Situation Recognition Using 3D Positional Information of Ball from Monocular Soccer Image Sequence

Takuro Nishino*, Yasuo Ariki**, Tetsuya Takiguchi**

* *Graduate School of Science and Technology, Kobe University, Japan*
** *Organization of Advanced Science and Technology, Kobe University, Japan*
E-mail: fantakuro@me.cs.scitec.kobe-u.ac.jp, {ariki, takigu}@kobe-u.ac.jp

Abstract

In this paper, we propose a system that tracks a ball stably and accurately and detects the events of the game by using the 3D positional information for automatic soccer video production. We use 3D particle filter with the state vector of nine dimensions in the ball tracking. Since the ball tracking by particle filter is a local search, it is difficult to continue tracking when it fails. Thus, we solve this problem by switching the local search to 3D global search, and by interpolating the lost coordinates. As a result, the tracking accuracy was improved by about 19.8%, and the events like the goal or the goal kick was detected with high accuracy.

1. Introduction

When professional cameramen shot or edit a sports game, they are required to do it attractively. In order to address this problem by computer, an automatic production system including digital camera work has been researched. This is a video production technology by shooting the sports game by a fixed camera, and then clipping and connecting the video frames. This technology enables us to edit them based on individual preference [4].

The automatic production system is composed of image recognition techniques to track the soccer ball and event recognition. Event recognition is the key issue for digital camera work as well as for retrieving the event and summarizing the whole soccer game. However, event recognition mainly depends on the ball tracking accuracy, because events such as free kick, goal kick, throw in, corner kick and penalty kick are strongly related to the ball.

Many tracking methods have been proposed previously such as mean-shift, Kalman filter, covariance tracker [1] and particle filter. Especially, particle filter was often employed in tracking because of strength to occlusion [2]. In paper [5], the soccer ball was tracked by 2D particle filter. However, 2D positional information is not useful to detect the events. In paper [2], a problem of particle filter, that it is difficult to discover the ball again if it is lost, is not addressed.

In this paper, we propose a system of tracking a ball stably and accurately and detecting the events of the soccer game by 3D location information for automatic soccer video production. Furthermore, in tracking phase, when the local search with a particle filter fails due to the disappearance of the ball, the local search replaced by the global search. Finally the 3D world coordinates are interpolated by assuming dynamical models and the tracking accuracy is improved.

2. Proposed Method

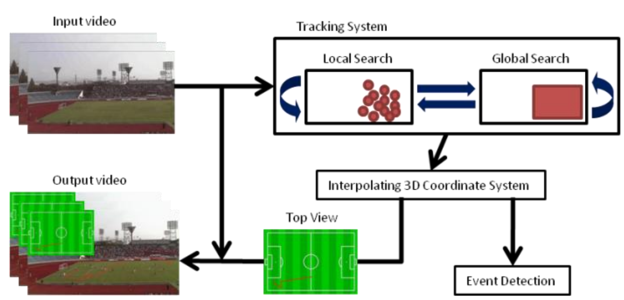

Figure 1: Processing flow of the system.

Figures 1 shows a processing flow of our proposed method. At first, in tracking phase, 3D position of a ball is tracked by switching the local search and the global search on the input video. Secondly, in interpolating phase, the 3D coordinates are interpolated by modeling the motion of the ball. Thirdly, using 3D position interpolated, the events are detected and the converted top view is appended to the output video.

Next, we describe a process of each part.

3. Tracking System

3.1 Local Search

In the local search for the ball, the particle filter is employed. The state x(t) at time t is defined as follows:

$$x = [p_x, p_y, p_z, v_x, v_y, v_z, a_x, a_y, a_z,]^T \qquad (1)$$

ISSN 1975-4736 ©MITA 2009

Here, p, v and a are the positions, velocities and accelerations.

Then, the state transition of the ball from t-α to t is modeled based on three types of ball height; parabolic flight φ_high, bounce φ_bound and decelerating rolling φ_ground due to friction. Furthermore, limiter φ_limit prevents physically impossible particles:

$$x(t) = ((\phi_{high} \otimes \phi_{bound} \otimes \phi_{ground}) \circ \phi_{\lim it})$$
$$x(t-\alpha) + \omega(t-\alpha)$$

$$\phi_{high}(x) = \begin{bmatrix} I_{3*3} & \alpha I_{3*3} & (\alpha^2/2)I_{3*3} \\ O_{3*3} & I_{3*3} & \alpha I_{3*3} \\ O_{3*3} & O_{3*3} & I_{3*3} \end{bmatrix} x$$

$$\phi_{bound}(x) = [p_x, p_y, p_z, v_x, v_y, -ev_z, a_x, a_y, a_z,]^T$$

$$\phi_{ground}(x) = x + \left[0, \cdots, 0, \frac{[-v_x, -v_y]}{\sqrt{{v_x}^2 + {v_y}^2}} \mu g, 0 \right]^T$$

where ω is the Gaussian noise term, $\otimes$ is exclusive OR operation, $\circ$ is composite function, e is coefficient of restitution, μ is dynamic coefficient of friction and g is gravitational acceleration. This model refers to [3].

The likelihood of the ball is computed at each particle by the normalized cross-correlation between the image feature at the particle position and the ball template. The normalized cross-correlation R is defined by Eq. (2). Here, I(x, y) and T(x, y) is gray value at the point (x, y) in the search area and the template image respectively. Ī and Ū are averaged gray value of the search area and the template image respectively.

$$R = \frac{\sum_{i,j} \{I(i,j) - \bar{I}\} \cdot \{U(i,j) - \bar{U}\}}{\sqrt{\sum_{i,j} \{I(i,j) - \bar{I}\}^2 \cdot \sum_{i,j} \{U(i,j) - \bar{U}\}^2}} \quad (2)$$

The perspective projection of the pinhole camera model is needed to project world coordinates (Xw, Yw, Zw) of state x to image coordinates xf, yf as shown in Eq. (3). A is the intrinsic parameters, R is the rotation matrix and t is the translate vector. R and t are called the extrinsic parameters.

$$\begin{bmatrix} x_f \\ y_f \\ 1 \end{bmatrix} = A[R \mid t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3)$$

3.2 Global Search

The particle filter in the local search can robustly track the ball under the occlusion for a short time. However, if it loses the ball, tracking has to be restarted because particles tends to stay at the same position. In this paper, we solve this problem by utilizing template matching of global search.

The system detects the area, where the normalized cross-correlation with the template image is higher than some threshold, as ball candidate. If the multiple candidates are detected, the area where the correlation value is the highest is assumed to be a ball. Then, the system switches global search to local search from the next frame. However, unless the ball is detected, it continues global search.

4. Interpolating Coordination

In a soccer game, a ball sometimes moves randomly and occlusions or frame out occur frequently. As a result, the system sometimes fails in tracking. However, to detect the events, the accurate 3D positional information of the ball is indispensable. In this paper, we propose a system that interpolates the interval when it fails to track the ball or the likelihood is lower than the threshold by assuming dynamical models. We call the interval "interpolated interval".

When the interpolated interval is very short, since the acceleration based on frictional force, gravity and resistive force is able to be ignored, the ball assumed to be in a linear uniform motion. Consequently, the interpolated interval from t+1 to t+α-1 is obtained by the following formula using the velocity vx, vy and vz:

$$\begin{aligned} p_x(t+k) &= p_x(t) + v_x k \\ p_y(t+k) &= p_y(t) + v_y k \\ p_z(t+k) &= p_z(t) + v_z k \\ v_{x,y,z} &= \frac{p_{x,y,z}(t+\alpha) - p_{x,y,z}(t)}{\alpha} \end{aligned} \quad (4)$$

Here, $0 \le k < \alpha$.

The interpolated interval becomes long when the background of the ball is audience and the likelihood degrades rapidly as shown in **Figures 2** .

Figure 2: The ball is lost for a long duration.

ISSN 1975-4736 ©MITA 2009

This figure shows a trajectory of a goal kick and the likelihood of the yellow part is degrades. In this interpolated interval, the ball assumed to be in a parabolic flight. Consequently, the interpolated interval from t+1 to t+α-1 is obtained by the following formula. Coordinates x and y are same as shown in Eq. (4):

$$p_z(t+k) = p_z(t) + v_x(t)k - \frac{1}{2}gk^2$$

5. Event Detection

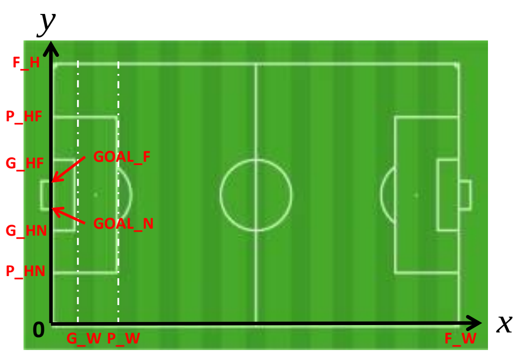

Figure 3: Absolute coordinates of the field.

The events, such as throw-in, goal kick and goal are detected using IF-THEN rule and the absolute coordinates as shown in **Figures 3**. By performing logical operation to px and py in state x, the rules are found manually to judge the occurrence of events denoted in **Table 1**. Then, the events are detected, which occurs in left half of the field because the camera shot the game only in left half of the field as shown in **Figures 2** due to the camera resolutions.

Throw-in is detected when the y position of the ball is less (py < 0) or greater (py > F_H) than the field.

Goal kick is detected when the ball goes out from the field and its y position is not goal ((px < 0) ∩ ((0 < py < GOAL_N) ∪ (GOAL_F < py < F_H))) or when the ball is put down on the goal area for a long time ((0 < px < G_W) ∩ (G_HN < py < G_HF)).

Goal is detected when the ball goes out from the field and its y position is on the goal ((px < 0) ∩ (GOAL_N < py < GOAL_F)) or when it is put down on the center of field for a long time ((p_x, p_y) = (F_W / 2, F_H / 2)).

Table 1: Events

Event	Symbol
Throw-in	TH
Goal kick	GK
Goal	GL

The judge rules are summarized as follows:

$$TH = (p_y < 0) \cup (p_y > F\_H)$$
$$GK = ((p_x < 0) \cap ((0 < p_y < GOAL\_N)$$
$$\cup (GOAL\_F < p_y < F\_H)))$$
$$\cup ((0 < p_x < G\_W) \cap (G\_HN < p_y < G\_HF))$$
$$GL = ((p_y < 0) \cap (GOAL\_N < p_y < GOAL\_F))$$
$$\cup ((p_x, p_y) = (F\_W/2, F\_H/2))$$

6. Experiments

6.1 Experimental Condition

We selected a soccer game that was played during the 38th National High School Soccer Championship (Kyoto area final) in Japan. The size of the image was 1280×720 pixels with 24-bit color. The ball template image size was 15×15 pixels.

6.2 Ball Tracking Experiment

In ball tracking, we compared the proposed method using the previous method with only particle filter for 9 videos clipped from the soccer video. The number of particles were 300, and the frame rate was 30 fps. The results are shown in **Table 2**. In the table, "Tracking accuracy" is the ratio of the number of correctly tracked frames to the number of the total frames. The initial state of the particle filter was given manually.

Table 2: Tracking accuracy

Video clip	Frame	Tracking accuracy	
		Previous(%)	Proposed(%)
shot1	480	18.9	86.5
Shot2	120	39.7	57.9
Shot3	450	95.6	97.7
Shot4	480	53.4	53.6
Shot5	420	78.6	90.6
Shot6	270	12.2	57.3
Shot7	270	80.1	96.4
Shot8	420	60.8	72.2
Shot9	270	80.1	85.8
Average		57.7	77.5

As a result, The tracking accuracy of the proposed method improved by about 19.8¥% on the average. In particular, for shot1 and shot6, the previous method completely lost the ball but the proposed method improved it drastically. It was seen that the condition of the shot (e.g. many

ISSN 1975-4736 ©MITA 2009

occlusions occur, the ball goes out from the field for a long time) heavily influenced the result.

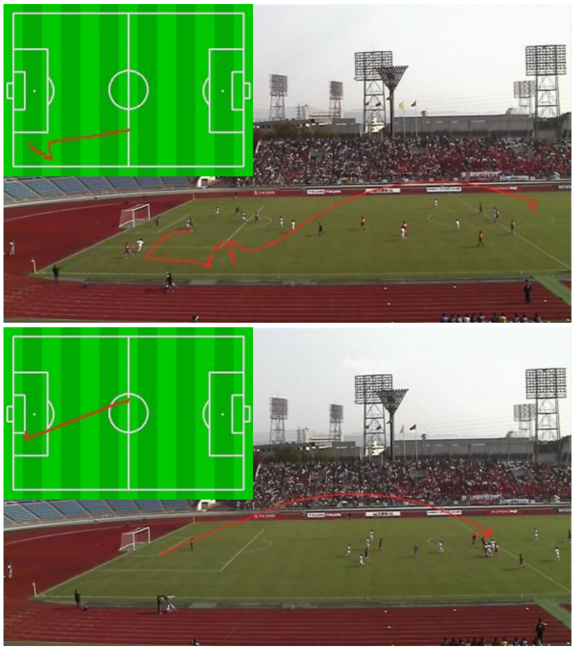

Figure 4: Trajectory of the ball.

Figure 4 shows the trajectory of the ball and its top view. The ball trajectory is successfully drawn even when the back ground is the audience. In the top view drawn by using px and py in state x, it can be seen that the ball trajectory is straight line owing to the coordinates interpolation.

6.3 Event Detection Experiment

The event detection was carried out for the 20 minutes video of the soccer game. The events were detected based on the rules and evaluated by the precision and the recall.

Table 3: Event detection result

Event	Precision	Recall
Throw-in	54%	70%
Goal kick	100%	100%
Goal	100%	100%

Table 3 shows the result. Here, the correctly detected events indicate that the interval detected by the system overlap the interval where the event truly occurred in the game. When a event was repeatedly detected in a short time, they were judged as false events.

As a result, the successful result was obtained for "Goal kick" and "Goal". This is because the position of the ball is clear in the event and the time duration of the events is long. By contrast, the precision as well as the recall of "Throw-in" are not so good because of miss tracking of the ball when it moved around the touchline side or the far side of the field due to the crowded players.

7. Conclusion

In this paper, for automatic soccer video production, we proposed a system that tracked a small ball accurately by 3D particle filter on the fixed Hi-vision video and detected the events by using its 3D positional information. Accurate events recognition was achieved by using particle filter with 3D world coordinates instead of 2D image coordinates. The problem of particle filter, that was difficult to recover again when the ball was lost, was also solved by switching the local search and the global search. Furthermore, by interpolating the lost interval, the tracking accuracy was improved by about 19.8%.

The events such as throw-in, goal kick and goal were detected by logical operations and 3D positional information. The evaluation of the proposed method showed the possibility of the event detection.

We are planning to track the players and detect more difficult events, such as offside, pass, dribbling and foul. IF-THEN rule was subjectively defined, but we would like to challenge the statistical learning of the event rule mining.

8. References

[1] Helio Palaio, Jorge Batista "Multi-Objects Tracking Using an Adaptive Transition Model Particle Filter with Region Covariance Data Association" *IEEE International Conference on Pattern Recognition*, 2008

[2] J. Ren, J. Orwell, G.A. Jones, M. Xu "A general framework for 3D soccer ball estimation and tracking" *IEEE International Conference on Image Processing, pp. 1935-1938*, 2004

[3] Toshihiko Misu, Atsushi Matsui, Masahide Naemura, Mahito Fujii, Nobuyuki Yagi "Distributed Particle Filtering for Multiocular Soccer-ball Tracking" IEEE International Conference on Acoustics, Speech and Signal Processing, 2007

[4] Yasuo Ariki, Shintaro Kubota and Masahito Kumano "Automatic Production System of Soccer Sports Video by Digital Camera Work Based on Situation Recognition" *IEEE international workshop on Multimedia Information Processing and Retrieval, pp.851-858,*, 2006

[5] Yasuo Ariki, Tetsuya Takiguchi and Kazuki Yano "Digital Camera Work for Soccer Video Production with Event Recognition and Accurate Ball Tracking by Switching Search Method" *International Conference on Multimedia and Expo), WE-PM2-L4.1, pp.889-892*, 2008

ISSN 1975-4736 ©MITA 2009