship based multi-view elimination with least square fitting prediction for volleyball players tracking in 3d space," *IIEEJ*

transactions on image electronics and visual computing, vol. 4, no. 2, pp. 145–155, 2016.

[35] F. Esteves and I. Mesquita, "Study on the setting zone in elite men's volleyball according to setter player and set type [estudo da zona de distribuição no voleibol de elite masculino em função do jogador distribuidor e do tipo de passe]," *Portuguese Jounal of Sport Sciences*, vol. 7, pp. 36–37, 2007.

Publications

Journals (with review)

[1] Ziwei Deng, Yilin Hou, **Xina Cheng** and Takeshi Ikenaga, ”Multi-peak Estimation for Real-time 3D Ping-Pong Ball Tracking with Double-queue Based GPU Acceleration”, IEICE Trans. Fundamentals, Vol.E101-D, No.5, pp. 1251-1259, May 2018.

[2] **Xina Cheng**, Norikazu Ikoma, Masaaki Honda, Takeshi Ikenaga, ”Ball State Based Parallel Ball Tracking and Event Detection for Volleyball Game Analysis”, IEICE Trans. Fundamentals, Vol. E100-A, No. 11, pp. 2285-2294, Nov. 2017.

[3] **Xina Cheng**, Norikazu Ikoma, Masaaki Honda, Takeshi Ikenaga, ”Multi-View 3D Ball Tracking with Abrupt Motion Adaptive System Model, Anti-Occlusion Observation and Spatial Density Based Recovery in Sports Analysis”, IEICE Trans. Fundamentals, Vol. E100-A, No. 5, pp. 1215-1225, May 2017.

[4] Shuyi Huang, Xizhou Zhuang, **Xina Cheng**, Norikazu Ikoma, Masaaki Honda and Takeshi Ikenaga, ”Player Feature Based Multi-Likelihood and Spatial Relationship Based Multi-View Elimination with Least Square Fitting Prediction for Volleyball Players Tracking in 3D Space”, IIEEJ Trans. on Image Electronics and Visual Computing, Vol. 4, No.

2, pp. 145-155, Dec. 2016.

International Conferences (with review)

[1] **Xina Cheng**, Takeshi Ikenaga, ”3D Space Motion Dense Based Team Tactical Status Detection in Volleyball Game Analysis”, The 2nd International Conference on Digital Signal Processing (ICDSP 2018), pp. 108-112, Feb. 2018.

[2] **Xina Cheng**, Norikazu Ikoma, Masaaki Honda and Takeshi Ikenaga, ”Simultaneous Physical and Conceptual Ball State Estimation in Volleyball Game Analysis”, Visual Communications and Image Processing (VCIP2017), pp. 1-4, Dec. 2017.

[3] Ziwei Deng, Yilin Hou, **Xina Cheng**, Takeshi Ikenaga, ”Vectorized Data Combination and Binary Search Oriented Reweight for CPU-GPU Based Real-time 3D Ball Tracking”, Pacific-Rim Conference on Multimedia (PCM2017), pp. 1-9, Sep. 2017.

[4] Fanglu Xie, **Xina Cheng**, Takeshi Ikenaga, ”Motion State Detection based Prediction Model for Body Parts Tracking of Volleyball Players”, Pacific-Rim Conference on Multimedia (PCM2017), pp. 1-9, Sep. 2017.

[5] **Xina Cheng**, Norikazu Ikoma, Masaaki Honda and Takeshi Ikenaga, ”Event State Based Particle Filter for Ball Event Detection in Volleyball Game Analysis”, 20th International Conference on Information Fusion (Fusion 2017), pp. 1-6, July 2017.

[6] Guannan Wu, **Xina Cheng**, Takeshi Ikenaga, ”Racket Color Judgment Based Template Selection and Ellipse Shape Likelihood for 3D Racket Tracking in Ping-Pong”, The 32nd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC2017), pp. 588-591, July 2017.

[7] Fanglu Xie, **Xina Cheng**, Takeshi Ikenaga, ”Mixture Particle Filter with Block Jump Biomechanics Constraint for Volleyball Players Lower Body Parts Tracking”, IAPR International Conference on Machine Vision Applications (MVA2017), pp. 302-305, May 2017.

[8] Ziwei Deng, **Xina Cheng**, Takeshi Ikenaga, ”Ball-like Observation Model and Multi-peak Distribution Estimation Based Particle Filter for 3D Ping-pong Ball Tracking”, IAPR International Conference on Machine Vision Applications (MVA2017), pp. 390-393, May 2017.

[9] Yilin Hou, **Xina Cheng**, Takeshi Ikenaga, ”Real-time 3D Ball Tracking with CPU-GPU Acceleration Using Particle Filter with Multi-command queues and Stepped Parallelism Iteration”, Second International Conference on Multimedia and Image Processing (ICMIP 2017), pp. 235-239, Mar. 2017.

[10] Yuan Wang, **Xina Cheng**, Norikazu Ikoma, Masaaki Honda, and Takeshi Ikenaga, ”Motion Prejudgment Dependent Mixture System Noise in System Model for Tennis Ball 3D Position Tracking by Particle Filter”, Joint 8th International Conference on Soft Computing and Intelligent Systems and 17th International Symposium on Advanced Intelli-

gent Systems (SCIS&ISIS2016), pp.124-129, Aug. 2016.

[11] **Xina Cheng**, Masaaki Honda, Norikazu Ikoma, Takeshi Ikenaga, ”Anti-occlusion Observation Model and Automatic Recovery for Multi-view Ball Tracking in Sports Analysis”, 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), pp. 1501-1505, Mar. 2016.

[12] **Xina Cheng**, Xizhou Zhuang, Yuan Wang, Masaaki Honda, Takeshi Ikenaga, ”Particle Filter with Ball Size Adaptive Tracking Window and Ball Feature Likelihood Model for Ball's 3D Position Tracking in Volleyball Analysis”, Pacific-Rim Conference on Multimedia (PCM2015), pp. 203-211, Sep. 2015.

[13] Xizhou Zhuang, **Xina Cheng**, Shuyi Huang, Masaaki Honda, Takeshi Ikenaga, ”Motion Vector and Players' Features based Particle Filter for Volleyball Players Tracking in 3D Space”, Pacific-Rim Conference on Multimedia (PCM2015), pp. 393-401, Sep. 2015.

[14] **Xina Cheng**, Yuhi Shiina, Xizhou Zhuang, Takeshi Ikenaga, ”Player Tracking using Prediction After Intersection based Particle Filter for Volleyball Match Video”, APSIPA Annual Summit and Conference (ASC 2014), pp. 1-4, Dec. 2014.