10

Digital Camera Work for Soccer Video Production with Event Detection and Accurate Ball Tracking by Switching Search Method

Yasuo Ariki and Tetsuya Takiguchi
Kobe University
Japan

1. Introduction

Due to the increasing number of special channels, digital broadcasting requires a tremendous amount of video content and work associated with new interactive services. Sports video production for small spectators is a key issue associated with this problem, because such videos have not been produced to date due to the cost of producing such videos. From this viewpoint, an efficient and automatic low-cost production system for sports video content is needed even if it sacrifices a degree of professionalism.

When we watch a sports game on TV, the camera work helps us to understand the game progress owing to the panning and zooming carried out by the cameraman. This means that the camera work is strongly associated with the game events and the most suitable camera work is selected according to the events. Through camera work based on event recognition, more interesting and intelligible video content can be produced. Camera work can be classified into real camera work and virtual camera work. In real camera work, event recognition has to be carried out in real time. At present, it is difficult to do this, so virtual camera work is now the best way to produce dynamic video content with panning and zooming. This virtual camera work is sometimes called ``digital shooting,'' which is a composite of digital camera work and digital switching techniques.

We have been developing an automatic production system for commentary soccer video using digital shooting techniques based on the event (situation) recognition (Ariki et al., 2006). ``Commentary'' means intelligible and the system can produce comments on the game's progress or events because the events are recognized and digital shooting is carried out based on the events. The system is composed of image recognition techniques to track the soccer ball, event recognition, and finally digital camera work using panning and zooming. Event recognition is the key issue for digital camera work as well as for retrieving the events and summarizing the whole soccer game. However, event recognition mainly depends on the ball tracking accuracy, because events such as free kicks, goal kicks, throw ins, corner kicks and penalty kicks are strongly related to the ball. So far, ball tracking has been a difficult task, so complete event recognition and automatic video production have not been possible.

In this chapter, we describe a new technique to track the ball accurately on the Hi-vision video. Since the Hi-vision camera was located far from the soccer field in order to film half the field, the ball looks too small for tracking. At first, in the first frame, the ball is searched for using a global search with normalized cross-correlation. Although it is possible to search for the ball using global searching for all frames, the ball is lost sooner or later because of its small size and because of occlusion by the players. To solve this problem, we employed a local search with a particle filter, which can continue tracking around the area where the ball is ``lost.'' When the local search fails due to the disappearance of the ball (when the ball passes on and along the white line), the real ball is located far from the particles. In this case, a global search replaces the local search. After detecting the ball again, the local search continues tracking the ball. This switching of the search strategies is automatically carried out depending on the situation in which the ball is lost.

The organization of this chapter is as follows. In Section 2, the related works are described and in Section 3, the overview of the automatic production system is presented. A new ball tracking method is presented in Section 4. The camera work and the situation recognition are described in Section 5 and 6. In Section 7, the experiments of ball tracking and subjective evaluation of soccer video production with AHP are described.

2. Related Works

There are many aspects to a TV program sports video content production system, including generating highlights (Yow et al, 1995) (Assfalg et al, 2003), creating a summary (Ekin et al, 2003) (Nguyen et al, 2003), reconstruction (Bebie & Bieri, 1998), creating a mosaic (Kim & Hong, 2000) and other elemental technologies (Utsumi, et al, 2002). However, these produced videos are greatly subjected to limitations related to the video production staff (cameraman, editor, switcher, etc.). Also, the content inherits any mistakes in camera work and switching. Therefore, such video content has absolutely no degree of freedom (for example, adjusting the content to fit the preferences of the viewer).

Digital shooting has degrees of freedom higher than secondary usage of sports TV program content, because it is able to generate varied camera work and switching from the material video content, obtained by filming the whole soccer field using a HD (High-Definition) camera. One related topic of research in the field of automatic video shooting systems that we call a ``one-source multi-production system'' is Virtual Soccer Stadium (Koyama et al, 2003), which makes it possible to show viewers a free view of the soccer field in 3D. Using this system, we can watch the game from any viewpoint, but the camera work or shooting techniques used in TV broadcasts can help the viewer to understand the game's progress more clearly.

The camera work such as panning and zooming are frequently utilized in classroom lectures, and content production systems or content summarization systems are proposed, employing the detection of lecturer-movement and voice-detection algorithms (Yokoi & Fujiyoshi, 2005) (Yokoi & Fujiyoshi, 2006). However, the event detection needed to switch between various camera works in a soccer sports video production system is much more difficult.

Event recognition depends mainly on the ball-tracking accuracy, because events such as free kicks, goal kicks, throw ins, corner kicks, and penalty kicks are strongly related to the ball. There are two procedures in soccer ball tracking. One is ball detection and the other is ball tracking. In paper (Huang et al, 2005), the ball candidates were first detected from several consecutive frames of broadcast TV video using color, shape and size information. Then a weighted graph was constructed with each node representing a candidate and each edge linking two candidates in adjacent frames. Then a Viterbi algorithm was carried out to extract the optimal path as the ball's locations. After ball was detected, it was tracked using Kalman filter-based template matching.

In paper (Tian et al, 2003), the ball candidates were also detected in each frame of the broadcast TV video by removing non-ball objects using color, shape and size information. After obtaining the ball candidates in all the frames, the Kalman filter was applied from the first frame to estimate the ball position on the succeeding frames. If the candidate was found in the next frame near the estimated point, it was used to update the Kalman filter. Otherwise (if no candidates were found), the estimated position was considered to be the ball position.

In paper (Liu et al, 2004), the soccer field was first extracted from the broadcast soccer video. Then several simple shape features and spatial context color were evaluated, and distinct non-ball regions were removed. Further examination determined the initial position of the ball in the first frame. The ball was then tracked using a CONDENSATION algorithm (particle filter) based on the similarity of histogram intersection of the object regions. This method hypothesizes the ball's occurrence on the game field and fails under some special conditions, such as occlusion of the ball by players or white field lines, too complex of a background, or the small size of the ball due to its distance from the camera. Although our approach also employs a particle filter, it is used for the local search only and switches to a global search when the ball is lost. This means that it is not necessary to hypothesize the ball's occurrence on the field.

3. Overview of the System

3.1 Digital Shooting

Digital shooting can be assumed to be an emulation of a virtual multiple-camera system that works by clipping a frame from HD material video content and mapping it roughly to frame at low resolution, such as SD (Standard Definition).

The digital shooting technique is composed of digital camera work and digital switching technique. The digital camera work is defined as virtual panning and virtual zooming. The virtual panning is a video production technique involving clipping a size-fixed frame by controlling frame location on HD video. The virtual zooming is a video production technique involving clipping a frame by controlling frame size. Also, digital switching is defined as the change of some virtual camera by controlling rapid change of frame location or size on HD video.

Although the camera work or switching by human beings during live sports events cannot retake the scene, digital shooting is able to repeatedly produce various camera work and switching from the recorded video material, which involved the filming of the entire soccer field with a HD camera.

In our experiments, we used a Victor GR-HD1 Hi-vision camera. Figure 1 shows half the field taken by Hi-vision camera and a SD image clipped from the fixed HD image. Digital camera work is produced by changing the coordinates of the clipping window on every HD frame. For example, Figure 2 and Figure 3 show digital panning down from right to left down and digital zooming in, respectively.

Fig. 1. Clipping from the HD image to a SD image using digital camera work

Fig. 2. Panning by digital camera work

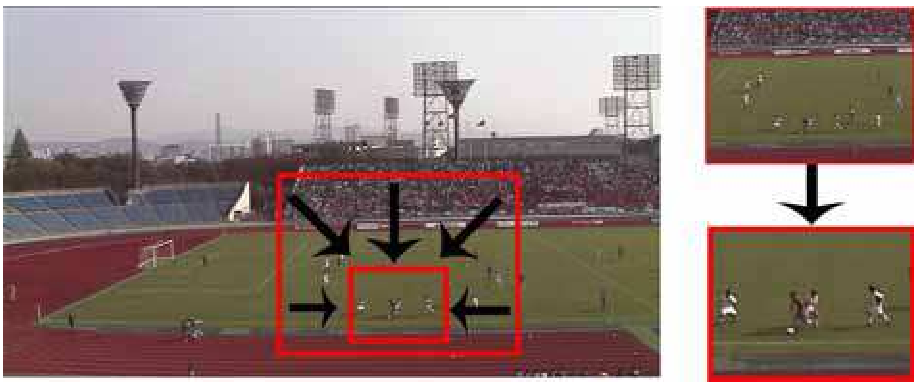

Fig. 3. Zooming by digital camera work

3.2 Processing Flow

Figure 4 shows the processing flow of the digital camera work system. At first, the entire image sequence is captured by the fixed Hi-vision camera.

In the image processing module, the players and the ball are tracked, and their coordinates are extracted. The image sequence is captured by the fixed Hi-vision camera so that no camera work is included in the original video content. Therefore, the background subtraction method can be applied to the material video to extract the players and ball. Background subtraction is a simple but effective method to detect moving objects in video images.

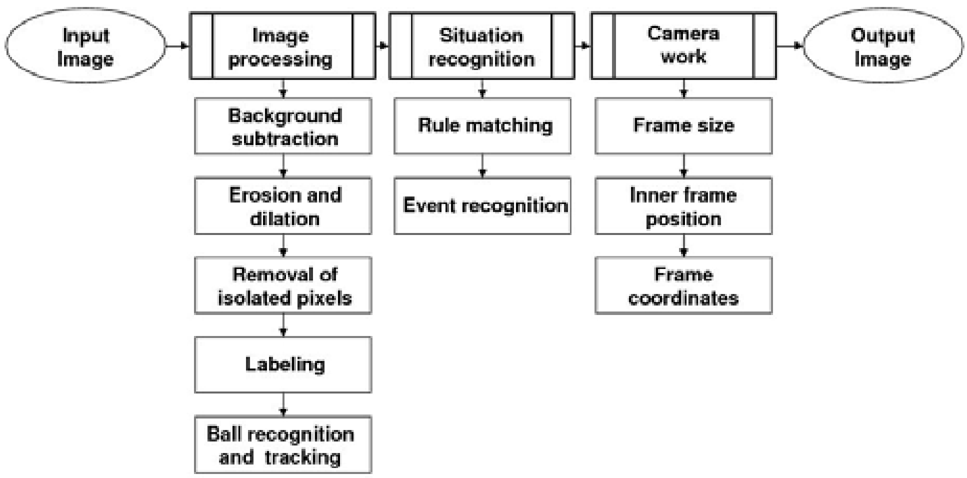

Fig. 4. Processing flow of the system

These background subtracted images do not depend on the image color. Hence, this method can be applied not only to the grass field but also to a dirt field. Furthermore, it is robust enough to handle the slow changes caused by sunshine or illumination if the background image to be subtracted is updated. The background image is updated by averaging the *M* frames at every *N* frames.

In the background subtraction process, each binary image is preprocessed by a morphological operator (erosion and dilation) to extract a player region, and noise reduction processing is applied. After region labeling, the ball is recognized and tracked. The players are extracted on every frame but not tracked or matched between consecutive frames.

In the situation recognition module, the events (such as throw ins, free kicks and goal kicks) are recognized based on the coordinates of the players and the 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.

The event recognition depends mainly on the ball tracking accuracy, because events (such as free kicks, goal kicks, throw ins, corner kicks and penalty kicks) are strongly related to the ball. So far, ball tracking was a difficult task so that complete event recognition and automatic video production were not successful.

In the following section, we propose an efficient and stable ball-tracking method by switching search methods back and forth 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 has failed by receiving a sequence of low probabilities among the ball-tracking results, the search method is switched to the global search.

4. 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 sometimes it is confused with other similar objects due to its featureless properties.
2. In the case where the ball is kicked high and passes over the spectators, similar objects such as people's faces, caps or bags are confused with the ball.
3. The ball is sometimes occluded by players and confused with the players' white shoes.
4. In the case where the ball passes on the white line or white pole, it disappears for a short time.
5. The ball movement is sometimes irregular when it touches players.

To solve problems 1 and 2, 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 a 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 the whole image.

This global search is time-consuming, so a 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 a unimodal Gaussian probability distribution function. On the other hand, the particle filter does not have such a hypothesis, so 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's location area, and the ball is searched for within this area.

Fig. 5. Organization of a ball tracking system

A local search can solve 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 a big white character written on the wall in front of the spectators. In this case, after a few seconds, the global search should be applied to detect the ball position again.

Figure 5 shows the organization of the proposed ball-tracking system using 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 a particle filter continues the tracking. Once the system has recognized that the local search has failed by receiving a sequence of low probabilities of the ball tracking results, the search method is switched to the global search.

4.1 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:

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

$$(\hat{x},\hat{y})=\arg\max_{x,y} R(x,y) \tag{2}$$

where $\bar{T}$ and $\bar{I}$ are the averaged intensity of the 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.

4.2 Local Search with Particle Filter

4.2.1 Particle Filter

In the local search for the ball, a particle filter (Isard & Blake, 1998) is employed. The ball is tracked by estimating *a posterior* probability $p(x_t \mid Z_t)$ of the ball state x_t at time *t* after observing the image feature sequence $Z_t=(z_1,...,z_t)$ up to current time *t*.

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

$$p(x_t \mid Z_t)=k_t p(z_t \mid x_t)p(x_t \mid Z_{t-1}) \tag{3}$$

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

$$p(x_t \mid Z_t) \approx \sum_{n=1}^{N} \pi_t^{(n)} \delta(s_t^{(n)}) \tag{4}$$

Here, δ indicates the delta function.

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

$$p(x_t \mid Z_{t-1}) = \int_{x_{t-1}} p(x_t \mid x_{t-1}) p(x_{t-1} \mid Z_{t-1}) dx_{t-1} \tag{5}$$

In this study, we employ motion dynamics.

4.2.2 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:

$$x_t = \begin{bmatrix} I & I \\ 0 & I \end{bmatrix} x_{t-1} + \omega \tag{6}$$

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

$$x_t = [p_{tx}, p_{ty}, v_{tx}, v_{ty}]^T \tag{7}$$

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

4.2.3 Likelihood Estimation

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

5. Camera Work Module

In the camera work module, the digital panning and zooming are controlled. Digital panning is performed on the HD image by moving the coordinates of the clipping window, and digital zooming is performed by changing the size of the clipping window.

5.1 Zooming and Clipping Size

After analyzing the professional camera work on a TV soccer game, it was found that during soccer games, three sizes of the clipping window are used in zooming in: 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 when the ball movement is slight. If the tight shot is used frequently, the video becomes not intelligible so the tight shot is used only for important play with duration more than two seconds.

The loose shot or middle shot camera work is selected for normal play or situations, such as free kicks, 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. The transition is continuously done within 0.5 seconds. If the duration of the loose shot or middle shot after the transition is less than 0.5 seconds, then a transition does not take place.

Based on these parameters, in our experiment, three clipping sizes are prepared as shown in Table 1. Figure 6 shows examples of these shot sizes on the HD image. They are continuously or abruptly switched back and forth between each other according to the game situation, as shown in Figure 7.

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

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

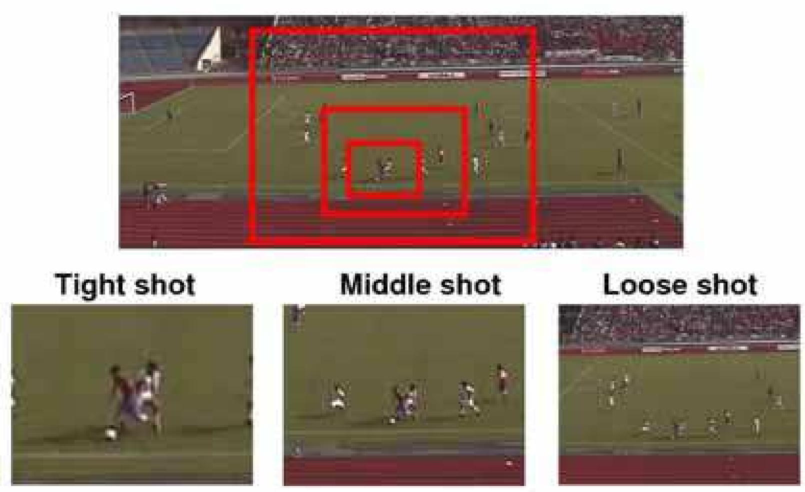

Fig. 6. Example of clipping window sizes

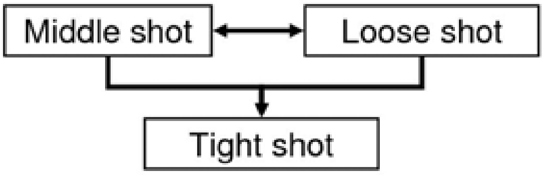

Fig. 7. Size switching of clipping window

5.2 Panning and Clipping Coordinates

The ball location is important because the soccer game progresses following the ball's location. However, ball trajectory cannot adequately be used directly to trigger a commencement of panning because smooth panning cannot be obtained due to the erratic movement of the ball.

In other words, the panning operation has to satisfy two contrary conditions, namely that the clipping window (frame) must follow the ball quickly if the ball moves fast, but if the ball moves erratically, the frame must not follow the ball.

To meet these two conditions, we set an inner frame within the clipping window as shown by the black frame in Figure 8. Even if the ball is moving erratically, the clipping window remains still if the ball exists inside the inner frame. If the ball moves out of the inner frame, the centroid of the clipping window moves toward the ball location as shown in Figure 9.

Fig. 8. Clipping window (white) and inner frame (black)

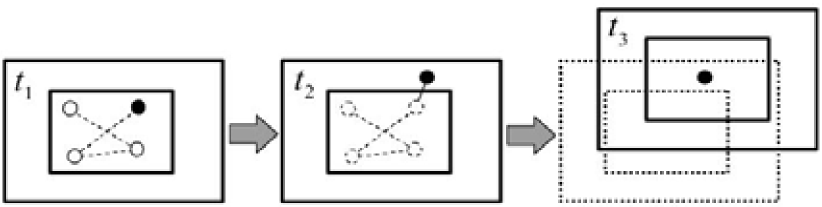

Fig. 9. Control method of clipping window

5.3 Inner Frame Position

The inner frame usually locates itself at the center of the clipping window. However, if the play is something like a free kick, long pass or dribbling etc., then the forward direction should be included in the clipping window to make the play formation or players remain visible.

Therefore, it is necessary to set the inner frame at the position opposite the center of the clipping window toward the ball movement direction. Figure 10 shows the inner frame in such a situation. The inner frame is shown as the small rectangle and the clipping window is shown as the large rectangle. Usually the left clipping window and the inner frame are employed. However, in the situation of a free kick by the goalkeeper, the right clipping window is selected to set the inner frame at the left position (opposite to the ball direction).

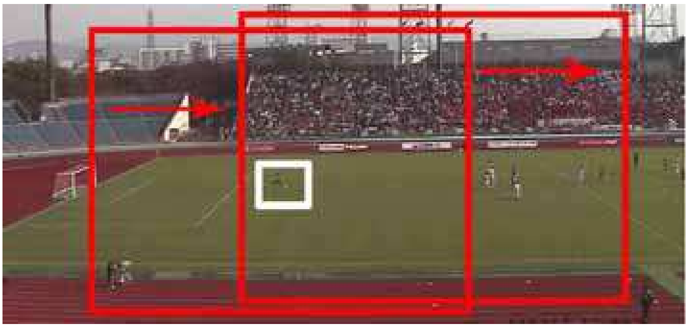

Fig. 10. Inner frame position control

6. 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 are forwarded to the camera work module. The game situations for controlling the camera work are classified into two groups. One changes the camera work among loose shots, middle shots and tight shots according to ball-player relationships. The other sets the typical camera work according to the events such as goal kicks, corner kicks and free kicks.

In order to clarify the game situation for changing the camera work, we analyzed the professional camera work of a soccer game broadcast on TV. From this analysis, rules were found for causing the camera work changes based on the game situation and the events. In this section, the rules for starting zooming and the rules for detecting events are described.

6.1 Zooming Rule

Rules for changing the camera work based on the game situation are described. The camera work change has already been described in Figure 7.

6.1.1 Rules for Switching from Loose Shot to Middle Shot

Changing the camera work from the loose shot to the middle shot occurs when a more detailed game situation is required in order to understand play. Through analyzing TV soccer games, two types of such situations are found. One is the case where the ball approaches the goal and the other is the case where the players crowd around the ball.

One of the most important scenes in a soccer game is a goal scene so that the camera must be zooming in at this time. When the ball passes over the penalty line toward the goal, a goal

scene may be expected, so the system changes the camera work to the middle shot from the loose shot. The situation is recognized by computing the horizontal coordinate of the ball.
When a lot of players are around the ball, it is effective to zoom in to see clearly. This situation is recognized by counting the number of players around the ball on middle shot resolution. If the number is over six or seven players, then the situation is interpreted as crowded.
The accuracy of the player detection is not so high due to the overlap of the players, however, so the crowded situation is determined by counting the number of players including the player overlap.

6.1.2 Rules for Switching from Middle Shot to Loose Shot

Changing the camera work from middle shot to loose shot occurs when the camera field does not include the ball or the minimum number of players. Through analyzing broadcast soccer games on TV, two such situations are found. One is the case where the ball moves too fast and the other is where the players scatter very fast.
A fast-moving ball situation is detected by computing the displacement of the ball coordinates at every frame. If it is over the threshold (six pixels), the fast movement of the ball is detected. This rule may be applied even when the ball moves forward and backward around the same place. In this case the camera work may change frequently. To avoid this situation, the coordinate of the ball is checked and if it is increasing or decreasing in a constant direction for more than two seconds, the rule is applied.
When players scatter, the camera work changes from the middle shot to loose shot to catch the ball and the minimum number of players. This situation is recognized by counting the number of players around the ball.
There are some conflicts between rules for changing camera work from the middle shot to loose shot and vice versa. One conflict is when a player kicks the ball toward the goal; i.e. when the ball moves very fast near the goal. In this case, two rules may be applicable. One is the rule to change the camera work from loose shot to middle shot because the ball is located near the goal as described in Section 6.1.1. The other rule is to change the camera work from middle shot to loose shot because of the very fast ball movement. In this case, the rule to change the camera work from the middle shot to the loose shot is inhibited.
The other conflict is when the ball moves very fast over a crowd of players. In this case, two rules may be applicable. One is the rule to change the camera work from middle shot to loose shot because the ball moves very fast. The other is the rule to change the camera work form loose shot to middle shot because the crowded players are detected. In this case, the rule to change the camera work from the loose shot to middle shot is inhibited.

6.1.3 Rules for Tight Shots

In the professional camera work on TV soccer games, tight shots, such as famous players or players scrambling for the ball and dribbling, are sometimes inserted in the videos. These tight shots make the game interesting. The proposed system does not recognize players' faces at present, so the scrambling situation is only zoomed in on as a tight shot in our system.

This situation is recognized when the ball movement is below some threshold and there is at least one player around the ball. The threshold is set to be 1.5 pixels through the preliminary experiment.

6.2 Event Detection Rule

The proposed system can recognize five events; free kicks, goal kicks, throw ins, corner kicks and penalty kicks. These events are detected using a feature that notices when the ball remains still for some duration under the tight shot camera work. The detection duration for such events is set at 6 seconds.
Once the events are detected, the system zooms out to view the entire situation and recognizes the event types according to rules using the ball position and the distance between the ball and the players' average position as shown in Table 2. For example, if the ball is on the corner arc and the distance between the ball and the players' average position is medium distance, the event is recognized as a 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.

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 bounds	Middle	MS
Corner kick	Corner spot	Middle	MS
Penalty kick	Penalty spot	Middle	TS

Table 2. Event recognition rules

7. Evaluation Experiment

7.1 Ball Tracking Accuracy

We selected consecutive 800 consecutive frames (26.6 sec) from 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 1,280 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, which is larger than the ball template image size, in order 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 switching search) were evaluated. The results are shown in Table 3. 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. The average ``processing time'' per image was calculated on a computer with a Xeon 3.06GHz processor and 1,024 MB of memory.
In Table 3, the local search achieved a very 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 in front of the spectators. In these cases, a 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 a high cross-correlation value, and the search was switched to the global search when 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 incorrect tracking caused by the particle filter was restored by the proposed switching search and the processing time was almost same as that of the particle filter.

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

Table 3. Experimental results

7.2 Subjective Evaluation by AHP

The final goal of this study is to construct a soccer video production system that meets viewer preferences and contains announcer commentary. Therefore, AHP (Analytic Hierarchy Process) (Saaty, 1997) can be used for the evaluation of the produced video contents because of its ability to represent human subjectivity.
AHP is a multi-criteria decision support method designed to select the best from a number of evaluated alternatives with respect to several criteria. It carries out pair-wise comparison judgments, which are used to decide the overall priorities for ranking the alternatives.
Three items were selected for the evaluation criteria of the video contents by AHP. They are naturalness, video quality and intelligibility. For naturalness, four camera work criteria are selected: zooming, panning, shot size, and shot duration. The video quality is adopted to judge whether or not the produced video content is inferior to TV or HD content. Also, the intelligibility of the game process is also adopted. These criteria are shown in Table 4.

Criteria	Evaluation
1. Zooming	Good <-> Poor
2. Panning	Good <-> Poor
3. Shot size	Good <-> Poor
4. Shot duration	Proper <-> Improper
5. Video quality	Fine <-> Coarse
6. Intelligibility	High <-> Low

Table 4. Evaluation criteria used in AHP

The video content to be evaluated by AHP are HD content, TV content, the produced content by our proposed method and the content produced by our conventional method with only panning from the original HD content (Ariki et al, 2004). All were produced from the same soccer game (38th National High School Soccer Championship).
The HD content was taken for a wide-angle half of the field by HD camera because of the limitation of camera resolution. The TV content was recorded by a video recorder when the

game was broadcast on TV. The reason why the HD content was compared in this chapter was to investigate which was fundamentally more comprehensible, TV content with camera work or wide-angle HD video content.

Figure 11 shows the AHP tree map used in this experiment. The middle row shows the six evaluation criteria. The bottom row shows the four materials to be compared.

The preference weights of AHP for each type of content are shown in Figure 12. 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 a high score. It also indicates that the TV content with panning camera work by professional cameramen is important compared with the HD content with no camera work.

The proposed method showed an 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 (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.

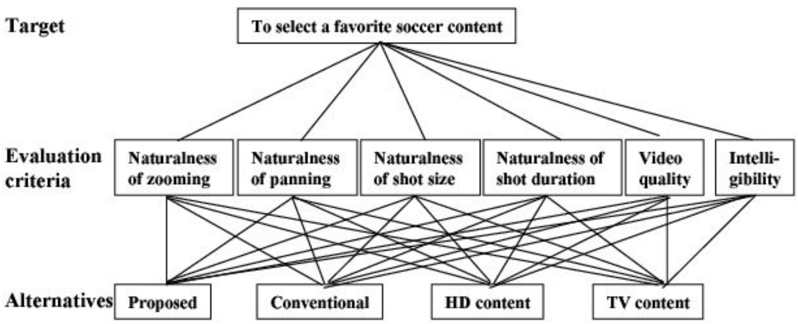

Fig. 11. AHP treemap

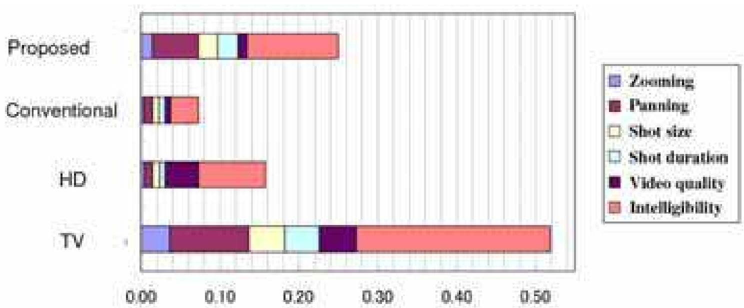

Fig. 12. Evaluation result by AHP

8. Conclusions

In this chapter, we proposed a system to track a small ball accurately on the fixed Hi-vision video by switching between global search and local search. This switching in search strategies is carried out automatically, depending on the situation when the ball is lost. Through the experiment, we showed that the proposed method can achieve continuous and stable ball tracking.
We also proposed a method to produce soccer video content with digital camera work based on situation recognition. AHP evaluation for our method showed a lower preference score than the TV content. However, the basic composition of the AHP scores is almost the same between the content produced by the proposed method and the TV content in terms of evaluation criteria, such as panning and zooming. This indicates that the improvement of our proposed method will lead to a degree of quality close to the technique of the TV cameramen in the future. Multiple cameras will be required to film the various shots that need to be taken from different angles rather than just the three shot types we tested. This problem will be also dealt with in the future.

9. References

Ariki, Y.; Kubota, S. & Kumano, M. (2006). Automatic Production System of Soccer Sports Video by Digital Camera Work Based on Situation Recognition, *Proceedings of Eight IEEE International Symposium on Multimedia*, pp. 851-858, 2006

Yow, D.; Yeo, B-L.; Yeung, M. & Liu, B. (1995). Analysis and Presentation of Soccer Highlights from Digital Video, *Proceedings of 2nd Asian Conference on Computer Vision*, pp. 499-503, 1995

Assfalg, J.; Bertini, M.; Colombo, C.; Bimbo, A. D. & Nunziati, W. (2003). Automatic Interpretation of Soccer Video for Highlights Extraction and Annotation, *Proceedings of ACM symposium on Applied computing*, pp. 769-773, 2003

Ekin, A.; Tekalp, A. M. & Mehrotra, R. (2003). Automatic soccer video analysis and summarization, *IEEE Trans. on Image Processing*, 12-7, pp. 796-807, 2003

Nguyen, N. T.; Thang, T. C.; Bae, T. M. & Ro, Y. M. (2003). Soccer Video Summarization System Based on Hidden Markov Model with Multiple MPEG-7 Descriptors, *Proceedings of the International Conference on Imaging Science, Systems and Technology*, pp. 673-678, 2003

Bebie, T. & Bieri, H. (1998). SoccerMan - Reconstructing Soccer Games from Video Sequences, *Proceedings of 2007 IEEE International Conference on Image Processing*, pp. 898-902, 1998

Kim, H. & Hong, K-S. (2000). Soccer Video Mosaicing Using Self-Calibration and Line Tracking, *Proceedings of International Conference on Pattern Recognition*, pp. 1592-1595, 2000

Utsumi, O.; Miura, K.; Ide, I.; Sakai, S. & Tanaka, H. (2002). An object detection method for describing soccer games from video, *Proceedings of 2002 IEEE International Conference on Multimedia and Expo*, pp. 45-48, 2002

Koyama, T.; Kitahara, I. & Ohta, Y. (2003). Live Mixed-Reality 3D Video in Soccer Stadium, *Proceedings of International Symposium on Mixed and Augmented Reality*, pp. 178-187, 2003

Yokoi, T. & Fujiyoshi, H. (2005). Virtual Camerawork for Generating Lecture Video from High Resolution Images, *Proceedings of 2005 IEEE International Conference on Multimedia and Expo*, pp. 45-48, 2002

Yokoi, T. & Fujiyoshi, H. (2006). Generating a Time Shrunk Lecture Video by Event Detection, *Proceedings of 2006 IEEE International Conference on Multimedia and Expo*, pp. 641-644, 2006

Huang, Q.; Liang, D.; Liu, Y. & Gao, W. (2005). A scheme for ball detection and tracking in broadcast soccer video, *Proceedings of PCM2005*, pp. 864-875, 2005

Tian, Q.; Yu, X.; Xu, C. & Leong, H. W. (2003). A ball tracking framework for broadcast soccer video, *Proceedings of 2003 IEEE International Conference on Multimedia and Expo*, pp. 273-276, 2003

Liu, Q-S.; Tong, X. F. & Lu, H. Q. (2004). An effective and fast soccer ball detection and tracking method, *Proceedings of ICPR*, pp. 795-798, 2004

Isard, M. & Blake, A. (1998). Condensation-conditional density propagation for visual tracking, *International Journal Computer Vision*, 29 (1), pp. 5-28, 1998

Saaty, T. (1997). A scaling method for priorities in hierarchical structures, *Journal of Mathematical Psychology*, 15, pp. 234-281, 1997

Ariki, Y.; Kumano, M. & Tsukada, K. (2004). A method of digital camera work focused on players and a ball, *Proceedings of 2004 Pacific-Rim Conference on Multimedia*, pp. 466-473, 2004

Multimedia
Edited by Kazuki Nishi

ISBN 978-953-7619-87-9
Hard cover, 452 pages
Publisher InTech
Published online 01, February, 2010
Published in print edition February, 2010

Multimedia technology will play a dominant role during the 21st century and beyond, continuously changing the world. It has been embedded in every electronic system: PC, TV, audio, mobile phone, internet application, medical electronics, traffic control, building management, financial trading, plant monitoring and other various man-machine interfaces. It improves the user satisfaction and the operational safety. It can be said that no electronic systems will be possible without multimedia technology. The aim of the book is to present the state-of-the-art research, development, and implementations of multimedia systems, technologies, and applications. All chapters represent contributions from the top researchers in this field and will serve as a valuable tool for professionals in this interdisciplinary field.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821