PAPER

Player Tracking in Far-View Soccer Videos Based on Composite Energy Function

Kazuya IWAI^{†a)}, *Student Member*, Sho TAKAHASHI^{†b)}, Takahiro OGAWA^{†c)},
and Miki HASEYAMA^{†d)}, *Members*

SUMMARY In this paper, an accurate player tracking method in far-view soccer videos based on a composite energy function is presented. In far-view soccer videos, player tracking methods that perform processing based only on visual features cannot accurately track players since each player region becomes small, and video coding causes color bleeding between player regions and the soccer field. In order to solve this problem, the proposed method performs player tracking on the basis of the following three elements. First, we utilize visual features based on uniform colors and player shapes. Second, since soccer players play in such a way as to maintain a formation, which is a positional pattern of players, we use this characteristic for player tracking. Third, since the movement direction of each player tends to change smoothly in successive frames of soccer videos, we also focus on this characteristic. Then we adopt three energies: a potential energy based on visual features, an elastic energy based on formations and a movement direction-based energy. Finally, we define a composite energy function that consists of the above three energies and track players by minimizing this energy function. Consequently, the proposed method achieves accurate player tracking in far-view soccer videos.

key words: soccer video, player tracking, energy function, elastic model

1. Introduction

Due to the widespread use of video distribution services and high-capacity recording media, there is a demand for efficient video retrieval systems. To develop such systems, various studies for analyzing videos semantically have been conducted [1]–[8]. In sports videos, which attract a huge number of viewers around the world, many researchers have reported methods for indexing or extracting specific events [3]–[8]. In order to perform the semantic analysis in soccer videos, player positions need to be provided.

For this purpose, various player tracking methods in soccer videos have been proposed [9]–[15]. Some conventional methods track players in broadcast soccer videos [9]–[12]. Furthermore, player tracking methods in multiple view soccer videos have been proposed [13]–[15]. These conventional methods first perform processing that utilizes only visual features, such as estimation of uniform colors and player detection, and then track each player based on the obtained results. As well as broadcast soccer videos and multi-

Fig. 1 Example of far-view soccer videos.

ple view soccer videos, there are far-view soccer videos, an example of which is shown in Fig. 1. These videos are more useful for advanced video analysis since almost all of the players are included in the videos. Therefore, most professional soccer teams use far-view soccer videos for coaching and scouting. However, there is no player tracking method in such videos. Furthermore, since each player region becomes small and color bleeding between player regions and the soccer field is caused by video coding, it becomes difficult to accurately track players in these videos by using conventional methods that first perform processing based only on visual features. Specifically, the conventional methods lose track of target players in far-view soccer videos. Furthermore, when a target player is close to another player of the same team, the conventional methods mix up them. For example, different target players are incorrectly tracked with each other and one player is simultaneously tracked by multiple trackers.

In order to solve the above problem, an accurate player tracking method in far-view soccer videos based on a composite energy function is proposed in this paper. Since the use of only visual features is not sufficient to accurately track players in far-view soccer videos, the proposed method also utilizes the following two characteristics. In soccer games, players play in such a way as to maintain a formation, which is a positional pattern of players. Thus, we focus on this characteristic and call it a positional characteristic hereafter. Furthermore, since the movement direction of each player tends to change smoothly in successive frames of soccer videos, we also focus on this characteristic and call it a movement characteristic hereafter. In order to track players on the basis of visual features and these char-

Manuscript received September 9, 2013.

Manuscript revised January 3, 2014.

[†]The authors are with the Graduate School of Information Science and Technology, Hokkaido University, Sapporo-shi, 060–0814 Japan.

a) E-mail: iwai@lmd.ist.hokudai.ac.jp

b) E-mail: sho@lmd.ist.hokudai.ac.jp

c) E-mail: ogawa@lmd.ist.hokudai.ac.jp

d) E-mail: miki@ist.hokudai.ac.jp

DOI: 10.1587/transinf.E97.D.1885

acteristics, the proposed method defines a composite energy function that consists of the following three energies. The first one is a potential energy, which utilizes visual features. This energy is defined on the basis of uniform colors and player shapes, and it represents the possibility of a target player region. Secondly, we adopt an elastic energy, which utilizes the positional characteristic. In order to consider maintenance of a formation, the proposed method introduces an elastic model obtained by connecting players with springs into player tracking. Then we define an elastic energy based on a change in this model's shape, which enables us to achieve player tracking using the characteristic of formations. Thirdly, we also adopt a movement direction-based energy, which utilizes the movement characteristic. This energy inhibits a sudden change in the movement direction of each player in successive frames. By minimizing the energy function defined by the above three energies, i.e., potential energy, elastic energy and movement direction-based energy, we perform player tracking. Therefore, we can track players based on visual features, positional characteristic and movement characteristic. Consequently, the proposed method achieves accurate player tracking in far-view soccer videos. It should be noted that the method presented in this paper is an extended version of our previous method [16]*.

This paper is organized as follows. In Sect. 2, a soccer field detection method based on camera parameter estimation as preprocessing for our method is explained. In Sect. 3, we describe the player tracking method based on the composite energy function. Sect. 4 shows experimental results to verify the effectiveness of the proposed method.

2. Preliminaries

In this section, we explain a soccer field detection method based on camera parameter estimation as preprocessing for our method. Before performing player tracking, we have to detect the actual field area (world coordinates) corresponding to the captured field area (screen coordinates) in the video. By using our previously reported method [17], the correspondence between the actual field area and the captured field area is obtained. This method first extracts field lines from a soccer video. Then camera parameters are estimated by matching the extracted field lines with a wire frame model, which represents the official layout of the soccer field lines. By using the estimated parameters, the correspondence between the actual field area and the captured field area is obtained.

In the proposed method, we restrict the area for player tracking to the area within the field lines on the wire frame model obtained by using the camera parameters estimated by the method [17]. It should be noted that the proposed method requires player positions not only in screen coordi-

nates but also in world coordinates, since we focus on formations and movement directions of players. Therefore, we obtain player positions in the world coordinates based on transformation from screen coordinates to world coordinates using the estimated camera parameters. In this paper, we describe $\mathcal{T}(\cdot)$ as the function that performs the above transformation. The input of $\mathcal{T}(\cdot)$ is a player position in the screen coordinates and the output is the player position in the world coordinates. The details of the procedures are shown in [17]. Note that, in this paper, we manually modify incorrect camera parameters estimated by the method [17] beforehand in order to avoid their influence on accuracy of the proposed method.

3. Player Tracking Based on Composite Energy Function

In this section, we explain the player tracking method based on a composite energy function. The proposed method utilizes not only visual features but also positional and movement characteristics. In order to track players utilizing the positional characteristic, we adopt an elastic model obtained by connecting players with springs. Furthermore, for realizing player tracking on the basis of multiple elements, we define a composite energy function. Specifically, we adopt potential energy $E^p(\mathbf{Y}^t)$ calculated on the basis of uniform colors and player shapes (see 3.1), elastic energy $E^s(\mathbf{Y}^t)$ calculated on the basis of a change in the elastic model's shape (see 3.2) and movement direction-based energy $E^d(\mathbf{Y}^t)$ (see 3.3). Note that $\mathbf{Y}^t$ is a set of search points in the screen coordinates, which is defined as $\mathbf{Y}^t = \{\mathbf{y}_1^t, \mathbf{y}_2^t, \dots, \mathbf{y}_{10}^t\}^\dagger$. The search point $\mathbf{y}_i^t$ ($i = 1, 2, \dots, 10$) corresponds to player i in the t -th frame ($t = 1, 2, \dots, T$; T is the number of frames).

By using the above three energies, we define the composite energy function $E(\mathbf{Y}^t)$ as follows:

$$E(\mathbf{Y}^t) = E^p(\mathbf{Y}^t) + \alpha E^s(\mathbf{Y}^t) + \beta E^d(\mathbf{Y}^t), \quad (1)$$

where α and β are parameters that control the influence of elastic energy $E^s(\mathbf{Y}^t)$ and movement direction-based energy $E^d(\mathbf{Y}^t)$. In the proposed method, we track players by minimizing the energy function $E(\mathbf{Y}^t)$ (see 3.4).

3.1 Potential Energy Based on Uniform Colors and Player Shapes

In this subsection, we explain potential energy, which represents the possibility of a target player region. In soccer games, players of one team wear uniforms with different colors from the uniform colors of the other team according to Fédération Internationale de Football Association (FIFA) laws of the game [18]. Therefore, by taking uniform colors into account, we can distinguish players of the two teams and track players of the target team. Furthermore, we also

*The proposed method introduces the following two novel approaches: i) restricting an area for player tracking by detecting the field area and ii) representing the possibility of a target player region by using both uniform colors and player shapes.

[†]In this paper, target objects are all 11 players of the target team except for the goal keeper.

focus on player shapes, since it is difficult to accurately represent the possibility of a target player region using only uniform colors in far-view soccer videos. Thus, in the proposed method, we utilize not only hue in the HSV color space but also Histograms of Oriented Gradients (HOG) descriptor [19] for representing color information and shape information.

In the proposed method, we first provide hue values of the previously known target team's uniform H_0^n ($n = 1, 2, \dots, N$; N is the number of colors included in the target team's uniform). Next, we calculate the dissimilarity between H_0^n and $H(\mathbf{y}^t)$, where $H(\mathbf{y}^t)$ is the hue value at position $\mathbf{y}^t$ in the t -th frame. Then we obtain the dissimilarity based on uniform colors $D^{color}(\mathbf{y}_i^t)$ as follows:

$$D^{color}(\mathbf{y}_i^t) = \sum_{\mathbf{y}^t \in R_i^t} Dissim(\mathbf{y}^t), \quad (2)$$

where R_i^t is a set of pixels in a rectangle, which is located with $\mathbf{y}_i^t$ in the center of its base. The size of the rectangle is $r_w \times r_h$ [pixels]. Furthermore,

$$Dissim(\mathbf{y}^t) = \min_{n=1,2,\dots,N} \left\{ 1 - \cos(H_0^n - H(\mathbf{y}^t)) \right\}. \quad (3)$$

Next, given a rectangle centered at $\hat{\mathbf{y}}_i^{t-1}$, which is player i 's position in the $(t-1)$ -th frame already obtained by applying our tracking method to the previous frame $t-1$, we calculate the dissimilarity of the HOG descriptor between the regions within the respective rectangles in the current frame. Specifically, the dissimilarity is calculated by using the Bhattacharyya coefficient as follows:

$$D^{hog}(\mathbf{y}_i^t) = \sqrt{1 - \rho(\mathbf{h}(\mathbf{y}_i^t), \mathbf{h}(\hat{\mathbf{y}}_i^{t-1}))}, \quad (4)$$

$$\rho(\mathbf{h}(\mathbf{y}_i^t), \mathbf{h}(\hat{\mathbf{y}}_i^{t-1})) = \sum_{u=1}^M \sqrt{h_u(\mathbf{y}_i^t) h_u(\hat{\mathbf{y}}_i^{t-1})}, \quad (5)$$

where $\mathbf{h}(\mathbf{y}_i^t)$ is the HOG descriptor in the region within the rectangle located at $\mathbf{y}_i^t$, which is defined as $\mathbf{h}(\mathbf{y}_i^t) = \{h_u(\mathbf{y}_i^t)\}_{u=1,\dots,M}$, $\sum_{u=1}^M h_u(\mathbf{y}_i^t) = 1$. Note that M is the number of histogram bins.

Finally, we define potential energy $E^p(\mathbf{Y}^t)$ based on the two dissimilarities as follows:

$$E^p(\mathbf{Y}^t) = \sum_{i=1}^{10} e^p(\mathbf{y}_i^t), \quad (6)$$

$$e^p(\mathbf{y}_i^t) = D^{color}(\mathbf{y}_i^t) + \gamma \frac{\mu_i^{color}}{\mu_i^{hog}} D^{hog}(\mathbf{y}_i^t), \quad (7)$$

where μ_i^{color} and μ_i^{hog} are respectively the mean values of $D^{color}(\mathbf{y}_i^t)$ and $D^{hog}(\mathbf{y}_i^t)$ in a search area for player i , where the details of the search area are shown in 3.4. Furthermore, γ is a parameter that controls the influence of $D^{hog}(\mathbf{y}_i^t)$. By using potential energy calculated above, the proposed

method can focus on players of the target team on the basis of uniform colors and player shapes.

3.2 Elastic Energy Based on Formations

In this subsection, we explain an elastic model and then define elastic energy to track players based on formations. In soccer games, players play in such a way as to maintain a formation. For realizing player tracking utilizing this characteristic, we adopt an elastic model obtained by connecting players with springs. Note that, in order to perform various tactics in soccer games, players generally form a group with nearby players. Therefore, the proposed method defines the elastic model using tactics-based groups, which are obtained from the previous tracking results by using the player clustering method [20]. Then elastic energy is calculated as follows:

$$E^s(\mathbf{Y}^t) = \sum_{i=1}^9 \sum_{j=i+1}^{10} \frac{1}{2} K(G_i^t, G_j^t) \left\{ \|\mathcal{T}(\mathbf{y}_i^t) - \mathcal{T}(\mathbf{y}_j^t)\| - \|\mathcal{T}(\hat{\mathbf{y}}_i^{t-1}) - \mathcal{T}(\hat{\mathbf{y}}_j^{t-1})\| \right\}^2, \quad (8)$$

where G_i^t represents the group that player i belongs to in the t -th frame. Furthermore, $K(G_i^t, G_j^t)$ is the strength of the spring between player i and player j defined as follows:

$$K(G_i^t, G_j^t) = \begin{cases} 1 & (G_i^t = G_j^t) \\ \exp\left(-\frac{\|\mathcal{T}(\hat{\mathbf{y}}_i^{t-1}) - \mathcal{T}(\hat{\mathbf{y}}_j^{t-1})\|}{\eta}\right) & (G_i^t \neq G_j^t) \end{cases}, \quad (9)$$

where η is a constant. By using Eq. (9), a spring between the players belonging to the same group has greater strength. On the other hand, the strength of a spring between players belonging to different groups becomes weaker as their distance becomes greater. This represents the fact that players who are located near each other are more related in terms of soccer tactics than players who are located far from each other.

Generally, it becomes difficult to accurately track players in far-view soccer videos from only the visual features shown in the previous subsection. Therefore, we adopt player tracking based on the positional characteristic by using elastic energy.

3.3 Movement Direction-Based Energy

In this subsection, we explain movement direction-based energy. In successive frames of soccer videos, the movement direction of each player tends to change smoothly. Thus, the proposed method defines an energy that maintains the movement direction of each player in successive frames. Specifically, we first calculate an average movement vector $\bar{\mathbf{v}}_i^{t-1}$ based on tracking results of player i in the past F frames as follows:

$$\bar{v}_i^{t-1} = \frac{1}{F} \sum_{l=1}^F \frac{\mathcal{T}(\hat{y}_i^{t-l}) - \mathcal{T}(\hat{y}_i^{t-l-1})}{\|\mathcal{T}(\hat{y}_i^{t-l}) - \mathcal{T}(\hat{y}_i^{t-l-1})\|}. \quad (10)$$

Next, we calculate a movement vector v_i^t based on the tracking result of player i in the $(t-1)$ -th frame and a position of a search point corresponding to the player in the t -th frame as follows:

$$v_i^t = \mathcal{T}(y_i^t) - \mathcal{T}(\hat{y}_i^{t-1}). \quad (11)$$

Finally, by using the average movement vector $\bar{v}_i^{t-1}$ and the movement vector v_i^t , we define movement direction-based energy $E^d(Y^t)$ as follows:

$$E^d(Y^t) = \sum_{i=1}^{10} e^d(y_i^t), \quad (12)$$

$$e^d(y_i^t) = 1 - \frac{\bar{v}_i^{t-1} \cdot v_i^t}{\|\bar{v}_i^{t-1}\| \|v_i^t\|}, \quad (13)$$

where $\bar{v}_i^{t-1} \cdot v_i^t$ denotes the inner product of $\bar{v}_i^{t-1}$ and v_i^t . By using the above energy, we can track players inhibiting a sudden change in the movement direction of each player in successive frames. Therefore, the proposed method performs robust player tracking even when a target player is close to another player of the same team or occluded.

3.4 Player Tracking Algorithm

In this subsection, we explain player tracking using the composite energy function. In the proposed method, we manually give an initial position of a search point for each player in the first frame. Then we utilize tracking results obtained in each target frame as initial positions of search points in the following frame. Tracking results are obtained by finding positions of search points $\hat{Y}^t (= \{\hat{y}_1^t, \hat{y}_2^t, \dots, \hat{y}_{10}^t\})$ that minimize the composite energy function $E(Y^t)$ in the area within the field lines shown in Sect. 2. Since a distance that a player can cover between two successive frames is limited, we set the search area with size of $s_w \times s_h$ [pixels] for each player. Specifically, $\hat{Y}^t$ can be obtained as follows:

$$\hat{Y}^t = \arg \min_{Y^t} E(Y^t). \quad (14)$$

We use simulated annealing [21] to minimize the energy function. Then player tracking using the above procedure becomes feasible.

In the proposed method, by minimizing the composite energy function based on visual features, positional characteristic and movement characteristic, accurate player tracking in far-view soccer videos can be expected.

4. Experimental Results

In this section, experimental results are shown to verify the effectiveness of the proposed method. For the experiments,

Table 1 Details of comparative methods.

Method	Description of the method
Method 1	Player tracking method using the standard particle filter algorithm [22]
Method 2	Player tracking method based only on uniform colors in the proposed method
Method 3	Player tracking method using potential energy in the proposed method
Method 4	Player tracking method using potential energy and elastic energy in the proposed method
Method 5	Player tracking method using potential energy and movement direction-based energy in the proposed method
Method 6	Our previous method [16]

Table 2 Details of target videos for experiments shown in [23]–[26].

Reference	Total number of frames	Frame rate
[23]	11,705	29.97 fps
[24]	3,599	25 fps
[25]	1,500	15 fps
[26]	815	25 fps

Table 3 Details of the target team for each clip.

Clip	Team	Uniform colors(shirt / shorts / socks)
Clips 1-12	Team A	Red & Black / Black / Red
Clips 13-15	Team B	Light blue / Light blue / Light blue
Clips 16-18	Team C	Yellow / Yellow / Yellow

we captured video clips from three kinds of far-view soccer videos, where each captured frame was a 24-bit color image of 720×480 pixels, and the frame rate was 30 fps. The proposed method and the comparative methods shown in Table 1 were applied to 18 clips, where the total number of frames was 15,638. Note that [23]–[25] and [26] show experimental results of player tracking obtained by applying their methods to soccer videos shown in Table 2 in order to evaluate their effectiveness. In contrast, we conducted the experiments using many more frames of videos than their test data.

For each clip of far-view soccer videos, the tracking target was set as players belonging to the team shown in Table 3. The parameters of the proposed method were set as $\alpha = 0.8$, $\beta = 0.2$ and $\gamma = 0.3$. We experimentally determined the best parameters by using a clip that was not included in the target clips. In the experiments, the size of rectangles was set as $r_w = 12$, $r_h = 24$ in such a way that they circumscribe player regions. Furthermore, the size of the search area for each player was set as $s_w = 4$, $s_h = 4$ in terms of the distance that players can cover between any two successive frames. In addition, the HOG descriptor was calculated with $M = 360$. We also set $\eta = 70$ and $F = 30$. Note that the same parameters were adopted for all of the target clips in the experiments. Introduction of an automatic parameter setting is our future work.

In the experiments, we checked tracking results every 30 frames for the target clips. If some target players were not tracked in a frame, we manually gave their correct positions and then performed player tracking using the given positions

Fig. 2 Tracking results obtained by the proposed method. The tracking targets are players belonging to Team A.

Fig. 3 Tracking results: (a-1)-(a-3) Ground truth, (b-1)-(b-3) Results obtained by method 1, (c-1)-(c-3) Results obtained by the proposed method. The tracking targets of (*-1), (*-2) and (*-3) are players belonging to Teams A, B and C, respectively.

after that frame. Note that the number of modifications that are necessary for each method is examined in the following quantitative evaluation. In the manual checks for correcting positions and evaluating the performance, we actually distinguished tracked players. A part of the tracking results obtained by using the proposed method is shown in Fig. 2.

In this figure, we tracked players belonging to team A. We also show rectangles, with size of $r_w \times r_h$ pixels, located on positions of the tracking results. From this figure, we can see that all of the target players are accurately tracked by using the proposed method in far-view soccer videos, although each player region is small and color bleeding be-

Table 4 Accuracy of the tracking results obtained by the proposed method and the comparative methods.

Target video	Frame	Target team	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Proposed method
Clip 1	300	Team A	0.900	0.900	1.000	1.000	1.000	1.000	1.000
Clip 2	300	Team A	0.700	0.700	0.800	1.000	0.800	1.000	1.000
Clip 3	570	Team A	0.579	0.579	0.579	0.842	0.737	0.842	0.895
Clip 4	580	Team A	0.684	0.789	0.789	0.895	0.842	0.947	1.000
Clip 5	582	Team A	0.526	0.684	0.684	0.789	0.737	0.947	1.000
Clip 6	630	Team A	0.810	0.714	0.857	0.905	0.857	0.857	1.000
Clip 7	780	Team A	0.692	0.808	0.846	0.846	0.846	0.962	1.000
Clip 8	807	Team A	0.654	0.577	0.692	0.846	0.769	0.885	0.962
Clip 9	930	Team A	0.710	0.774	0.839	0.839	0.839	1.000	1.000
Clip 10	990	Team A	0.727	0.818	0.818	0.848	0.818	1.000	1.000
Clip 11	1062	Team A	0.571	0.771	0.800	0.886	0.829	0.886	1.000
Clip 12	1350	Team A	0.689	0.689	0.778	0.867	0.800	0.911	0.978
Clip 13	511	Team B	0.588	0.647	0.706	0.706	0.706	0.706	0.882
Clip 14	720	Team B	0.625	0.708	0.792	0.792	0.792	0.833	0.917
Clip 15	1136	Team B	0.649	0.703	0.784	0.811	0.784	0.838	0.919
Clip 16	1027	Team C	0.765	0.735	0.794	0.941	0.824	1.000	1.000
Clip 17	1639	Team C	0.741	0.759	0.759	0.796	0.759	0.907	0.907
Clip 18	1724	Team C	0.667	0.772	0.825	0.860	0.825	0.930	0.982
Total	15638		0.681	0.733	0.785	0.851	0.805	0.915	0.969

Table 5 Numbers of frames and players that are necessary to correct for the proposed method and the comparative methods.

Method	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Proposed method
Number of frames	165	138	111	77	101	44	16
Number of players	327	200	151	101	145	89	43

tween player regions and the soccer field occurs.

Next, in order to verify the effectiveness of the proposed method, we compare our method with method 1. Many researchers have proposed player tracking methods using the particle filter [24], [27], [28]. Therefore, we use method 1 based on the standard particle filter algorithm [22] as a comparative benchmarking method. According to the method [27], method 1 tracks players by initializing an independent tracker for each player, and the likelihood of each particle is calculated on the basis of a HSV color histogram. Tracking results obtained by using method 1 and the proposed method are shown in Fig. 3. Note that each image is a zoomed portion of the target frame, in which remarkable players are included. From Figs. 3 (b-1)-(b-3), we can see that some target players are not tracked by using method 1. Since method 1 utilizes only uniform colors for player tracking, it is difficult to obtain accurate tracking results in far-view soccer videos. On the other hand, as shown in Figs. 3 (c-1)-(c-3), the proposed method accurately tracks all of the target players. This is due to player tracking focusing on not only visual features but also positional and movement characteristics.

We also quantitatively evaluate the performance of the proposed method. If all of the target players were correctly tracked in a checked frame, we regarded this frame as a correct frame. Then we define *Accuracy* as follows:

$$Accuracy = \frac{\text{Num. of correct frames}}{\text{Num. of all checked frames}}. \quad (15)$$

In the experiments, the number of all checked frames was

517 and the number of target players in each frame was 10, i.e., the number of target players in all checked frames was 5,170. Table 4 shows *Accuracy* of the proposed method and the comparative methods. We also show how often we corrected the positions through the experiments to achieve perfect tracking of all of the target players, i.e., *Accuracy* in Eq. (15) is 1.000. Table 5 shows the numbers of frames and players that are necessary to correct for the proposed method and the comparative methods. From these tables, we discuss the effectiveness of each proposed energy and the proposed method below.

Discussion of the effectiveness of introducing the HOG descriptor into potential energy

We can see that method 3 performs player tracking more accurately than does method 2. This is because method 3 represents the possibility of a target player region by focusing on player shapes in addition to uniform colors. Thus, the effectiveness of introducing the HOG descriptor into potential energy in the previous method [16] (method 6) was verified.

Discussion of the effectiveness of elastic energy

We can see that method 4 performs player tracking more accurately than does method 3. In scenes where a target player is close to another player of the same team and the target players are far from the camera, tracking results obtained by using method 4 are often more accurate than results obtained by using method 3. This is because method 4 can track players maintaining a formation by using elastic energy, although there is not enough visual information to accurately track players. Thus, the effectiveness of adopting

Table 6 Average computational time (sec) per frame of the proposed method and the comparative methods.

Method	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Proposed method
Computational time	0.01	0.33	0.53	0.55	0.54	0.42	0.56

elastic energy based on the elastic model was verified.

Discussion of the effectiveness of movement direction-based energy

We can see that method 5 performs player tracking more accurately than does method 3. Particularly in scenes where a target player is close to another player of the same team or occluded, tracking results obtained by using method 5 are more accurate than results obtained by using method 3 since method 5 can track players inhibiting a sudden change in the movement direction of each player in successive frames by using movement direction-based energy. Thus, the effectiveness of adopting movement direction-based energy was verified.

Discussion of the effectiveness of the proposed method

We can see that the proposed method performs player tracking more accurately than does method 6. In scenes where the target players are close to signboards with colors similar to the uniform colors of the target team, performance improvement of player tracking is confirmed. Thus, the effectiveness of introduction of restricting an area for player tracking by using a soccer field detection method was verified. Furthermore, we can see that *Accuracy* of the proposed method is higher than that of all of the comparative methods. This result verifies the effectiveness of player tracking based on uniform colors, player shapes, maintenance of a formation and the movement direction of each player, which is achieved by using the composite energy function.

From these experimental results, we can conclude that the proposed method achieves accurate player tracking in far-view soccer videos by tracking players on the basis of visual features, positional characteristic and movement characteristic. On the other hand, it becomes difficult to accurately track players when the proposed method is applied to highly complex scenes such as corner kicks. In [27], when a player is close to other players, the center of gravity of each player is resampled on the basis of detected results of the players obtained by Real AdaBoost [29] and movement directions and velocity of the players. By introducing similar approaches into our method, accurate player tracking can be expected in such scenes. This will be our future work.

Finally, we compare the computational time between the proposed method and the comparative methods. Table 6 shows the average computational time per frame of the proposed method and the comparative methods. This result was obtained by applying the proposed method and the comparative methods to clips 1-18. Note that these experiments were performed on a computer using Inter(R) Core(TM) i7-950 (3.06GHz) with 8 Gbytes RAM. From this table, we can see that the proposed method requires more time than other comparative methods. Then we will need to reduce the computational time of the proposed method in our future work.

5. Conclusions

In this paper, we have proposed an accurate player tracking method in far-view soccer videos based on a composite energy function. The proposed method utilizes not only visual features but also positional and movement characteristics. Particularly, in order to track players utilizing the positional characteristic, we adopt an elastic model obtained by connecting players with springs. Furthermore, for realizing player tracking on the basis of multiple elements, we define a composite energy function. Then we perform player tracking by minimizing this energy function. Consequently, the proposed method achieves accurate player tracking in far-view soccer videos. Our experimental results verified the effectiveness of the proposed method. By using our tracking method, we can obtain all of the player positions of a target team included in far-view soccer videos. Therefore, more effective semantic analysis of soccer videos can be expected by using the obtained results.

Acknowledgements

The video images used in this paper were provided by Hokkaido Football Club Corporation. This work was partly supported by Grant-in-Aid for Scientific Research (B) 25280036 from JSPS.

References

- [1] I. Ide, K. Yamamoto, R. Hamada, and H. Tanaka, "An automatic video indexing method based on shot classification," IEICE Trans. Inf. & Syst. (Japanese Edition), vol.J82-D-II, no.10, pp.1543-1551, Oct. 1999.
- [2] H.R. Naphide and T.S. Huang, "A probabilistic framework for semantic video indexing, filtering and retrieval," IEEE Trans. Multimed., vol.3, no.1, pp.141-151, March 2001.
- [3] H. Miyamori, "Automatic annotation of tennis action for content-based retrieval by collaborating audio and visual information," IEICE Trans. Inf. & Syst. (Japanese Edition), vol.J86-D-II, no.4, pp.511-524, April 2003.
- [4] M.H. Hung and C.H. Hsieh, "Event detection of broadcast baseball videos," IEEE Trans. Circuits Syst. Video Technol., vol.18, no.12, pp.1713-1726, Dec. 2008.
- [5] B. Li, H. Pan, and I. Sezan, "A general framework for sports video summarization with its application to soccer," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, vol.3, pp.169-172, April 2003.
- [6] A. Ekin, A.M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Trans. Image Process., vol.12, no.7, pp.796-807, July 2003.
- [7] R. Leonardi, P. Migliorati, and M. Prandini, "Semantic indexing of soccer audio-visual sequences: A multimodal approach based on controlled Markov chains," IEEE Trans. Circuits Syst. Video Technol., vol.14, no.5, pp.634-643, May 2004.
- [8] G. Zhu, C. Xu, Q. Huang, Y. Rui, S. Jiang, W. Gao, and H. Yao,

- “Event tactic analysis based on broadcast sports video,” *IEEE Trans. Multimed.*, vol.11, no.1, pp.49–67, Jan. 2009.
- [9] M. Beetz, S. Gedikli, J. Bandouch, B. Kirchlechner, N. von Hoyningen-Huene, and A. Perzylo, “Visually tracking football games based on TV broadcasts,” *Proc. Int. Joint Conf. Artificial intelligence*, pp.2066–2071, 2007.
- [10] V. Pallavi, J. Mukherjee, A.K. Majumdar, and S. Sural, “Graph-based multiplayer detection and tracking in broadcast soccer videos,” *IEEE Trans. Multimed.*, vol.10, no.5, pp.794–805, Aug. 2008.
- [11] S. Takahashi, W. Lim, and M. Haseyama, “Players tracking approach using level-set method based on color components of players in soccer videos,” *Proc. Work. Picture Coding and Image Processing*, pp.67–68, Dec. 2010.
- [12] S. Takahashi, W. Lim, and M. Haseyama, “Player tracking by using level-set method in soccer video,” *IEICE Trans. Inf. & Syst. (Japanese Edition)*, vol.J96–D, no.3, pp.695–703, March 2013.
- [13] Y. Ohno, J. Miura, and Y. Shirai, “Tracking players and estimation of the 3D position of a ball in soccer games,” *Proc. Int. Conf. Pattern Recognition*, vol.1, pp.145–148, 2000.
- [14] S. Iwase and H. Saito, “Parallel tracking of all soccer players by integrating detected positions in multiple view images,” *Proc. Int. Conf. Pattern Recognition*, vol.4, pp.751–754, Aug. 2004.
- [15] M. Xu, J. Orwell, and G. Jones, “Tracking football players with multiple cameras,” *Proc. Int. Conf. Image Processing*, vol.5, pp.2909–2912, Oct. 2004.
- [16] K. Iwai, S. Takahashi, T. Ogawa, and M. Haseyama, “Performance improvement of soccer player tracking using elastic model and its evaluation,” *Proc. Int. Work. Advanced Image Technology*, pp.686–691, 2013.
- [17] T. Watanabe, M. Haseyama, and H. Kitajima, “A soccer field tracking method with wire frame model from TV images,” *Proc. Int. Conf. Image Processing*, vol.3, pp.1633–1636, Oct. 2004.
- [18] FIFA, *Lows of the Game 2012/2013*, Available: <http://www.fifa.com>
- [19] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, vol.1, pp.886–893, 2005.
- [20] H. Kon and M. Haseyama, “Players clustering based on graph theory for tactics analysis purpose in soccer videos,” *IEICE Trans. Fundamentals*, vol.E90–A, no.8, pp.1528–1533, Aug. 2007.
- [21] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, “Optimization by simulated annealing,” *Science*, vol.220, no.4598, pp.671–680, May 1983.
- [22] I. Michael and B. Andrew, “CONDENSATION—conditional density propagation for visual tracking,” *Int. Journal of Computer Vision*, vol.29, no.1, pp.5–28, 1998.
- [23] M.C. Hu, M.H. Chang, J.L. Wu, and L. Chi, “Robust camera calibration and player tracking in broadcast basketball video,” *IEEE Trans. Multimed.*, vol.13, no.2, pp.266–279, April 2011.
- [24] G. Zhu, C. Xu, Q. Huang, and W. Gao, “Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter,” *Proc. IEEE Int. Conf. Multimedia and Expo*, pp.1629–1632, July 2006.
- [25] H. Li and M. Flierl, “Sift-based multi-view cooperative tracking for soccer video,” *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp.1001–1004, March 2012.
- [26] X. Tong, J. Liu, T. Wang, and Y. Zhang, “Automatic player labeling, tracking and field registration and trajectory mapping in broadcast soccer video,” *ACM Trans. Intelligent Systems and Technology*, vol.2, no.2, pp.15:1–15:32, Feb. 2011.
- [27] H. Kataoka, K. Hashimoto, and Y. Aoki, “Player position estimation by monocular camera for soccer video analysis,” *Proc. SICE Annual Conference*, pp.1985–1990, Sept. 2011.
- [28] T. Zhang, B. Ghanem, and N. Ahuja, “Robust multi-object tracking via cross-domain contextual information for sports video analysis,” *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp.985–988, March 2012.

- [29] Y. Freund and R.E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Computational Learning Theory*, vol.904, pp.23–27, 1995.

Kazuya Iwai received his B.S. degree in Electronics and Information Engineering from Hokkaido University, Sapporo, Japan in 2012. He is currently pursuing an M.S. degree at the Graduate School of Information Science and Technology, Hokkaido University. His research interests include semantic analysis in videos.

Sho Takahashi received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Sapporo, Japan in 2008, 2010 and 2013, respectively. He is currently an assistant professor in the Graduate School of Information Science and Technology, Hokkaido University. His research interests include semantic analysis in videos and its visualization. He is a member of the IEEE and Institute of Image Information and Television Engineers (ITE).

Takahiro Ogawa received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2003, 2005 and 2007, respectively. He is currently an assistant professor in the Graduate School of Information Science and Technology, Hokkaido University. His research interests are digital image processing and its applications. He is a member of the IEEE and Institute of Image Information and Television Engineers (ITE).

Miki Haseyama received her B.S., M.S. and Ph.D. degrees in Electronics from Hokkaido University, Japan in 1986, 1988 and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University as an associate professor in 1994. She was a visiting associate professor of Washington University, USA from 2005 to 2006. She is currently a professor in the Graduate School of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She is a member of the IEEE, Institute of Image Information and Television Engineers (ITE) and Acoustical Society of Japan (ASJ).