

DIGITAL CAMERA WORK FOR SOCCER VIDEO PRODUCTION WITH EVENT RECOGNITION AND ACCURATE BALL TRACKING BY SWITCHING SEARCH METHOD

Yasuo Arika, Tetsuya Takiguchi

Organization of Advanced Science and Technology
Kobe University
{ariki,takigu}@kobe-u.ac.jp

Kazuki Yano

Graduate School of Engineering
Kobe University
yano@me.cs.scitec.kobe-u.ac.jp

ABSTRACT

In this paper, we propose a method of digital zooming by automatically recognizing the soccer game events such as penalty kick and free kick based on player and ball tracking. We also propose an efficient and stable ball tracking method by switching search methods between global search and local search. In the frames where the ball is lost as well as in the first frame, the global search with normalized cross-correlation is performed. Then the local search with particle filter continues the tracking. Once the system has recognized that the local search had failed by receiving a sequence of low probabilities of the ball tracking results, the search method is switched to the global search. We carried out three ball tracking experiments; global search, local search and the proposed switching search method. As a result, the experiments showed the effectiveness of the proposed method.

Index Terms— Digital camera work, situation detection, ball tracking, particle filter, normalized cross correlation

1. INTRODUCTION

A digital broadcasting requires a lot of video contents due to the increasing number of special channels. Sports video contents for the small audience are key issue to this problem, but they have not been produced so far due to the production cost. From this view point, we have been developing an automatic production system of commentary soccer video by digital camera work based on the event (situation) recognition on Hi-vision video images. "Commentary" means intelligible and the system can produce the comments on the game progress or game events because the events are recognized and the digital camera work is carried out based on the events.

The system is composed of image recognition techniques to track a ball and players, event recognition and finally digital camera work with panning and zooming. The event recognition is the key issues for digital camera work as well as for retrieving the events and also summarizing the whole soccer game.

However, the 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 with the ball. So far, a ball tracking was a difficult task so that complete event recognition and automatic video production failed.

In this paper, we propose a new technique to track the ball accurately on the Hi-vision video. Since the Hi-vision camera is set far from the soccer field to catch the half field, the ball looks so small for tracking. At first, on the first frame, the ball is searched for by using a global search with normalized cross-correlation. Although it is possible to search for the ball by the global search for all frames,

the ball is lost sooner or later because of its small size and occlusion by the players.

To solve this problem, we employ a local search with a particle filter which can keep tracking around the area where it is lost. When the local search fails due to the ball disappearance in a case where the ball passes on and along the white line, the real ball locates far from the particles. In this case, the global search starts instead of local search. After detecting the ball again, the local search continues tracking. This switching of the search strategies is automatically carried out depending on the situation when the ball is lost.

2. RELATED WORKS

Mainly, there are two procedures in soccer ball tracking. One is the ball detection and the other is ball tracking. In paper [1], the ball candidates are first detected from several consecutive frames of the broadcast TV video, using color, shape and size information. Then a weighted graph is constructed with each node representing a candidate and each edge linking two candidates in adjacent frames. On this graph, Viterbi algorithm is carried out to extract the optimal path as ball's locations. After ball detection, it is tracked by using Kalman filter based template matching.

In paper [2], the ball candidates are also detected on each frame of the broadcast TV video, by removing non-ball objects using color, shape and size information. After obtaining the ball candidates on all frames, Kalman filter is applied from the first frame to estimate the ball position on the succeeding frames. If the candidate is found on the next frame near the estimated point, it is used to update the Kalman filter. Otherwise (if no candidates are found), the estimated position is considered as the ball position.

In paper [3], game field is firstly extracted on the broadcast soccer video. Then several simple shape features and spatial context color are evaluated and distinct non-ball regions are removed. Further examination determines the initial position of the ball on the first frame. Then it is tracked by CONDENSATION algorithm (particle filter) based on the similarity of histogram intersection of the object regions. This method hypothesizes the ball occurrence in the game field and fails under some special conditions such as occlusions of the ball by players or line mark, too complex background, or too small size of the ball due to far distance. Although our approach also employs particle filter, it is used as the local search and switched to the global search when the ball is lost. This means that it is not necessary to hypothesize the ball occurrence in the field.

3. OVERVIEW OF THE SYSTEM

3.1. Processing flow

Fig.1 shows a processing flow of automatic video producing system. At first, the entire image sequence is captured by the fixed Hi-vision camera.

In the image processing module, players and a ball are tracked and their coordinates are extracted. The image sequence is captured by the fixed Hi-vision camera so that any camera work is not included. Therefore the background subtraction method can be applied to the material video to extract players and ball. After the background subtraction process, the ball is tracked by switching the search method. The players are extracted on every frame after labeling the background subtracted images, but not tracked or matched between consecutive frames.

In the situation recognition module, the events such as throw in, free kick and goal kick are recognized based on the coordinates of players and a ball, using the rules to recognize the event situation. In the camera work module, proper clipping size and frame coordinates are decided according to the recognized events.

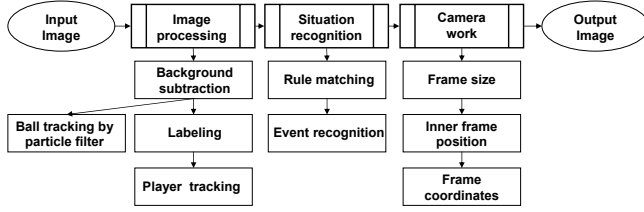


Fig. 1. Processing flow of the system.

3.2. Ball Tracking

In tracking a soccer ball on videos, there are several problems to be solved. The main problems are summarized as follows;

1. A ball is small and round so that it is sometimes misrecognized to other similar objects due to its featureless property.
2. In a case where the ball is kicked high and passing over the audience, similar objects such as audience faces, caps or bags are confused with the ball.
3. The ball is sometimes occluded by players and misrecognized to player's white shoes.
4. In a case where the ball passes on the white line or white pole, it disappears for a shot time.
5. The ball movement is sometimes irregular when it touches players.

To solve the problems 1 and 2, the background subtraction is effective because the stationary objects similar with the ball are removed. Then ball detection methods such as normalized cross-correlation between the detected moving objects and ball template can be applied on the first frame. We call this the global search because the system does not know where the ball is on the first frame so that it searches on the whole image.

This global search is time consuming so that more efficient search algorithm is required. One of the most plausible methods will be a local search such as Kalman filter or particle filter. The Kalman tracker hypothesizes unimodal Gaussian probability distribution function. On the other hand, the particle filter does not have



Fig. 2. Organization of a ball tracking system

such hypothesis so that the particle filter is thought to be better than the Kalman filter. This ball search method is called a local search because the system predicts the ball existing area and the ball is searched within the area.

Local search can solve the problems 3, 4 and 5 to some degree, but it fails due to ball disappearance when the ball passes on and along the white line or it passes along the big white character written on the wall before the audience. In this case, after a few seconds, the global search should be applied to detect the ball position again.

Fig.2 shows an organization of the proposed ball tracking system with the global search and local search. In the frames where the ball is lost as well as in the first frame, the global search with normalized cross-correlation is performed. Then the local search with particle filter continues the tracking. Once the system has recognized that the local search had failed by receiving a sequence of low probabilities of the ball tracking results, the search method is switched to the global search.

4. GLOBAL SEARCH

In the global search for the ball, the normalized cross-correlation $R(x, y)$ is computed between the ball template $T(i, j)$ and the small region within the search area I based on the following equation;

$$R(x, y) = \frac{\sum_{i,j} \{T(i, j) - \bar{T}\} \cdot \{I(x + i, y + j) - \bar{I}\}}{\sqrt{\sum_{i,j} \{T(i, j) - \bar{T}\}^2 \cdot \sum_{i,j} \{I(x + i, y + j) - \bar{I}\}^2}} \quad (1)$$

$$(\hat{x}, \hat{y}) = \arg \max_{x,y} R(x, y)$$

where \bar{T} and \bar{I} are the averaged intensity of ball template image and the underlying small region respectively. (\hat{x}, \hat{y}) is the point where the ball is matched best within the search area I . In this study, the search area I is set to a whole image in the video. The ball template is prepared manually in advance.

5. LOCAL SEARCH WITH PARTICLE FILTER

5.1. Particle Filter

In the local search for the ball, the particle filter[4] is employed. The ball is tracked by estimating a *posterior* probability $p(x_t|Z_t)$ of the ball state x_t at time t after observing the image feature sequence $Z_t = (z_1, \dots, z_t)$ up to current time t .

This *posterior* probability is computed by Bayesian theory as follows, using a *prior* probability $p(x_t|Z_{t-1})$ and the likelihood $p(z_t|x_t)$ of the image feature z_t at the state x_t .

$$p(x_t|Z_t) = k_t p(z_t|x_t) p(x_t|Z_{t-1}) \quad (2)$$

If the *posterior* probability $p(x_t|Z_t)$ is complicated, it can be computed by sampling randomly the states x_t as particles $s_t^{(n)}$ ($n = 1, \dots, N$) according to the *prior* probability $p(x_t|Z_{t-1})$ and estimating the likelihood $p(z_t|x_t)$ of the image feature z_t at the state x_t as $\pi_t^{(n)}$. It is expressed as follows;

$$p(x_t|Z_t) \approx \sum_{n=1}^N \pi_t^{(n)} \delta(s_t^{(n)}) \quad (3)$$

Here, δ indicates the delta function.

The *prior* probability $p(x_t|Z_{t-1})$ is computed by transferring the *posterior* probability $p(x_{t-1}|Z_{t-1})$ at previous time $t-1$, using the state transition probability $p(x_t|x_{t-1})$ as follows;

$$p(x_t|Z_{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|Z_{t-1})dx_{t-1} \quad (4)$$

In this study, we employ motion dynamics.

5.2. Motion Dynamics

The motion dynamics used to move the particles from the *posterior* probability at time $t-1$ to the *prior* probability at time t is expressed by the following equation of motion;

$$x_t = \begin{bmatrix} I & I \\ 0 & I \end{bmatrix} x_{t-1} + \omega \quad (5)$$

where ω is Gaussian noise term, and x_t is the state of the particle and expressed as follows;

$$x_t = [p_{tx}, p_{ty}, v_{tx}, v_{ty}]^T \quad (6)$$

Here, p_{tx} , p_{ty} , v_{tx} and v_{ty} are the positions of x and y , velocities in x and y direction in the image at time t respectively. The velocity is estimated using the past tracked positions.

5.3. Likelihood Estimation

The likelihood of the ball is estimated at each particle by the normalized cross-correlation expressed in Eq.(1) between the image feature at the position of the particle and the ball template. When the normalized cross-correlation is applied directly, many particles are mis-recognized to the non-ball objects such as audience faces, caps and bags. To avoid such a situation, moving objects are only extracted and tracked by the particle filter. The moving objects are extracted by subtracting the background image from the present.

6. CAMERA WORK MODULE

In the camera work module, the digital panning and zooming are controlled. The digital panning is performed on the HD (High Definition) image, taken by Hi-vision camera, by moving the coordinates of the clipping window and the digital zooming is performed by changing the size of the clipping window.

In zooming, three sizes of the clipping window are found to be used in the soccer game, after analyzing the professional camera work on TV soccer game; tight shot, middle shot and loose shot size. These sizes of the clipping window are selected according to the game situation. For example, the tight shot is selected for the play near the goal and the ball movement is small. If the tight shot is used frequently, the video becomes not intelligible so that the tight shot is used only for the important play.

The loose shot or middle shot is selected for the normal play or the situation like free kick where the ball is expected to move fast. Transition from the loose shot to middle shot or vice versa is performed according to the game situation. If the duration of the loose shot or middle shot after transition is less than 0.5 second, then the transition is not caused. According to this fact, in our experiment, three clipping sizes are prepared as shown in Table1. They are continuously or abruptly switched each other according to the game situation as shown in Fig.3.

Table 1. Three sizes of the clipping window on the HD image (pixel)

Tight shot size	Middle shot size	Loose shot size
120 × 90	240 × 180	480 × 360

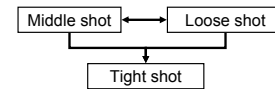


Fig. 3. Size switching of clipping window.

7. SITUATION RECOGNITION MODULE

In the situation recognition module, the game situation is recognized based on the ball and players extracted in the image processing module and this game situation is forwarded to the camera work module. The game situation to control the camera work is classified into two groups. One is to change the camera work among loose shot, middle shot and tight shot according to the ball and players relations. The other is to set the typical camera work according to the events such as goal kick, corner kick and free kick.

The proposed system can recognize five events; free kick, goal kick, throw in, corner kick and penalty kick. These events are detected using the feature that the ball keeps still for some duration under the tight shot camera work. The duration to detect the events is set to be 6 seconds.

Once the events are detected, the system zooms out the camera to watch the entire situation and recognizes the event types according to rules using the ball position and the distance between the ball and players average position as shown in Table2. Distance "far" and "middle" is decided by the predetermined threshold. For example, if the ball is on the corner spot and the distance between ball and players average position is middle, the event is recognized as the corner kick. After the event recognition, the clipping size is determined according to the rules and the center of the clipping window is set to the players average position.

Table 2. Event recognition rules

Events	Ball position	Distance between ball and players	Clipping size
Free kick	Field area	Far	LS or MS
Goal kick	Goal area	Far	LS
Throw in	Out of touch line	Middle	MS
Corner kick	Corner spot	Middle	MS
Penalty kick	Penalty spot	Middle	TS

8. EVALUATION EXPERIMENT

8.1. Ball tracking accuracy

We selected consecutive 800 frames (26.6 secs) from the soccer game of 38th National high school soccer championship Kyoto area

final in Japan. The size of the image was 1280 x 720 pixels with 24 bit color. The background image was produced in advance by averaging the randomly selected images in the video. The ball template image size was 15 x 15 pixels and the particle image size was 30 x 30 pixels that is larger than the ball template image size to search the best position within the particle image. The number of particles re-sampled was 100.

Three tracking methods, global search, local search and switched search were evaluated. The result is shown in Table3. In the table, "Tracking rate" is the ratio of the number of correctly tracked frames to the number of total frames 800. "Restart" indicates the necessity of manual restart when the tracking failed. "Processing time" was measured per an image at average on a computer with Xeon 3.06GHz and 1024 MB memory.

In Table3, the local search achieved drastically high tracking rate with low processing time owing to the particle filter. Two failures occurred when the ball passed on and along the white line and it passed along the big white character written on the wall before the audience. In these cases, the manual restart was required by specifying the true ball position.

On the other hand, in the proposed method, the search was switched to the local search after the global search obtained high cross-correlation value, and the search was switched to the global search when the 40 frames continued with the low probability of the ball tracking results. The threshold of high cross-correlation and the low probability were empirically determined. It showed that the failures caused by the particle filter were saved by the proposed switched search and the processing time was almost same as the particle filter.

Table 3. Experimental result

Method	Tracking rate (%)	Restart	Processing time (sec)
Global search	61.1	Not required	3.2
Local search	91.5	Required	0.8
Switched search (proposed)	91.8	Not required	0.9

8.2. Subjective evaluation by AHP

Our final goal of this study is to construct the commentary soccer video production system for preferences of small audience. Therefore, AHP(Analytic Hierarchy Process)[5] can be used for the evaluation of the produced video contents because of its ability to represent human's subjectivity.

AHP is a multi-criteria decision support method designed to select the best from a number of alternatives evaluated with respect to several criteria. It carries out pair wise comparison judgments which are used to decide overall priorities for ranking the alternatives.

The video contents to be evaluated by AHP are original HD content, TV content, the produced contents by our proposed method and the contents produced by our conventional method with only panning from the original HD content[6]. The TV content is the same soccer game as HD content and broadcast by the broadcasting company. All were produced from the same soccer game of 38th National high school soccer championship Kyoto area final in Japan.

The preference weights of AHP for each content are shown in Fig.4. The TV content showed high score followed by the contents of the proposed method, HD and the conventional method. The result shows that the presentation of game process is the most important and therefore the TV content obtained high score. It also indicates that the TV content with the panning camera work by the

professional cameraman is important compared with the HD content with no camera work.

The proposed method showed the AHP weight similar to the TV content and improved the intelligibility of the game process compared with the conventional method because of the camera work such as panning and zooming based on the situation recognition. The content produced by the proposed method is still inferior to the TV contents since the digital camera work is not so high compared with the professional cameraman and the video quality is also lower than the TV and HD contents.

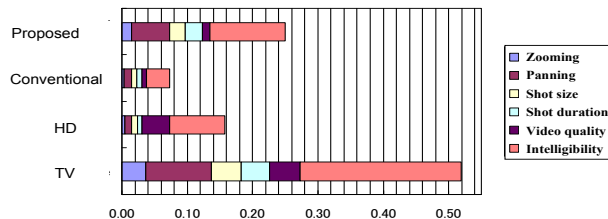


Fig. 4. Evaluation result by AHP.

9. CONCLUSIONS

In this paper, we proposed the system to track the small ball accurately on the fixed Hi-vision video by switching the global search and local search. This switching in the search strategies is automatically carried out, depending on the situation when the ball is lost. Through the experiment, we showed that the proposed method can achieve the continuous and stable ball tracking.

We also proposed a method to produce soccer video contents with digital camera work based on the situation recognition. AHP evaluation for our method showed less preference score than the TV contents. However, a basic composition of the AHP scores is almost same between the contents produced by the proposed method and the TV contents in terms of evaluation criteria such as panning and zooming. This indicates that the improvement of our proposed method will approach the techniques of the TV camera man in future. Multi cameras will be required to serve the various shots taken from different angles rather than three shot types. This problem will be also dealt with in future.

10. REFERENCES

- [1] Q. Huang D. Liang, Y. Liu and W. Gao, "A scheme for ball detection and tracking in broadcast soccer video," *PCM2005*, pp.864-875, 2005.
- [2] Q. Tian X. Yu, C. Xu and H. W. Leong, "A ball tracking framework for broadcast soccer video," *ICME03*, pp.273-276, 2003.
- [3] Qing-Shan Liu Xiao-Feng Tong, Han-Qing Lu, "An effective and fast soccer ball detection and tracking method," *ICPR04*, pp.795-798, 2004.
- [4] Michael Isard and Andrew Blake, "Condensation-conditional density propagation for visual tracking," *Int. J. Computer Vision*, vol. 29(1), pp.5-28, 1998.
- [5] T. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology*, vol. 15, pp.234-281, 1997.
- [6] Yasuo Ariki Masahito Kumano and Kiyoshi Tsukada, "A method of digital camera work focused on players and a ball," *PCM2004*, pp.466-473, 2004.