

# ANTI-OCCLUSION OBSERVATION MODEL AND AUTOMATIC RECOVERY FOR MULTI-VIEW BALL TRACKING IN SPORTS ANALYSIS

Xina Cheng<sup>\*</sup>    Masaaki Honda<sup>\*\*</sup>    Norikazu Ikoma<sup>†</sup>    Takeshi Ikenaga<sup>\*</sup>

<sup>\*</sup> Graduate School of Information, Production and Systems, Waseda University, Japan

<sup>\*\*</sup> Faculty of Sports Science, Waseda University, Japan

<sup>†</sup> Faculty of Engineering, Kyushu Institute of Technology, Japan

## ABSTRACT

The 3D position of the ball plays a crucial role in professional sport analysis. In ball sports, tracking of ball's precise position accurately is highly required, whose performance is affected by inaccurate 3D coordinates and occlusion problem. In this paper, we propose anti-occlusion observation model and automatic recovery by 3D ball detection based on multi-view videos to track the ball in 3D space. The anti-occlusion observation model evaluates each camera's image and eliminates the influence of the cameras in which the ball is occluded. The automatic recovery method detects the ball's 3D position by homography relation of the multi-video and generates a new distribution to initiate the tracker when tracking failure is detected. Experimental results based on the HDTV video sequences, which were captured by four cameras located at the corners of the court, show that the success rate of the 3D ball tracking achieves 99.14%.

**Index Terms**— Sports analysis, multi-view ball tracking, occlusion, automatic recovery

## 1. INTRODUCTION

3D Ball tracking is a crucial part of sports analysis since the trajectory and movement of balls are key elements in analyzing the behaviors of players and the performances of teams. Because that the ball always moves in 3D space, the velocity, direction and trajectory of a ball can only be described accurately by 3D information. Therefore, the tracking rate and the precision of the ball's 3D position affect the reliability of game analysis. Our research focuses on tracking ball's 3D position with high tracking success rate.

We targets two problems due to the unique environment of games and the limitation of the shooting conditions, such as volleyball games. To solve these problems, a lot of methods have been proposed.

The first one is occlusion problem. As in volleyball games, the ball is occluded by players frequently. Under such situations, there is no tracking target in the image and it is impossible to track it. To solve this problem, Scharcanski [1] switches different sampling model and Guo [2] enlarges

the search region based on the algorithm of particle filter. However, like most occlusion handling method [3] [4] [5], their work cannot track the ball's true position due to the lack of multi-view information. Rezaee [6] uses the homography relation of two-view videos to declare as well as cancel the occlusion, but this work cannot obtain the targets' 3D position. Harguess [7] tracks objects by multi-view videos and solves the occlusion problems by deleting the camera with occlusion. This work has the limitation that it cannot deal with occlusion occurs in several cameras.

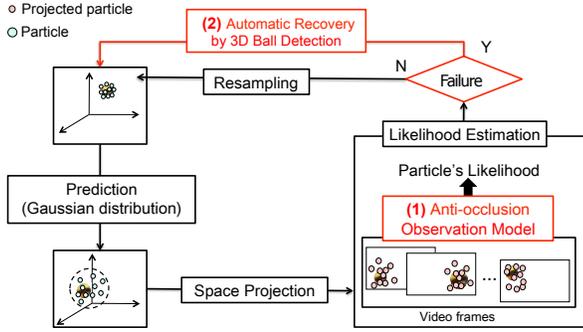
The second problem is that the precise 3D coordinate of a ball is difficult to obtain. Chen [8] has put forward an automated system to detect ball candidates and to fit a trajectory from a single camera's information. Lacking of multiple space information, his approximate result is unreliable. With multi-camera videos, Cheng [9] has proposed a 3D framework of particle filter [10] with ball feature likelihood for 3D ball tracking in volleyball analysis. This work can avoid the error produced by 3D structure reconstruction [11], but once the tracker loses the tracking targets, it is hardly to recover.

To cope with those problems in 3D ball tracking, this paper puts forward two proposals by using multi-view videos. The anti-occlusion observation model makes occlusion decision of each camera and switches the measurement model according to the number of cameras with occlusion. The automatic recovery by 3D ball detection detects tracking failures and recovers tracking by both 2D image features and homography relation of multiple cameras.

This paper is arranged as follows. Section 2 and section 3 cover the detail of the proposals and experiment results, respectively. Section 4 is the conclusion.

## 2. PROPOSALS

We use synchronization videos captured from different perspectives. The multiple cameras not only can obtain 3D information but also are capable of being checked and corrected by each other. Our proposals are implemented on Cheng's particle filter framework [9] for 3D ball tracking whose overall structure is shown in **Fig. 1**. The state space in our formula-



**Fig. 1:** Overall structure of 3D ball tracking implemented with proposals

tion is the ball's position in 3D space and the state at discrete time  $t$  is defined as  $\mathbb{X}_k = [x_k, y_k, z_k]$ ,  $k \in \mathbf{N}$ .  $(x_k, y_k, z_k)$  is the center coordinate of the ball's position. In prediction step, time evolution of  $\mathbb{X}_k$  is shown in equation (1)

$$\mathbb{X}_k = f(\mathbb{X}_{k-1}, \mathbf{v}_k) = \mathbb{X}_{k-1} + \mathbf{v}_k, \quad k \in \mathbf{N} \quad (1)$$

The system noise  $\mathbf{v}_k$  follows the Gaussian probability distribution.

Our first proposal is employed in likelihood estimation step. The likelihood  $L_i(\mathbb{X}_k)$  of each particle (sampling) is indicated as equation (2).

$$L_i(\mathbb{X}_k) = g \left[ L_i(\mathbb{X}_k; \mathbb{I}_k^1), \dots, L_i(\mathbb{X}_k; \mathbb{I}_k^m), \dots, L_i(\mathbb{X}_k; \mathbb{I}_k^M) \right] \quad (2)$$

Where,  $M$  is the total number of the cameras and  $\mathbb{I}_k = \{\mathbb{I}_k^1, \mathbb{I}_k^2, \dots, \mathbb{I}_k^m, \dots, \mathbb{I}_k^M\}$  is the observation space that is a collection of image frames at discrete time  $k$ . The  $L_i(\mathbb{X}_k; \mathbb{I}_k^m)$  is the likelihood value of the  $i_{th}$  particle estimated from the  $m_{th}$  camera's frame at state  $\mathbb{X}_k$ , which is calculated by color, circle and moving likelihood model [9]. We define it as the image likelihood of the  $m_{th}$  camera.  $g(\cdot)$  is a function to combining elements from each camera.

The proposed anti-occlusion observation model switches the definition of  $g(\cdot)$  adaptively and the detail is introduced in the following subsection.

Then, we propose the second proposal after state estimation step, which estimates the state according to the posterior distribution  $p(\mathbb{X}_k | \mathbb{I}_k)$ .

$$p(\mathbb{X}_k | \mathbb{I}_k) \approx \sum_i^I \mathbf{w}_k^i \delta(\mathbb{X}_k - \mathbb{X}_k^i) \quad (3)$$

Where  $\delta(\cdot)$  is the Dirac delta function,  $\mathbf{w}_k^i$  is the normalized importance weight of each particle. Here, the proposed automatic recovery method will replace the resampling step to re-initialize the tracker when tracking failure is detected.

## 2.1. Anti-occlusion Observation Model

We estimate the occlusion in each image by a threshold of image likelihood. For each particle, when the ball is occluded in

one frame, the corresponding image likelihood  $L_i(\mathbb{X}_k; \mathbb{I}_k^m)$  is low. The occlusion judgment of a camera is shown as equation (4).

$$L_i(\mathbb{X}_k; \mathbb{I}_k^m) < tr_o \quad (4)$$

Here  $tr_o$  is a threshold of  $L_i(\mathbb{X}_k; \mathbb{I}_k^m)$ . If the  $L_i(\mathbb{X}_k; \mathbb{I}_k^m)$  of one particle makes the equation (4) true, the probability of this particle being projected on ball's region is low and we assume the ball is occluded in this camera.

Based on the number of cameras with occlusion, we switch the measurement function  $g(\cdot)$  mentioned in equation (2).

$$g(\cdot) = \begin{cases} \sqrt[M]{\prod_{m=1}^M L_{i,m}(\mathbb{X}_k; \mathbb{I}_k^m)} & \# = 0 \\ \sqrt[M-\#]{\prod_{\alpha=1}^{M-\#} L_{i,\alpha}(\mathbb{X}_k; \mathbb{I}_k^\alpha)} & 0 < \# \leq M-2 \\ \sqrt[M-1]{\prod_{\beta=1}^{M-1} L_{i,\beta}(\mathbb{X}_k; \mathbb{I}_k^\beta)} & \# > M-2 \end{cases} \quad (5)$$

Where  $\#$  is the number of cameras with occlusion and  $M$  is the total number of cameras.

(1) When  $\# = 0$ , each camera's image likelihood owns high probability to be a ball so all of them can be used to calculate the likelihood of a particle.

(2) When  $0 < \# \leq M-2$ , it can be known that the ball is occluded in some cameras. To eliminate the influence of low likelihoods caused by occlusion, we just combine the higher image likelihoods by  $g(\cdot)$ . In this situation, although some camera's image likelihoods are low, there are still enough high image likelihoods showing the 3D position of the ball. So we can remove the low likelihood directly.

Here, the index  $\alpha$  is the mark of the image likelihood which are higher than threshold  $tr_o$ .

(3) When  $\# > M-2$ , it is hardly to judge the low likelihood is caused by occlusion problem or not. So we rearrange all the image likelihoods by their numeric value and reduce the lowest one to measure the likelihood.

Here the index  $\beta$  is the camera number after rearrangement and the smaller number represents the larger image likelihood.

## 2.2. Automatic Recovery by 3D Ball Detection

The proposed automatic recovery method re-initializes the tracker automatically when tracking failed by three steps.

### 2.2.1. Tracking Failure Detection

We detect tracking failure after the state estimation in particle filter algorithm by the weight distribution, which can reflect the tracking situation of the tracker. If the tracker loses

its target, the weight distribution of the particles  $\mathbf{w}_k^i$  will be smooth without obvious peak. So that the maximum value of the weight will be small. Thus, equation (6) is a necessary condition of tracking failure detection at the discrete time  $k$ .

$$\max(\mathbf{w}_k^i) < tr_{fail} \quad (6)$$

Here  $tr_{fail}$  is the threshold of tracking failure. In this way, if in state  $\mathbb{X}_k$  the tracking failure is detected, we skip resampling step and detect the ball in image space to obtain some ball candidates.

### 2.2.2. Ball Candidate Detection

At first, we raster scan pixels in the video frames with a certain interval. For each scanning pixel, we assign it with a value as equation (7).

$$p_i = \sum_{d_{i,j} < \rho} L_j \frac{d_{i,j}}{\rho} \quad (7)$$

Where the index  $i$  and  $j$  represents two pixels in the image.  $p_i$  is the assigned value for the  $i_{th}$  pixel.  $L_j$  is the ball feature based likelihood value of the  $j_{th}$  pixel calculated in a circle region. Radius  $\rho$  of this circle region is calculated by the physical position and size of the ball.  $d_{i,j}$  is the distance between the two pixels.

By setting a likelihood threshold, the pixels with high possibility to be the ball's position can be obtained. Then we classify these pixels into groups according to: (1) if the distance between two pixels is smaller than the given distance threshold, they are assumed belonging to same group; (2) if a pixel belongs to more than one groups, these groups are assumed as one group.

After grouping pixels, we fuse these pixel groups into ball candidates by equation (8) and (9).

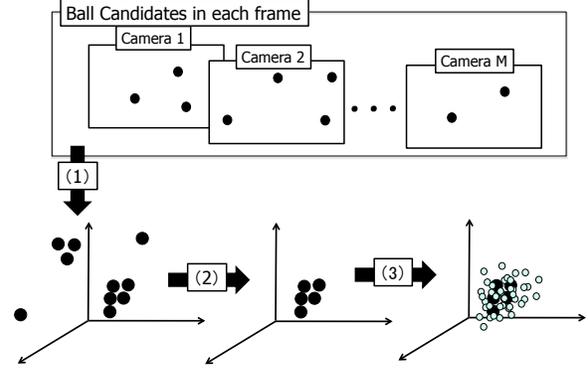
$$x_g = \eta \sum_{\phi=0}^{\Phi} x_i p_i \quad (8)$$

$$y_g = \eta \sum_{\phi=0}^{\Phi} y_i p_i \quad (9)$$

$(x_g, y_g)$  is the coordinate of ball candidates fused from  $g_{th}$  group and the  $(x_i, y_i)$  is the coordinate of the  $i_{th}$  pixel in  $g_{th}$  group.  $\Phi$  is the number of pixels in the  $g_{th}$  group and  $\eta$  is a normalization factor of  $p_i$ .

### 2.2.3. Recovery (Re-initialization)

To recover the tracker, a new 3D probability distribution  $P(\mathbb{X})$  of particles is required for estimating the state  $\mathbb{X}_k^i$ . We distribute  $P(\mathbb{X})$  not only based on the 2D ball candidates of the multi-view images but also using the 3D homography relation of every camera.



**Fig. 2:** The process of recovering the tracker with the detected ball candidates consists 3 steps. (1) Each pair of ball candidates is reconstructed to 3D space to obtain 3D ball candidates; (2) Elimination the noise points by distance judgments; (3) New probability distribution is generated to re-initialize the tracker.

In **Fig. 2**, we reconstruct all the combination of 2D coordinates as step (1), a sequence of reconstructed 3D coordinates  $\mathbf{x}_k^n, n = 1, 2, \dots, N$  is obtained. Here  $N$  is the total number of reconstructed 3D coordinates.

To eliminate the obvious noise, we evaluate each coordinate by the following condition equations.

$$\min(|\mathbf{x}_k^n - \mathbf{x}_k^{n_0}|) > tr_{dis1} \quad (10)$$

$$|\mathbf{x}_{k-1} - \mathbf{x}_k^{n_0}| > tr_{dis2} \quad (11)$$

In equation (11), the  $\mathbf{x}_{k-1}$  is the estimated tracking position of the ball in discrete time  $k-1$  before tracking failed. The  $tr_{dis1}$  and the  $tr_{dis2}$  are two different threshold of distance. If the coordinate  $n_0$  satisfies the two equations simultaneously, we judge this  $n_0$  as noise point and delete it.

As the step (3) in **Fig. 2**, we arrange the particles for each coordinate according to its weights after removing the noise. So the  $\mathbb{X}_k$  consists of a set of states  $\{\mathbb{X}_k^1, \dots, \mathbb{X}_k^n, \dots, \mathbb{X}_k^N\}$  which are derived from the coordinates respectively. For each coordinate,

$$\mathbb{X}_k^n = \mathbb{X}_k^n + \sigma_k \quad (12)$$

$\sigma_k$  is the system noise that follows the Gaussian distribution.

The particle number distributed to every coordinate  $PN_n$  is depending on its weight  $\mathbf{W}_k^n$ .

$$PN_n = I \cdot \mathbf{W}_k^n \quad (13)$$

$I$  is the total amount of particles. The  $\mathbf{W}_k^n$  of each coordinate is calculated by equation (14). For coordinate  $n_0$ ,

$$\mathbf{W}_k^{n_0} = 1 - \frac{1}{\lambda} \sum_{n=1}^N (\mathbf{x}_k^n - \mathbf{x}_k^{n_0})^2 \quad (14)$$

Here,  $\lambda$  is a coefficient to ensure

$$\sum_{n=1}^N \mathbf{W}_k^n = 1 \quad (15)$$

**Table 1:** Tracking success rate of anti-occlusion observation model comparing with conventional work.

Method	Cheng’s [9] tracking framework		proposed anti-occlusion observation model		Combination of proposals	
Sequence	Successful HIT	Success rate	Successful HIT	Success rate	Successful HIT	Success rate
set1	148	65.49%	203	89.82%	221	97.79%
set2	142	59.92%	208	91.63%	226	99.58%
set3	146	63.20%	221	95.67%	231	100%
total	436	62.82%	622	91.07%	688	99.14%

In addition, the distribution of  $\mathbb{X}_k$  is obtained to re-initialize the tracker. Sometimes, because of occlusion event or other interference, there is no detection result. Under this situation, this frame will be skipped and the automatic recovery method will be executed in next frame.

### 3. EXPERIMENT AND DISCUSSIONS

#### 3.1. Experimental Sequences and Evaluation Method

The experiment is based on videos recording the total 3 sets of the final game of an official volleyball match which contains several kinds of volleyball scenes (2014 Japan Inter High school Games of Men’s Volleyball held in Tokyo Metropolitan Gymnasium in Aug. 2014) by four cameras located at corners of the court. The video’s resolution is  $1920 \times 1080$ , the frame rate is 60 fps and cameras’ shutter speed is 1000 per second so there is no motion blur in the videos.

To evaluate the performance of our proposals, we give a definition of *HIT* that is a time period between two consecutive ball hittings. So in the game for experiment there are totally 684 *HIT*s and in each set there are 226, 227 and 231 *HIT*s. We judge a *HIT* is successful if the ball can be tracked continuously during it and the success rate is calculated as:

$$success\ rate = \frac{\sum\ success\ful\ HIT}{\sum\ HIT} \times 100\% \quad (16)$$

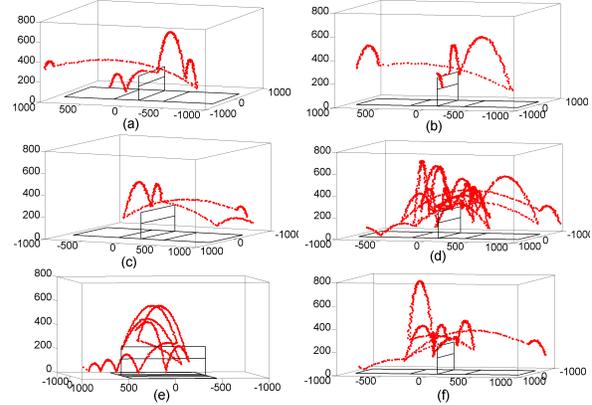
#### 3.2. Experimental Result and Comparison

We use Cheng’s tracking framework [9] without occlusion handling method and recovery method as the experimental baseline. To test and present the contribution of each proposal, we implement the anti-occlusion observation model first and then combine it with automatic recovery method.

**Table. 1** gives the comparison data. Our anti-occlusion observation model increases the tracking success rate to 91.07% and automatic recovery method makes the success rate reach over 99%. **Fig. 3** shows some ball trajectory plotted by the tracked 3D position of the ball.

#### 3.3. Analysis and Discussions

The experimental results shows that the proposed two methods achieve significant improvement in tracking success rate



**Fig. 3:** Example of plotted trajectory of tracking results

because of using multi-view information. The anti-occlusion observation model devotes a lot to the success rate by eliminating the image likelihood of the cameras in which occlusion problem occurs. The presented automatic recovery by 3D ball detection recovers the failed tracking. While improving the tracking rate so much, these methods have to pay some compensate, such as the deleting of some cameras’ information makes the estimated positions of the particles unstable. On the whole, although there are still some points need to be improved, the proposals own a high performance in 3D ball tracking for sports analysis.

### 4. CONCLUSION

This paper presents anti-occlusion observation model and automatic recovery by 3D ball detection to achieve high tracking rate of ball’s 3D position tracking with multi-view videos. The anti-occlusion observation model eliminates the interference of the occlusion problem that occurs in some cameras. It handles the occlusion problems by arranging 3D information reasonably so that the multi-view videos are given full play. The automatic recovery re-initialize the particle filter when tracking failure occurs automatically by the ball’s 3D position detection. Both the proposals have been implemented on prepared test sequences. The success rate has been raised up to 99.14% compared with conventional work.

## 5. REFERENCES

- [1] J. Scharcanski, A.B. De Oliveira, P.G. Cavalcanti, Y. Yari, "A particle-filtering approach for vehicular tracking adaptive to occlusions. Vehicular Technology," *IEEE Transactions on*, 60(2), 381-389, 2011.
- [2] C. Guo, Y Lu, T. Ikenaga, "Multiple Likelihood based Particle Filter for Long-term Full Occlusion," *The Journal of IIEEEJ*, vol. 39, pp. 580-589, 2010.
- [3] Z. Duan, Z. Cai, J. Yu, "Occlusion detection and recovery in video object tracking based on adaptive Particle filters," *Control and Decision Conference*, 2009. CCDC'09. Chinese. IEEE, 2009.
- [4] G. Di Caterina, J. J. Soraghan, "Robust complete occlusion handling in adaptive template matching target tracking," *Electronics letters* 48.14: 831-832, 2012.
- [5] I. Iman, K. Faez, "Object tracking with occlusion handling using mean shift, Kalman filter and Edge Histogram," *Pattern Recognition and Image Analysis (IPRIA)*, 2015 2nd International Conference on. IEEE, 2015.
- [6] H. Rezaee, A. Aghagolzadeh, M.H. Seyedarabi, S.A. Zu'bi, "Tracking and occlusion handling in multi-sensor networks by particle filter," *In GCC Conference and Exhibition (GCC)*, 2011 IEEE (pp. 397-400), 2011.
- [7] J. Harguess, CHu, J. K. Aggarwal, "Occlusion robust multi-camera face tracking," *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011 IEEE Computer Society Conference on. IEEE, 2011.
- [8] H. Chen, W. Tsai, Lee. S, Yu. J, "Ball Tracking and 3D Trajectory Approximation with Applications to Tactics Analysis from Single-camera Volleyball Sequences," *Multimedia Tools and Applications*, 60(3):641-667.S, 2012.
- [9] X. Cheng, X. Zhuang, Y. Wang, M. Honda, T. 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 (PCM)*, 2015.
- [10] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *Signal Processing, IEEE Transactions on*, 50(2), 174-188, 2002.
- [11] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision," *In Cambridge university press*, 2003