Automatic Production System of Soccer Sports Video by Digital Camera Work Based on Situation Recognition

Yasuo Ariki and Shintaro Kubota
Dept. of Computer and System Engineering
Kobe University
1-1 Rokkodai, Kobe, 657-8501, Japan
ariki@kobe-u.ac.jp

Masahito Kumano
Dept. of Electronics and Informatica
Ryukoku University
1-5 Yokotani, Seta, Oe-cho, Otsu, 520-2194, Japan
kumano@rins.ryukoku.ac.jp

Abstract

We are studying about automatic production of soccer sports videos for easy understanding by using digital camera work on camera fixed videos. The digital camera work is a movie technique which uses virtual panning and zooming by clipping frames from hi-resolution images and controlling the frame size and position. We have studied so far digital panning. In this paper, we propose a method of digital zooming by automatically recognizing the game situation or events such as penalty kick and free kick based on player and ball tracking. These recognition results are used as key indices to retrieve the event scenes from soccer videos. We compared the proposed technique with a conventional technique by AHP method that can reflect an individual subjectivity.

1. Introduction

A digital broadcasting requires a lot of video contents and works to new interactive services 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, an efficient and automatic content production system for sports video contents is strongly required at low cost even at the expense 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 of the camera work. This means that the camera work is strongly associated with the game events and the suitable camera work is selected according to the events. Through the camera work together with the event recognition, more interesting and intelligible video contents can be produced.

The camera work may be classified into real camera work and virtual camera work. In the real camera work, event recognition has to be carried out in real time. At present, it is difficult so that the virtual camera work will be the best way to produce the dynamic video contents with panning and zooming. The virtual camera work is sometimes called digital shooting, composed of digital camera work and digital switching technique.

In this paper, we propose an automatic production system of commentary soccer video by digital shooting techniques based on the event (situation) recognition. "Commentary" means intelligible and the system can produce the comments on the game progress or game events in future because the events are recognized and the digital shooting is carried out based on the events.

One of the advantages of the digital shooting is that the camera work such as panning and zooming is adjusted to the user preference. This means that the user can watch his own soccer video produced by his own virtual editor, cameraman and switcher based on the user private preference. In this paper, we also propose a method to produce the preference based video contents.

The contribution of this paper is that the soccer events are recognized using the image recognition techniques and they can be used as the key indices to retrieve the events and also to summarize the whole soccer game. Furthermore, in this paper, we use the key indices to perform the digital shooting and to produce the soccer video contents with panning and zooming.

The organization of this paper is as follows. In Section 2, the related works are described and in Section 3, the overview of the automatic production system is presented. The camera work, the situation recognition and user preference control are described in Section 4, 5 and 6. In Section 7, the experiments of soccer video production and their subjective evaluation with AHP are described.

2. Related works

There are many approaches to the contents production system such as generating highlight [1, 2], summary [3, 4], reconstruction [5], mosaic [6] and these elemental technologies [7] to sports TV program contents. However, these produced contents are greatly subjected to restriction of contents production staff such as a cameraman, editor and switcher, etc. Also, these contents inherit mistakes of camera work and switching. Therefore, these contents have little degrees of freedom.

The digital shooting has the degrees of freedom higher than secondary usage of sports TV program contents, because it is able to generate variant camera work and switching from material video contents taking a whole soccer court by HD camera. As a related work in automatic shooting system that we call "one source multi-production system", Virtual Soccer Stadium [8] shows a free view in 3D space of soccer court to users. Using this system, we can watch the game from any viewpoint. However, the camera works or shooting techniques used in TV contents can help the user to understand the game progress more clearly.

The camera works such as panning and zooming are frequently utilized in classroom lectures and content production system or content summarization system is proposed employing the detection of lecturer movement and voice detection algorithm [9, 10]. However, the event detection to switch the camera works is further difficult in the soccer sports video production system.

3. Overview of the system

3.1. Digital shooting

The digital shooting can be assumed as an emulation of a virtual multiple camera system by clipping the frame from HD material video contents and by mapping roughly to frame with the resolution for example 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 of clipping a size-fixed frame by controlling frame location on a HD material video. The virtual zooming is a video production technique of clipping a frame by controlling frame size. Also, the digital switching is defined as the change of some virtual camera by controlling rapid change of frame location or size on a HD material video.

Although the camera work or switching by human in live sports cannot perform retaking due to environmental cause, the digital shooting is able to repeatedly produce various camera work and switching from material video contents taking a whole soccer court by HD camera. Therefore the digital shooting production system can perform as potential ability the various virtual taking and meet request of small audience.

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

Figure 1. Clipping from HD image to SD image by digital camera work.

Figure 2. Panning by digital camera work.

Figure 3. Zooming by digital camera work.

3.2. Processing flow

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

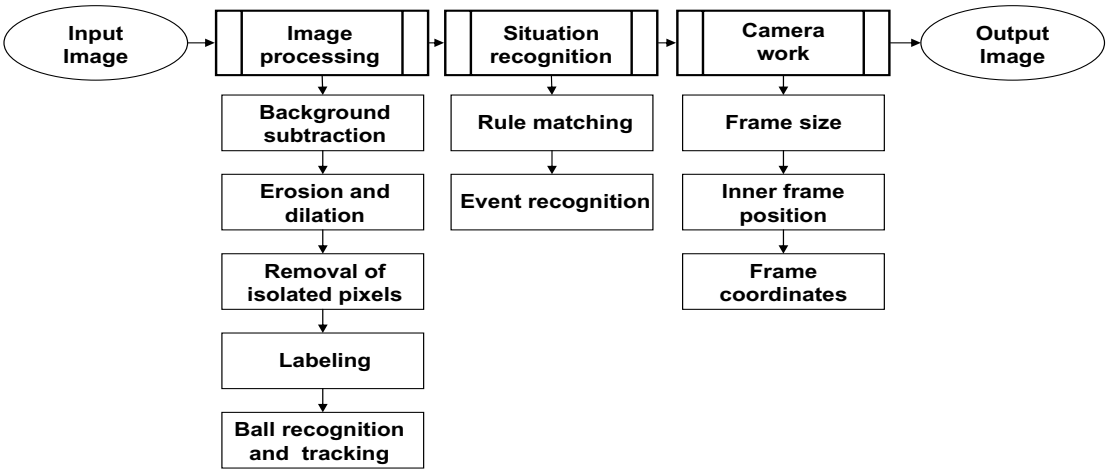

Figure 4. Processing flow of the system.

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. The background subtraction is a simple but effective method to detect moving objects in video images.

These background subtracted images do not depend on the image color. Hence this method can be applied not only to the grass court but also to the earthen court. Furthermore it is robust to the slow change 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 region of player, and a 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 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.

The event recognition and the camera work are strongly connected so that the camera work is described at first in the following section to clarify what types of camera works are required for what types of game situation (events). Then, the method to recognize the corresponding game situation is described together with the rules to detect the events.

4. Camera work module

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

4.1. Zooming and clipping size

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 with duration more than two seconds.

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 (described in Section 5.1). The transition is continuously done within 0.5 second. 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. Fig.5 shows the examples of these shot sizes on the HD image. They are con-

tinuously or abruptly switched each other according to the game situation as shown in Fig.6.

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

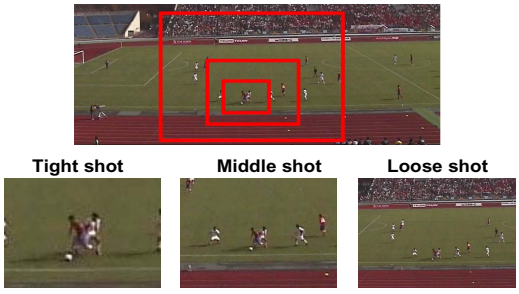

Figure 5. Example of clipping window sizes.

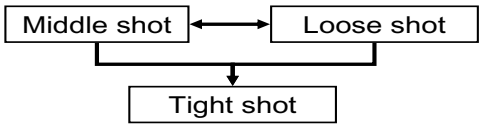

Figure 6. Size switching of clipping window.

4.2. Panning and clipping coordinates

The ball location is important because the soccer game progresses following a ball location. However, the ball trajectory is not adequate to be used directly for producing the panning because a smooth panning cannot be obtained due to wiggling movement of the ball.

In other words, the panning operation has to satisfy two contrary conditions such that the clipping window (frame) should follow the ball quickly if the ball moves fast, however if the ball wiggles, the frame should not follow the ball.

To meet these two conditions, we set an inner frame within the clipping window as shown by a black frame in Fig.7. Even if the ball is wiggling, the clipping window may be still if the ball exists inside the inner frame. If the ball is out of the inner frame, the centroid of the clipping window moves toward the ball location as shown in Figure.8.

Figure 7. Clipping window (white) and inner frame (black).

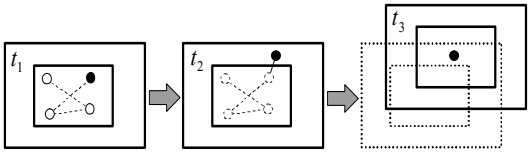

Figure 8. Control method of clipping window.

4.3. Inner frame position

The inner frame usually locates at the center of the clipping window. However, if the play is such as free kick, long pass and dribble etc, then the forward direction should be included in the clipping window to make the play formation or players be visible.

Therefore it is necessary to set inner frame at the opposite location from the center to the ball proceeding direction. Fig.9 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 free kick by the goal keeper, the right clipping window is selected to set the inner frame at the left position (opposite to the ball direction).

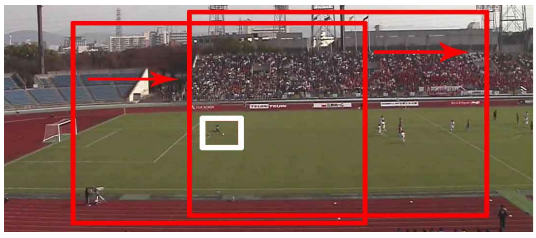

Figure 9. Inner frame position control.

5. 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.

In order to clarify the game situation for changing the camera work, we analyzed professional camera works of a soccer game broadcast on TV. From this analysis, rules are found to cause the camera work change based on the game situation and the events. In this section, rules to make zooming and rules to detect events are described.

5.1. Zooming rule

Rules to change the camera work based on the game situation are described. The camera work change is already described in Fig.6.

5.1.1 Rules from loose shot to middle shot

Changing the camera work from the loose shot to the middle shot is caused when the more detail game situation is required to understand the play. Through analyzing broadcast soccer games on TV, two types of such situations are found. One is the case where the ball approaches to 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. When the ball passes over the penalty line toward the goal, a goal scene may be expected so that 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 watch the detail. 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. Therefore the crowded situation or not is decided by counting the number of players including the player overlap.

5.1.2 Rules from middle shot to loose shot

Changing the camera work from middle shot to loose shot is caused in a case where the camera can not catch the ball or the minimum number of the players. Through analyzing broadcast soccer games on TV, two types of such situations are found. One is the case where the ball moves so fast and the other is where the players scatter away fast.

The situation of the ball fast movement is detected by computing the displacement of the ball coordinates at every frame. If it is over 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 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 to change camera work from the middle shot to loose shot and vice versa. One conflict is when a player shoots the ball to the goal. In other words, when the ball moves so 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 locates near the goal as describe in Section5.1.1. The other rule is to change the camera work from middle shot to loose shot because of the so 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 so fast over the crowded 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 so 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.

5.1.3 Rules to tight shot

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

5.2. Event detection rule

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 once 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. 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

6. User preference control

In the proposed system, the clipping size, zooming speed or duration balance of the clipping sizes can be changed or tuned to each user. However the tuning is very troublesome because there may be various combinations among them. In order to effectively select the most preferable camera work, we prepared seven types of the preference camera works. They are "telescope", "individual", "event", "zooming", "offence", "defense" and the normal mode proposed in this paper.

In the "telescope" mode, loose shot is always selected so that the game formation can be observed clearly. Since the panning is slow in this mode so that the intelligibility is better, but it is not so exciting. In the "individual" mode, the middle shot or tight shot is frequently selected. The panning becomes a little large so that the intelligibility decreases, but the detail of the players or ball movement can be caught clearly.

In the "event" mode, the events such as free kick, throw in and goal kick are zoomed in so that the position, the timing of the play and the ball movement are clearly observed. In the "zooming" mode, normally the loose shot is always selected, but when the players are scrambling for the ball, more tight zooming shot is employed than the normal camera work. Therefore both the formation and ball movement are both clearly observed.

In the "offence" and "defense" mode, the players of one team are mainly caught in the image so that the user favorite team is always observed. The "offence" and "defence" can be discriminated by computing the moving direction of the averaged position of the total players. The normal mode is the proposed one describe in Section5.1.

7. Evaluation experiment

7.1. Subjective evaluation by AHP

In this section, we present a method to evaluate the produced video contents. 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)[12] 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.

7.2. Evaluation criteria used in AHP

As the evaluation criteria of the video contents by AHP, three items were selected. They are naturalness, video quality and intelligibility. For the naturalness, four camera work criteria are selected ; zooming, panning, shot size and shot duration. The video quality is adopted to judge whether the produced video contents are inferior to TV or HD contents or not. Also, intelligibility of game process is adopted additionally. They are shown in Table3.

Table 3. Evaluation criteria used in AHP

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

7.3. Experimental setup for AHP

The video contents to be evaluated by AHP are HD content, TV content, the produced contents by our proposed method and the contents produced by our conventional method with only panning [13]. All were produced from the same soccer game of 38th National high school soccer championship Kyoto area final in Japan.

The HD content was taken for a wide-angle half of the court by HD camera in this paper because of limitation of

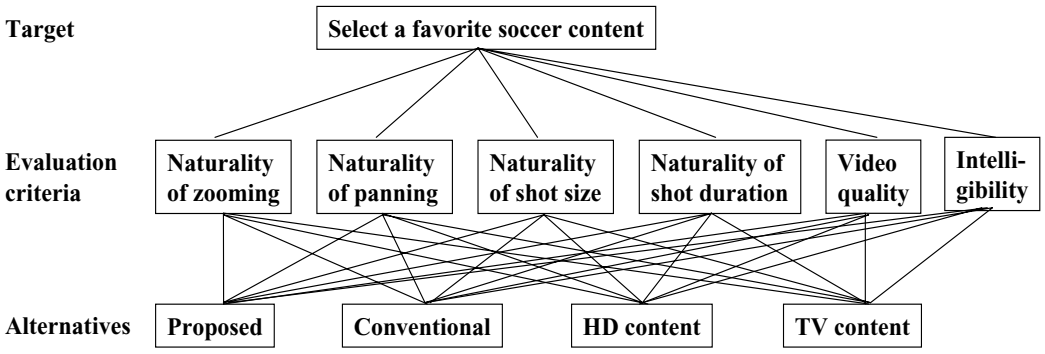

Figure 10. AHP treemap.

the 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 paper was to investigate which was fundamentally more comprehensible between TV content with camera work and HD content taking wide-angle of soccer court.

Fig.10 shows the AHP tree map used in this experiment. In the middle layer, the six evaluation criteria are placed. In the bottom, the four materials to be compared are placed.

7.4. Experimental result by AHP

As a result of the experiment, AHP showed the preference weights (0.058, 0.182, 0.087, 0.084, 0.105, 0.483) to the criteria items shown in Table3 in this order. This result indicates that the intelligibility of the game process is the most important criterion, then the panning and the resolution(video quality) follow.

The preference weights of AHP for each content are shown in Fig.11. The HD 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 shooting techniques are not so high compared with the professional cameraman and the video quality is also lower than the TV and HD contents.

As for event recognition, it was correctly recognized at about 90% except for the throw in. The recognition rate of the throw in was at about 50% because the event was recognized based on the duration of the ball stopping so that it fails when some player starts the game faster than normal pause at throw in.

As for video contents personalization, we carried out a questionnaire to 15 students about which video contents are more preferable, the personalized soccer video produced by the method described in Section 6 or the non-personalized video produced by the method described in Section 4. 10 students preferred the personalized soccer video and 5 students preferred the non-personalized soccer video. The latter 5 students told that they wanted to watch the video contents with players close-up, formation analysis, individual techniques and multi-camera angles. These video contents are out of our method so that, without these 5 students, the personalized soccer video produced based on the private preference has significant superiority.

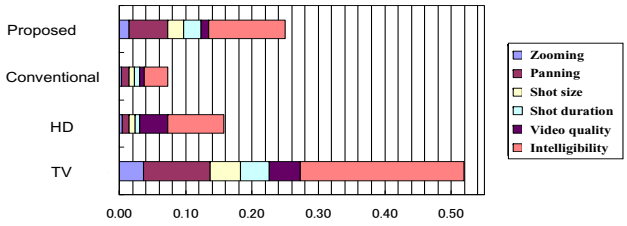

Figure 11. Evaluation result by AHP.

8. Conclusions

In this paper, we 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 can catch up with the techniques of the TV camera man in future.

The evaluation of video quality was less than TV contents. The digital zooming causes the serious deterioration on the video resolution. However, there are some techniques of digital zooming that can prevent the deterioration on video quality to some degree. As a future work, we will incorporate the techniques.

It is significantly meaningful that the video content produced by the proposed method is preferable to the HD content with fixed camera, because the produced contents improved the intelligibility compared to the fixed video contents. It becomes possible to solve the problems such as cost, lack of video contents as well as staff at broadcast stations, caused by the channel increase, since our proposed method can make use of the video contents taken by the fixed HD camera.

At present, the system does not work in real time. This means that SD videos are produced for individual person according to his preference after recording the soccer sports video in Hi-vision. However, if the special LSI is available, it will achieve the real time production system in future.

In addition, our proposed system can not only produce the video contents but also retrieve the scene in the video contents by utilizing the recognized situation or events as the indices. To make the system more advanced, we will develop the formation recognition, switching of multi-angle cameras and player's face recognition in future.

References

[1] Dennis Yow,Boon-Lock Yeo,Minerva Yeung,Bede Liu: "ANALYSIS AND PRESENTATION OF SOCCER HIGHLIGHTS FROM DIGITAL VIDEO", ACCV'95,pp.499-503,1995.

[2] Jurgen Assfalg, Marco Bertini, Carlo Colombo, Alberto Del Bimbo, Walter Nunziati: "Automatic Interpretation of Soccer Video for Highlights Extraction and Annotation", SAC 2003: pp.769-773, 2003.

[3] A. Ekin, A. M. Tekalp, and R. Mehrotra: "Automatic soccer video analysis and summarization", IEEE Trans. on Image Processing, vol. 12, no. 7, pp.796-807, July 2003.

[4] Ngoc Thanh Nguyen, Tuoung Cong Thang, Tae Meon Bae, Yong Man Ro: "Soccer Video Summarization System Based on Hidden Markov Model with Multiple MPEG-7 Descriptors", CISST 2003: pp.673-678, 2003.

[5] Thomas Bebie, Hanspeter Bieri: "SoccerMan - Reconstructing Soccer Games from Video Sequences". ICIP, pp.898-902, 1998.

[6] Hyunwoo Kim, Ki-Sang Hong: "Soccer Video Mosaicing Using Self-Calibration and Line Tracking". ICPR 2000: pp.1592-1595, 2000.

[7] Okihisa Utsumi, Koichi Miura, Ichiro Ide, Shuichi Sakai, Hidehiko Tanaka: "An object detection method for describing soccer games from video", Proc. 2002 IEEE Intl. Conf. on Multimedia and Expo (ICME2002), vol.1, pp.45-48, Aug. 2002.

[8] Takayoshi Koyama, Itaru Kitahara, Yuichi Ohta: "Live Mixed-Reality 3D Video in Soccer Stadium", ISMAR 2003: pp.178-187, 2003.

[9] T.Yokoi, H.Fujiyoshi: "Virtual Camerawork for Generating Lecture Video from High Resolution Images", Proc. 2005 IEEE Intl. Conf. on Multimedia and Expo (ICME2005), 2005.

[10] T.Yokoi, H.Fujiyoshi: "Generating a Time Shrunk Lecture Video by Event Detection", Proc. 2006 IEEE Intl. Conf. on Multimedia and Expo (ICME2005), pp.641-644, 2006.

[11] http://www.megavision.co.jp/

[12] Saaty, T.: "A scaling method for priorities in hierarchical structures", Journal of Mathematical Psychology, Vol. 15, pp.234-281 (1997)

[13] Masahito Kumano, Yasuo Ariki and Kiyoshi Tsukada: "A Method of Digital Camera Work Focused on Players and a Ball – Toward Automatic Contents Production System of Commentary Soccer Video by Digital Shooting –", 2004 Pacific-Rim Conference on Multimedia (PCM2004), pp.466-473, 2004-12.