21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

# 3D Tracking of Soccer Players Using Time-Situation Graph in Monocular Image Sequence

Hiroki ITOH† Tetsuya TAKIGUCHI‡ Yasuo ARIKI‡
† *Graduate School of System Informatics, Kobe University*
‡ *Organization of Advanced Science and Technology, Kobe University*
*itoh@me.cs.scitec.kobe-u.ac.jp* {*takigu,ariki*}*@kobe-u.ac.jp*

## Abstract

*In this paper, we propose a new method to track players using 3D particle filter guided by the time-situation graph in order to perform players tracking robust to occlusion in a soccer image sequence. In the conventional method using particle filter, there is a deficit that it is difficult to discover the players again once they are lost in an image sequence. Thus, we represents the position information of two or more players as the time-situation graph beforehand. Then, by running particle filter guided by this graph, the incorrect detection of players can be greatly reduced and the players be robustly tracked even when occlusion occurs. As a result, the tracking accuracy was improved by 7.15 points in comparison with the conventional method.*

## 1. Introduction

Recently, digital photographing by amateurs has been widely spread so that any one can shoot the sports game easily. Therefore, the need for image editing is getting strong, and the automatic image production systems are attracting attention, which support editing an image based on each individual preference.

These systems are composed of image recognition techniques to track the players and ball, event recognition and digital camera work. Event recognition[1] is the key issue for digital camera work as well as for retrieving the event and summarizing the whole soccer game. However, event recognition and digital camera work mainly depend on the players tracking accuracy.

Many tracking methods have been proposed previously such as mean-shift, Kalman filter, covariance tracker and particle filter[4]. Especially, particle filter was often employed in tracking the players in a soccer image sequence. However, when it loses the players once, it is difficult to discover them again since the position information of two or more players is not used between the frames in an image sequence.

In this paper, we represent the position information of two or more players as the time-situation graph beforehand. Then, by running particle filter in the form guided by this graph, we propose a method to reduce the incorrect detection of players greatly and to track players robustly even when occlusion occurs.

This paper is organized as follows. Section 2 describes the flow of the proposed method. In section 3, the details of time-situation graph is described. Section 4 describes the proposed method that tracks the players using 3D particle filter guided by the time-situation graph. In Section 5, the performance of the proposed method is evaluated with the actual image sequence. Section 6 is for paper summarization and discuss about the future work.

## 2. Flow of the Proposed Method

Figure 1 shows a flow of the proposed method which is composed of the time-situation graph construction process and the players tracking process. In the time-situation graph construction process, each player area is first extracted from an input frame by background subtraction. Then, the time-situation graph is constructed by storing into the node the information extracted in the player area. The time-situation graph is a graph whose node includes the number of players (henceforth, it is called the number of components) who exist in a player area extracted by background subtraction at every frame.

In the players tracking process, all players are first detected according to the node information in the time-situation graph. Then, each player is tracked by 3D particle filter in the detection area guided by the time-situation graph.

978-4-9906441-1-6  

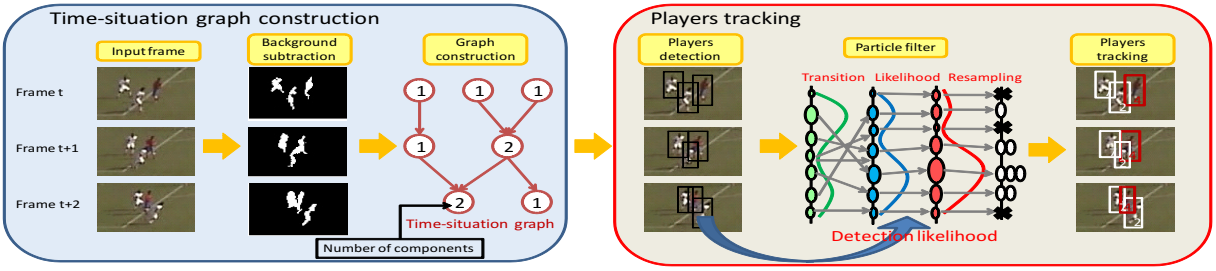


**Figure 1. Flow of the proposed method**

## 3. Time-Situation Graph

We represent the position information of two or more players as the time-situation graph beforehand using the method proposed by Figueroa [3]. In this method, the players, who are occluded each other, can be correctly detected since the number of players who exist in a player area is obtained in a time-situation graph shown in Figure 1(after background subtraction). Then, by running particle filter in the form guided by this graph, it is expected to reduce the incorrect detection of players greatly and to track players robustly even when occlusion occurs.

### 3.1. Time-Situation Graph Information

The time-situation graph is constructed based on a set of player areas extracted at every frame by background subtraction as shown in Figure 1(in graph construction process). Table 1 summarizes the node information and edge information. The node information is composed of the label to identify each node, player area(the number of pixels), player area size(width and height), the center point coordinates of the player area and the number of components. The edge information is composed of the label to identify each edge and the distance between nodes.

The color information was included in the edge information in the method proposed by Figueroa [3] because their method tracked players only using graph. However, the time-situation graph which we propose in this paper detects the players based on the number of components. Therefore, the necessary operation in this graph is only to make the number of components change. From this point, the color information is not included in the edge information in the proposed method.

### 3.2. Construction of the Time-Situation Graph

Let G be a time-situation graph which is a directed graph. $n_i(t)$ is a node labeled $i$ (identification number)

**Table 1. Time-situation graph information**

| Node information | Edge information |
|---|---|
| Label | Label |
| Area | Distance between nodes |
| Size(width and height) | |
| Position | |
| Number of components | |

at frame $t$. $d_{i,j}$ is the distance between nodes $n_i(t)$ and $n_j(t+1)$. $d_{max(i,j)}$ represents the estimated max distance to which a player can move within 1 frame, and this is estimated according to 3D world coordinate position. The distance between nodes is defined by the Euclid distance between the center point coordinates of the player area at frame $t$ and the center point coordinates of the player area at frame $t+1$.

In Figure 2, the distances between nodes about node $n_1(t)$ are shown. The yellow point at frame $t+1$ is the position of node $n_1(t)$ at frame $t$. The red line segments which stretch from the yellow point to the center point coordinates of each nodes at frame $t+1$(the blue points) are the distances between nodes. $e_{i,j}$ is the edge between nodes $n_i(t)$ and $n_j(t+1)$.The algorithm to construct the time-situation graph is defined according to the following steps.

1. Create a node $n_i(t)$ for the node labeled $i$ at the first frame(t = 1), and insert this node into a time-situation graph G.
2. Create a node $n_j(t+1)$ for the node labeled $j$ at frame $t+1$, and insert this node into the time-situation graph G.
3. Calculate the distance $d_{i,j}$ between nodes $n_i(t)$ and $n_j(t+1)$.
4. Create an edge $e_{i,j}$ satisfying the condition $d_{i,j} < d_{max(i,j)}$.
5. Group the nodes, and determine the number of components based on the player area.
6. Repeat steps 2-5 for the whole image sequence.

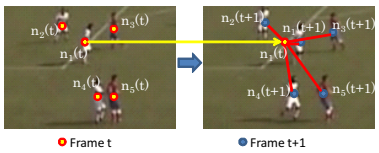


**Figure 2. Distance between nodes**

### 3.3. Determination on the Number of Components

#### 3.3.1 Grouping Nodes

This section describes the method of grouping the nodes by using the edge between them as shown in Figure 3. When node $n_{w_4}(t+1)$, to which edges $e_{v_3,w_4}$ and $e_{v_4,w_4}$ are linked, exists at frame $t+1$, two nodes$n_{v_3}(t)$ and $n_{v_4}(t)$ at frame $t$ belong to the same group as $n_{w_4}(t+1)$. In other words, all the nodes linked by edges belong to the same group.

A new group number is defined for every new group detected on the time-situation graph. For example in Figure 3, starting from node $n_{v_1}(t)$, Group1 is detected, and in the same way starting from node $n_{v_3}(t)$, Group2 is detected.

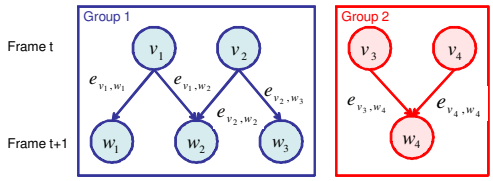


**Figure 3. Grouping nodes**

#### 3.3.2 Determination on the Number of Components

In the proposed method, the determination of the number of components plays on important role because the players are detected based on the number of components. The top of Figure 6 shows the example where three players occluded each other and the bottom of Figure 4 shows the flow of the determination on the number of components. The steps are summarized by the following algorithm.

Algorithm

**[Step1]**$num_X$(the number of components for GroupX) is first defined as the total number of components for each node $n_{v_i}(t)$ in GroupX at frame $t$. The number of components for each node at feame $t+1$ is still 0.

$$num_X = \sum_{v_i \in X} (num_{v_i})$$

Each player area is measured after background subtraction. In Figure 4(a), $num_X$ is set to 3, and the player area are 305, 242 and 139.

**[Step2]**The ideal player area $A_p$ of one player area is estimated by the position information of each node $n_{w_i}(t+1)$ in GroupX at feame $t+1$. Each node $n_{w_i}(t+1)$ is updated by using this ideal player area $A_p$ as follows.

$$A_{w_i} \leftarrow A_{w_i} - A_p$$

In Figure 4(b), The estimated ideal player area $A_p$ is 130, and each player area at frame $t+1$ is updated.

**[Step3]**The number of components $num_X$ is decremented by 1, and the number of components for each node $n_{w_i}(t+1)$ at frame $t+1$ is initialized by 1. This is repeated for every node at frame $t+1$.

$$num_X \leftarrow num_X - 1$$

$$num_{w_i} = 1$$

In Figure 4(c), the number of components $num_X$ for GroupX is decremented by 2, and the number of components for each node at frame $t+1$ is set to 1.

**[Step4]**The following process is repeated to the node $n_{w_i}(t+1)$ whose area $A_{w_i}$ is largest among the players at frame $t+1$ until the number of components $num_X$ for GroupX is 0.

$$A_{w_i} \leftarrow A_{w_i} - A_p$$

$$num_X \leftarrow num_X - 1$$

$$num_{w_i} \leftarrow num_{w_i} + 1$$

In Figure 4(d), the number of components $num_X$ is 0, and each number of components $num_{w_i}$ at frame $t+1$ is finally determined.

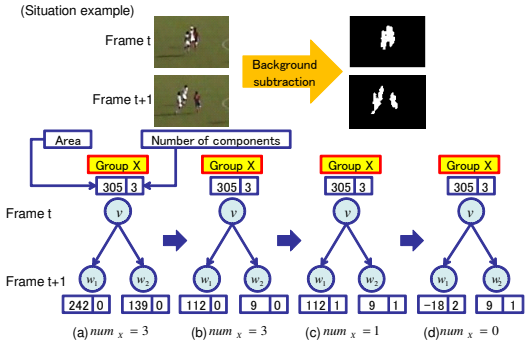


**Figure 4. Flow of the determination on the number of components**

## 4. Tracking Method

### 4.1. Players Detection Based on Node Information

Since the "tracking-by-detection" method [2] which combines players tracking with players detection is employed in this paper, the accuracy of players detection has a great effect on the tracking accuracy. The method of players detection differs according to the number of components contained in the node of the time-situation graph.

In the case where the number of components is 1 or 2, the players are detected based on using player area information of the node. In the case where the number of components is 3 or more, the players are detected by using player area information of the node and SVM. First, the perpendicular 2 players are detected based on the y coordinate between mini($y_{min}$) and max($y_{max}$), and the detection window size estimated by 3D world coordinates position with regards to the center coordinates of the player area included in the node as shown in Figure 5(a) and (b). Then, as shown in Figure 5(c), the other players are detected by using SVM[5] which is widely used with various applications and is known for recognition performance being high.

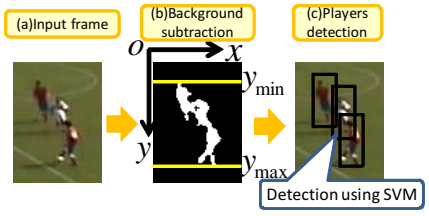


**Figure 5. Players detection**

### 4.2. Players Tracking by 3D Particle Filter Using Time-Situation Graph

We employed a 3D particle filter for each player tracking. The state $\vec{x}_p(t)$ at time $t$ is defined as follows:

$$\vec{x}_p(t) = [p_x, p_y, v_x, v_y, a_x, a_y]^T \quad (1)$$

Here, $p$, $v$ and $a$ are the positions, velocities and accelerations at time $t$. Then, the state transition of the player is modeled based on linear motion with uniform acceleration.

The likelihood $\omega$ of the player is computed at each particle by the weighted sum as shown in Eq (2) using three kinds of likelihoods (the likelihood $p_D$ based on the detection result by the node information and SVM, the likelihood $p_H$ based on the histogram and the likelihood $p_C$ based on the cross-correlation). Here, $\alpha$ is defined as the parameter showing the degree of occlusion, and computed based on the node information(area) obtained when the player was detected. $\beta$ is defined as the value obtained by converting the distance between the target player and the nearest player into the probability of the normal distribution.

$$\omega = \alpha \cdot p_D + \beta \cdot p_H + (1 - \beta) \cdot p_C \quad (2)$$

## 5. Experimental Evaluation

### 5.1. 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.

### 5.2. Result and Discussion

In players tracking, we compared the proposed method using the time-situation graph with the conventional method without the time-situation graph for 10 videos (the average number of frames : 350) clipped from the soccer video. The tracking space is the only left half of the field. Therefore, we selected the sample videos in which the tracking players almost play in the left half of the field since the players can not be tracked when they come out into the right half of the field.

The players detection by the conventional method employs SVM to the whole soccer field. The number of particles were 150, and the frame rate was 30 fps. The results are shown in Figure 6. In Figure 6, "Tracking accuracy" is the ratio of the number of correctly tracked frames to the number of the total frames. Tracking accuracy by the proposed method are almost higher than that of the conventional method which does not use the time-situation graph.

As a result, the tracking accuracy of the proposed method is improved by 7.15 points on the average as shown in Table 2. As can be seen from this result, since the accuracy of tracking players is higher by using time-situation graph when occlusion occurs, players tracking becomes more robust than the conventional method. Tracking failure occurs when the players of the same team cross so that the particles of the players are reversed each other or 2 particles overlap to 1 player. For the former, it is considered that the likelihood of particle filter does not decrease due to the same uniform color of the same team even if tracking is reversed. For the latter, it is considered that the number of components

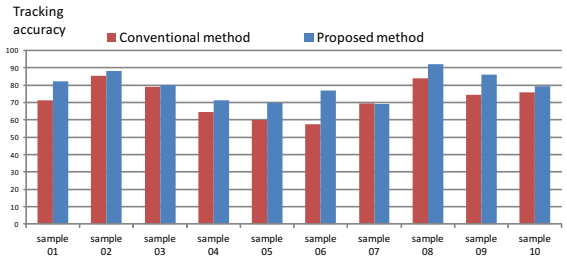


**Figure 6. Comparison between the two methods**

in the time-situation graph becomes wrong. Compared with the conventional method, the proposed method is inferior in the point the players with no movement for long frames may not be detected by background subtraction. However, the proposed method can more correctly detect the players than the conventional method when their occlusion occurs.

**Table 2. Averaged tracking accuracy**

| Method | $A_p(\%)$ |
|---|---|
| Conventional method | 72.15 |
| Proposed method | 79.50 |

Figure 7 shows the tracking results clipped at fixed intervals. It turns out that the occluded players are tracked correctly.

## 6. Conclusion

In this paper, we proposed a new method that tracks the players using 3D particle filter guided by the time-situation graph in order to perform robust players tracking to occlusion in a soccer image sequence. As a result, the tracking accuracy was improved by 7.15 points in comparison with the conventional method without the time-situation graph.

In the future, in order to deal with the failure in the number of components, we will improve tracking in the region where players are occluded each other, for example, inside the penalty area. It is easy to make a failure in the number of components since the occlusion between players tends to occur in this region. Therefore, we are planning to develop the algorithm to correct the number of components in the time-situation graph and introduce the constraint with the number of components (there are 22 players and 1 referee on soccer field) by tracking players on the whole field shot by multiple cameras.

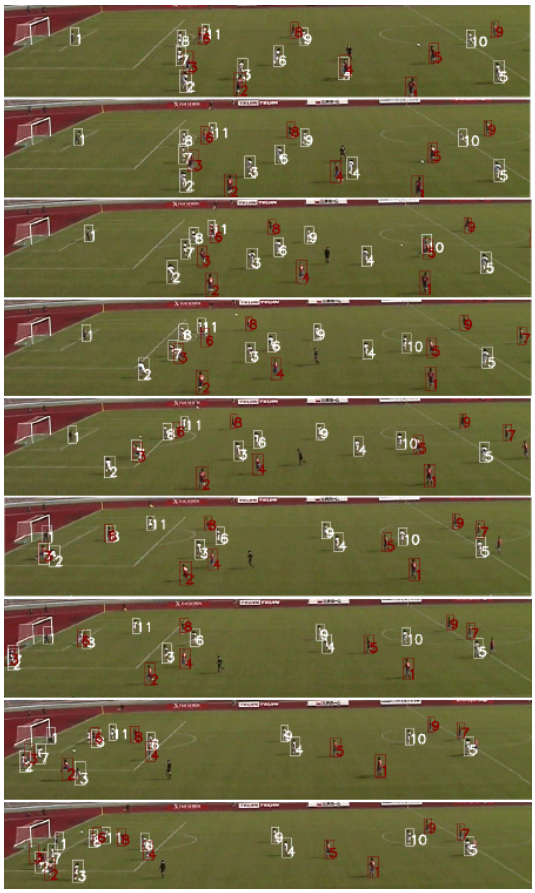

**Figure 7. Tracking result**

## References

[1] V. Tovinkere and R. J. Qian, "Detecting Semantic Events in Soccer Games: Towards A Complete Solution", IEEE ICME, pp. 1040-1043, 2001.

[2] M. Breitenstein, F. Reichin, B. Leibe, E. Koller-Meier and L. V. Gool, "Robust tracking-by-detection Using a Detector Confidence Particle Filter", The 12th IEEE ICCV, pp. 1515-1522, 2009-9.

[3] Pascual J. Figueroa a, Neucimar J. Leite and Ricardo M.L. Barros, "Tracking soccer players aiming their kinematical motion analysis", CVIU, pp. 122-135, 2005.

[4] Takuro Nishino, Yasuo Ariki and Tetsuya Takiguchi, "Tracking of Multiple Soccer Players Using a 3D Particle Filter Based on Detector Confidence", ACSE,pp. 93-104, 2011.

[5] M.J. Vapnik, "The Nature of Statistical Learning Theory", Springer, Heidelberg, 2001.