15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Ball-like Observation Model and Multi-peak Distribution Estimation Based Particle Filter for 3D Ping-pong Ball Tracking

Ziwei Deng[1] Xina Cheng Takeshi Ikenaga
Graduate School of Information, Production and Systems, Waseda University
Kitakyushu, Fukuoka, Japan
vivideng@toki.waseda.jp[1]

Abstract

3D ball tracking is of great significance to ping-pong game analysis, which can be utilized to applications such as TV content and tactic analysis. To achieve a high success rate in ping-pong ball tracking, the main problems are the lack of unique features and the complexity of background, which make it difficult to distinguish the ball from similar noises. This paper proposes a ball-like observation model and a multi-peak distribution estimation to improve accuracy. For the ball-like observation model, we utilize gradient feature from the edge of upper semicircle to construct a histogram, besides, ball-size likelihood is proposed to deal with the situation when noises are different in size with the ball. The multi-peak distribution estimation aims at obtaining a precise ball position in case the particles' weight distribution has multiple peaks. Experiments are based on ping-pong videos recorded in an official match from 4 perspectives, which in total have 122 hit cases with 2 pairs of players. The tracking success rate finally reaches 99.33%.

1 Introduction

Due to the development of computer vision, more people are interested in ping-pong game analysis, which is of great commercial value. 3D ball tracking plays a crucial role in ping-pong game analysis, which can obtain 3D ball position and velocity for applications such as TV content and tactic analysis. Therefore, our research aims at improving the success rate of 3D ping-pong ball tracking in real and complex conditions for ping-pong game analysis.

The lack of unique features of the ping-pong ball and the complex background in real conditions are the main problems that affect the tracking accuracy. In formal games such as the Olympic Games, the ping-pong ball is in white, which has less unique color features and is easily affected by luminance and background. The complex background of the video bringing a mass of similar noises also makes it difficult to track the ball accurately. Moreover, the occlusion problem, the high speed, the abrupt motion change and the camera calibration error increase the difficulty as well.

There are several works aiming at 3D ball tracking in recent years. The Hawk-Eye system [1] is applied in real matches to obtain the ball's trajectory these years. However, this system needs multiple high-speed cameras placed at high locations in order to provide high-quality videos against a simple background of the court, which makes it expensive and limited. As for [2] and [3], although they achieved some degrees of success in detecting or tracking a ball, but their scenes of the test videos are relatively simple. While [4]-[7] are tested in complex situations, but their tracking targets such as volleyball and orange ping-pong ball, have more unique color features.

Some of the works have already solved some difficult problems. Multiple features such as color and moving features are applied in [4]-[7] for ball tracking. To solve the occlusion problem, [1]-[7] use multiple cameras in different perspectives in order to capture the ball in every frames. In [6], an anti-occlusion observation model is proposed to solve the occlusion problem by removing the lowest likelihood of all the cameras. [7] has put forward a system model with mixture system noises in order to adapt the high speed and the abrupt motion change in 3D tennis ball tracking.

Some works use the gradient feature and the size feature to overcome the problem of the lack of unique features. The observation model of [5] only utilizes the magnitude of the gradient to judge the circle edge of the ball, which leads to a great deal of noises on complex objects. In [4], canny edge detection and Hough transform are used to detect a circle, which cost a large computation. The ball-size feature is also applied by [4] in order to reduce candidate circles with different radius. However, owing to [5], an adaptive radius of the tracking window can be obtained to decide each particle's ROI (region of interest) in our work. The estimation models of the particle filter in [5]-[7] cannot obtain a precise result when the particles' weight distribution has multiple peaks, which is a very serious problem to be solved.

This paper proposes the upper semicircle likelihood for a more appropriate use of the gradient feature in ping-pong ball tracking, which utilizes both the magnitude and the orientation of the gradient on the upper semicircle edge. And the ball-size feature is proposed to eliminate similar noises caused by complex background using the ball-size likelihood. Meanwhile, the multi-peak distribution estimation method is proposed to obtain a precise ball position when the particles' weight distribution has multiple peaks.

This paper is arranged as follows. Section 2 and section 3 cover the detail of the 3D ball tracking method and the proposals, and the experiment and the conclusion are in section 3 and section 4.

2 3D Ball Tracking Method

The Multi-view 3D ball tracking method using particle filter is implemented in our research. By using synchronous videos captured from 4 perspectives, the 3D information can be obtained and the occlusion problem can be solved to some extent. The camera calibration method we choose is from Hartley's work [8].

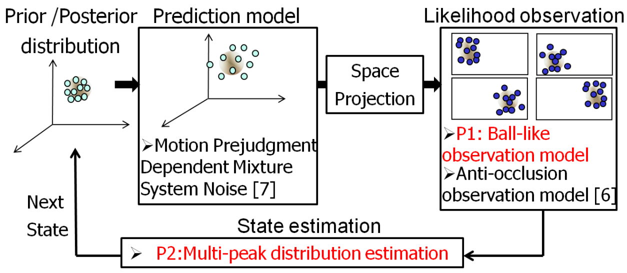

Figure 1. Overall structure of 3D ball tracking.

The overall structure of our framework is shown in Fig.1. The state vector is defined as below, which contains the 3D position of the ping-pong ball at discrete time k.

$$\mathbb{X}_k = [x_k, y_k, z_k], \quad k \in \mathbf{N} \tag{1}$$

In the system model, the time evolution of the state vector is given with a mixture system noise proposed in [7] which is adaptive to different motions.

In the space projection, the 3D position of particles are projected to 4 image planes, while the tracking window[5] is defined to be the ROI (region of interest) for each particles which is based on the ball's projected radius.

Our first proposal is proposed in the observation model. We define the observation of camera m as $\mathbb{I}_k^m$ at discrete time k. Then the likelihood $L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$ of the i_{th} particle is calculated based on the ball-like observation model.

For the state estimation, we propose the second proposal named multi-peak distribution estimation, which estimates the state according to more than the posterior distribution, as well as the influence from similar noises.

3 Proposals

3.1 Ball-like observation model

The Ball-like observation model totally includes 4 kinds of likelihood. One particle finally gets its likelihood of 1 perspective by the equation below.

$$\begin{aligned} L(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) =& L_{color}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) \cdot L_{move}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) \\ &\cdot L_{circle}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) \cdot L_{size}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) \end{aligned} \tag{2}$$

The upper semicircle likelihood $L_{circle}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$ and the ball-size likelihood $L_{size}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m)$ are described in detail in this section.

3.1.1 Upper semicircle likelihood

In the ball tracking, the circle-shape is an obvious feature to distinguish from noises in the complex background, which is robust to rotation problem as well. However, in ping-pong ball tracking, the ball's gradient information on the lower semicircle edge is easily affected by luminance because of its white color. Therefore, we choose to use the upper semicircle edge to represent its gradient feature.

To calculate the possibility that the tracking window is located at a circle object, the gradient histogram of the pixels on the upper semicircle edge is constructed. We use the orientation of the gradient to decide the bin of the histogram and divide it into 48 bins. As for the orientation which is between 2 bins, its corresponding magnitude is divided to 2 parts based on the different weight and is respectively added to 2 bins. So the histogram $h = {h_i^u}_{u=1,2,\ldots,48}$ can be calculated as below.

$$h_i^u = \lambda \sum_{p=1}^{P} |G_i(x_p, y_p)| \delta(\hat{\theta}_i(x_p, y_p) - u) \tag{3}$$

where $\hat{\theta}_i(x_p, y_p)$ is the quantized orientation and $|G_i(x_p, y_p)|$ is the magnitude. P is the number of pixels in the ROI (region of interest). λ is a normalization factor and $\delta(\cdot)$ is the Dirac delta function.

After generating the histogram for each particle, the distance between the particle's histogram and the template histogram $\hbar$ is calculated based on the Bhattacharyya coefficient $BC(h_i^u, \hbar)$.

$$L_{circle}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1 - BC(h_i^u, \hbar)}{2\sigma^2}) \tag{4}$$

3.1.2 Ball-size likelihood

Most noises have high likelihood in color and moving features, such as the player's arm and the label on the player's uniform. Fortunately, despite the high similarity with the ball, the size is another feature to eliminate these ball-like noises. Under this circumstance, the ball-size likelihood is designed to decrease the noises' likelihood.

The original ROI (region of interest) is a circle whose radius is the actual radius of the ball in image plane. Thus, it is suitable for the HSV histogram construction and the background subtraction since other noises is removed from consideration. When it comes to enlarging the ROI, the degree of the enlargement is set as 1.5 times of the original radius and the circle shape of the ROI is changed into a square shape. The number of pixels in the EROI (enlarged ROI) is defined as S_{EROI}.

The size ratio calculation starts with a scan in a color filter and a moving filter for every pixel. In Fig.2, the method to obtain the "ball-like mask" is clearly expressed. Once getting the ball-like mask, counting

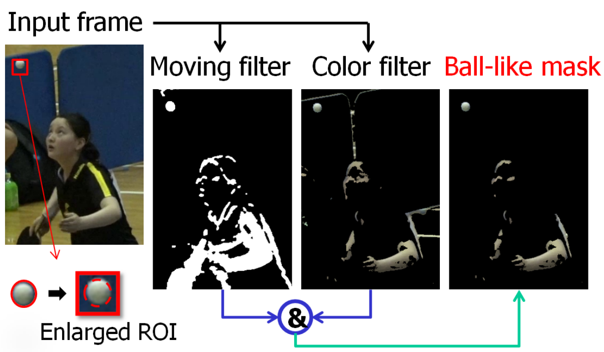

Figure 2. Ball-like mask.

the total number of the valued pixels ($S_{ball-like}$) in the EROI based on this mask is proceeded.

The size ratio R is evaluated by the below equation.

$$R = \frac{S_{ball-like}}{S_{EROI}} \tag{5}$$

For the real ball-size ratio r_b, its value is the original ROI area divided by S_{EROI}, which means that the EROI totally covers the whole ball. At this ratio, the ball-size likelihood definitely reaches the maximum. The function that transfer the size ratio R to the ball-size likelihood is shown as below.

$$L_{size}(\mathbb{X}_k^{(i)}; \mathbb{I}_k^m) = \begin{cases} RC & R < r_b - 0.1 \\ 1 & |R - r_b| \leq 0.1 \\ 1 - RC & R > r_b - 0.1 \end{cases} \tag{6}$$

where C is a constant decided by the enlarging degree of the EROI. To sum up, the farther the size ratio from the r_b, the lower the ball-size likelihood.

3.2 Multi-peak distribution estimation

For the situation that some similar noises exist in the complex background, the particles' weight distribution based on the observation model may have multiple peaks, as Fig.3 shows. As for the conventional estimation method in [5]-[7], the weighted average position of all the particles P_{wa}^I is calculated to be the target position, which is imprecise as we can conclude from Fig.3.

The multi-peak distribution estimation method is proposed to obtain a precise ball position under this circumstance. The main idea is that we divide particles into different groups based on their positions in order to separate the target with noises, and then we judge which group is the real target. Thus, this method includes a particles grouping method and a best group judgment.

After resampling, we can get several particles with relatively higher weight. These particles are scanned and divided into multiple groups according to the rules below.

(1) If the distance between 2 particles is smaller than the given threshold, they are assumed to belong to the same group.

(2) If a particle doesn't belong to any existed group, a new group is added based on this particle.

(3) If a particle belongs to more than 1 groups, these groups are assumed as one group.

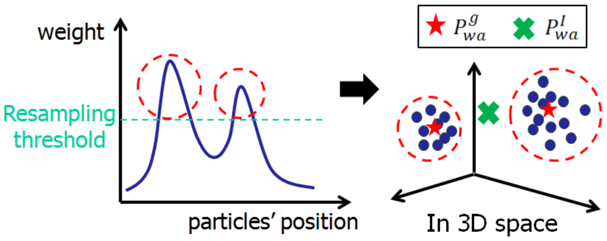

Figure 3. Multi-peak distribution estimation.

After grouping particles, the weighted average position P_{wa}^g of group $\mathbb{G}_k^g$ is calculated.

To judge which group is the real target, we calculate the "best-group likelihood" based on the particles likelihood and the quantity of the group members. Firstly, we calculate the distance $d_g^{(i)}$ between the P_{wa}^g and the particles belonging to $\mathbb{G}_k^g$. In our conclusion, the particle which is farther from the P_{wa}^g should have less influence on the judgment, so that the best-group likelihood is defined as below.

$$L_{bg}(\mathbb{G}_k^g) = \sum_{i=0}^{i<s_g} \kappa_g^{(i)} \cdot L(\mathbb{X}_k^{(i)}; \mathbb{I}_k) \tag{7}$$

where s_g is the total number of the particles in $\mathbb{G}_k^g$. Parameter $\kappa_g^{(i)}$ is designed as below to reduce the influence from the particles which are farther from the P_{wa}^g.

$$\kappa_g^{(i)} = \frac{e^{-d_g^{(i)}} - e^{-r_g}}{1 - e^{-r_g}} \tag{8}$$

where r_g is the radius of $\mathbb{G}_k^g$.

Afterwards, the group with the highest best-group likelihood is judged as the best one. By using this judging method, the effect of the particles which are closer to the P_{wa}^g is enlarged, and the influence from the quantity of the group members is reduced, in other words, not the more particles, the higher likelihood.

Finally, the weighted average position of the best group is treated as the ball's position, which is relatively more credible than the result of the conventional estimation method.

4 Experiment

4.1 Experiment sequences

The experiment is based on the videos recorded in an official ping-pong match (2016 Kanto University Student Ping-pong Championship) by 4 cameras located at 4 corners of the court. The video's resolution is 1920×1080, the frame rate is 60 fps and the cameras' shutter speed is set as 1000 per second. We randomly cut 15 sequences for each perspective from the videos, as Fig.4 shows, which in total include 3913 frames with 2 pairs of players. And these sequences contain 122 hit cases including smash, chop, serve and so on.

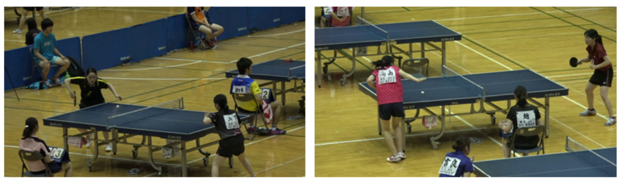

Figure 4. Example frames.

4.2 Evaluation method

To evaluate the performance of our proposals, we give a definition of *success frame* that the projected tracking window of the 3D position should exactly cover or cover part of the ball in at least 2 perspectives of the frame. This is because some error exists to affect the precision, which is caused by camera calibration. The calculation of the tracking success rate is shown as below.

$$Success\ rate = \frac{\sum Success\ frame}{Total\ frames} \times 100\% \qquad (9)$$

4.3 Experimental result and analysis

For comparison, the contributions of our 2 proposals are respectively evaluated. The experimental results are shown in Table 1, and some short 3D ball trajectory results plotted by the 3D ball position are shown in Fig.5. As a conclusion, our work achieves 99.33% tracking success rate and gains 67.49% improvement compared with conventional frame work, which indicates that our proposals are effective in 3D ping-pong ball tracking.

However, when the ball is occluded in 2 perspectives, similar noises still have some bad effect on the tracking result sometimes, as shown in the third trajectory in Fig.5. I think this kind of problems may be solved by using more video perspectives.

Table 1. Experimental results.

Experiment work	Success frame	Success rate
Conventional frame work	1246	31.84%
P1[1]	3796	97.01%
P1+P2[2]	3897	99.33%

[1]P1: Ball-like observation model
[2]P2: Multi-peak distribution estimation

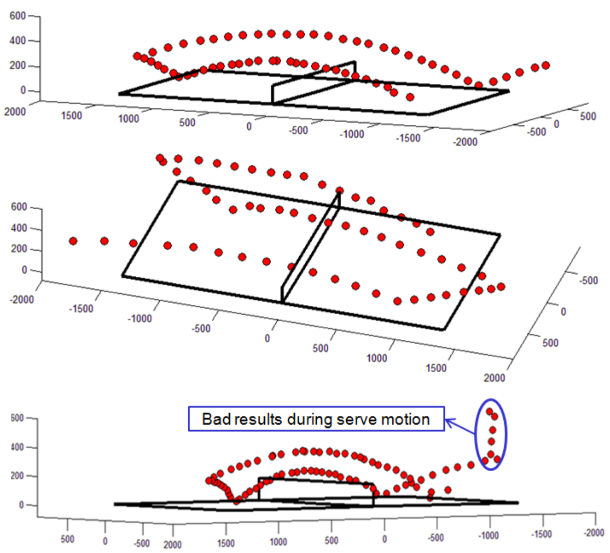

Figure 5. 3D ball trajectory.

5 Conclusion

This paper proposes a ball-like observation model and a multi-peak distribution estimation method, aiming at solving the problems in 3D ping-pong ball tracking, such as the lack of unique features and the complexity of background. The first proposal utilizes the gradient feature from the edge of upper semicircle to construct a histogram in the upper semicircle likelihood, besides, the ball-size likelihood is proposed to eliminate the similar noises which are different in the size with the ball. The multi-peak distribution estimation aims at obtaining a precise ball position in case the particles' weight distribution has multiple peaks. The experiment is based on 15 sequences shot in an official match with 4 perspectives. The tracking success rate reaches 99.33% and gains 67.49% improvement compared to the conventional frame work.

In the future, we plan to accelerate our 3D ball tracking method using GPU to meet the demand for real-time applications. And we may need to do some modifications to make it more hardware-friendly.

References

[1] N. Owens, C. Harries, C. Stennett, "Hawk-eye tennis system," *International Conference on Visual Information Engineering*, pp.182-185, 2003.

[2] X. Chen, Q. Huang, Member IEEE, W. Zhang, Z. Yu, R. Li, P. Lv, "Ping-pong Trajectory Perception and Prediction by a PC based High Speed Four-Camera Vision System," *8th World Congress on Intelligent Control and Automation*, pp.1087-1092, 2011.

[3] H. Bao, X. Chen, Z. Wang, M. Pan, F. Meng, "Bouncing Model For the Table Tennis Trajectory Prediction and the Strategy of Hitting the Ball," *2012 International Conferenceon Mechatronics and Automation (ICMA)*, pp.2002-2006, 2012.

[4] H. Myint, P. Wong, L. Dooley, A. Hopgood, "Tracking a table tennis ball for umpiring purposes," *14th IAPR International Conference on Machine Vision Applications (MVA)*, pp.170-173, 2015.

[5] X. Cheng, X. Zhuang, Y. Wang, M. Honda, T. Ikenaga, "Particle Filter with Ball Size Adaptive Tracking Window and Ball Feture Likelihood Model for Ball's 3D Position Tracking in Volleyball Analysis," *Pacific-Rim Conference on Multimedia (PCM)*, 2015.

[6] X. Cheng, M. Honda, N. Ikoma, T. 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)*, Mar. 2016.

[7] Y. Wang, X. Cheng, N. Ikoma, M. Honda, T. 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 Intelligent Systems (SCIS&ISIS2016)*, Aug. 2016.

[8] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press*, 2003.