AUTOMATIC PLAYER'S VIEW GENERATION OF REAL SOCCER SCENES BASED ON TRAJECTORY TRACKING

Nozomu Kasuya, Itaru Kitahara, Yoshinari Kameda, and Yuichi Ohta

Graduate School of Systems and Information Engineering, University of Tsukuba

ABSTRACT

This paper proposes a method to generate a player's view video of an actual soccer match using a 3D free-viewpoint video technique. Users can enjoy 3D video simply by choosing a target player and concentrating on a soccer match from the player's view by manually controlling the viewing position. The generated video provides an immersive sight as if the user is running on the pitch. To generate the player views, the target player's 3D trajectory must be estimated. We developed a novel computer vision technique for player tracking that robustly works in an actual soccer stadium. In the current system, the orientation of the virtual camera for the player's view does not follow the gaze direction of each player because the image resolution is too poor to acquire gaze direction by computer vision. Users can choose a favorite orientation control method. We applied the proposed method to an actual soccer match held in an outdoor stadium to confirm its effectiveness.

Index Terms—Computer vision, Computer graphics, Virtual reality, Object detection, Tracking

1. INTRODUCTION

3D video technique that virtually captures the appearance of real objects from arbitrary viewpoints is one of the most active topics in computer vision and graphics [1]-[4]. The broadcast of sport scenes will be an attractive application of 3D video, since each member of the audience has his/her own favorite way to watch such scenes. In ordinary 3D free-viewpoint video systems, however, concentrating on watching the soccer scene is difficult because users get distracted from the match by controlling the input devices (e.g., a mouse and a keyboard) to specify a desirable pose of a virtual camera. Our approach solves the problem by generating a "player's view" video stream.

2. PLAYER'S VIEW

As shown in Fig. 1, our proposed player's view is a virtual view from a player's viewpoint. When capturing such views with normal photographic devices, we need to put an actual camera on a player's body. However, this is obviously not feasible, since it interferes with the play. Instead, we generate a player's view using 3D free-viewpoint video techniques that can generate appearances virtually captured from arbitrary viewpoints. Although the area where observers can move the viewpoint is limited to the players' standing positions, the operation for controlling a virtual camera is very simple: choosing a player from whom to watch the desired sight. It is possible to concentrate on watching matches without being disturbed by the operation. The generated video also provides an immersive sight: standing near players on a soccer field. Furthermore watching a decisive moment of a soccer game from top players' view is useful for player development and training.

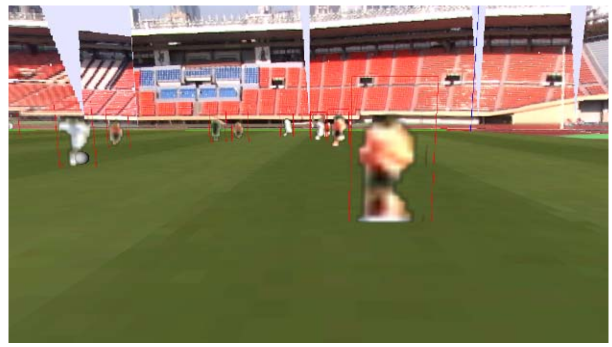

Fig. 1 Snapshot of player's view

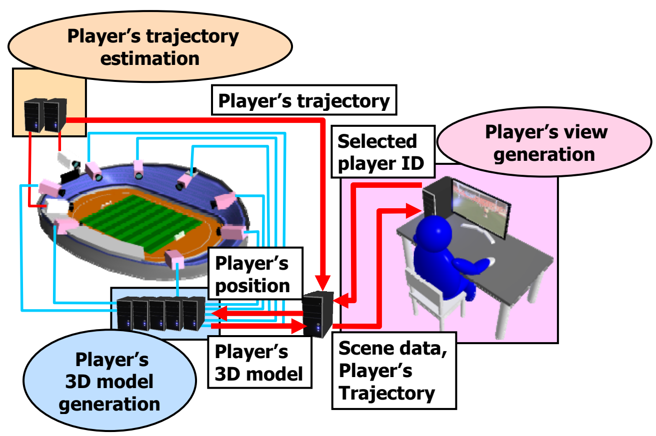

Fig. 2 Player's view generation system

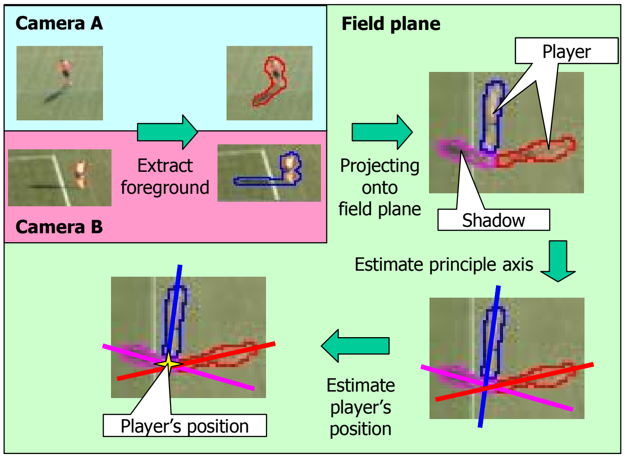

Fig. 3 Player's position estimation

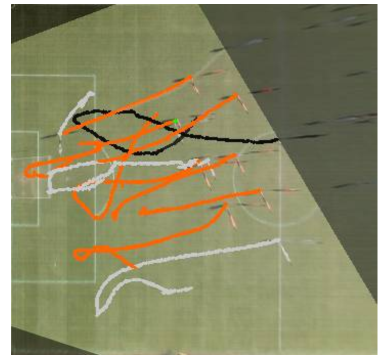

Fig. 4 Estimated trajectories of players

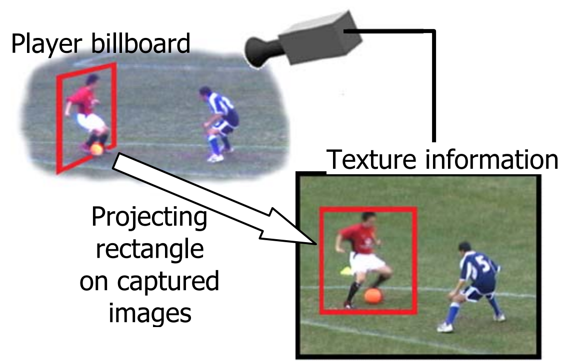

Fig. 5 3D modeling by player billboard

3. PLAYER'S VIEW GENERATION

Fig. 2 illustrates the system configuration of our proposed method. First, the players' 3D positions in the soccer scene are detected by multiple fixed cameras, and their trajectories are estimated by connecting the detected positions along with a time sequence. The 3D models of the players are generated by applying the estimated player position to a "player-billboard" technique [6]. The scene data set is constructed by bundling the models of all players in every frame. When a user selects a player from whom to observe the desired sight, the virtual viewpoint is moved to the player's viewpoint, and the sight is virtually captured.

3.1. Player's trajectory estimation

The information of the player's trajectories is essential to generate the sight from a player's view. When we estimate the player's trajectory in an actual soccer stadium, the following problems must be solved. (1) The region size in the captured image of each soccer player is small due to the limited resolution of the cameras. (2) Since soccer is a contact sport, many occlusions and interactions are observed in the images. (3) Sometimes the sun makes strong player shadows against position estimation. (4) To provide a player's view broadcasting service, a trajectory with short latency must be estimated. We proposed a method to estimate player trajectories under such difficult conditions [5].

We captured soccer scenes using two calibrated cameras to reduce the effects of occlusion and shadow problems. As shown in Fig. 3, we projected the foreground regions in the two captured images onto a soccer field and estimated a player's position as the intersection of the primary axes of the two projected body regions and a shadow region. Furthermore, to reduce false tracking during "close play," the method classifies the detected players by uniform color. Here, we assume it is not so often that the players of a same team play with standing close or conflicting each other.

Fig. 4 shows the player trajectories obtained by this method. Although the tracking accuracy, which is the detection ratio of trajectories with detecting a player's position in the whole frames, is only about 70%, it is enough to generate player views because some player trajectories are completely estimated in the sequence. In other words, more than 15 players' views are provided by our system.

3.2. Player's 3D model generation

Player models are generated with "player-billboard" proposed by Koyama et al. [6]. Player-billboard represents each player by one rectangle and an associated texture. We generate a 3D model by merging the textures extracted from multiple videos. Since we do not reconstruct precise 3D shape models, the processing cost to generate 3D models can be reduced. Moreover, the data size for describing the 3D model is small because each player is simply represented.

We can make it simple to extract the texture by using estimated player's position. We assume that all players are standing vertically and are almost the same size: 2-m tall and 1-m wide. As shown in Fig. 5, a rectangular plane is placed at the player's position in actual 3D space and the 3D

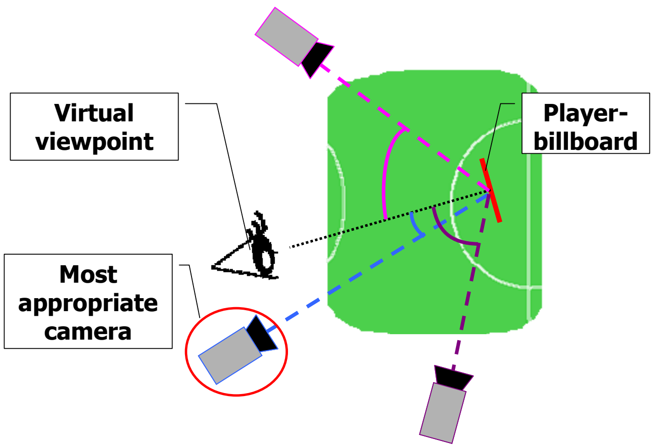

Fig. 6 Texture selection for a player billboard

coordinates of its vertices are projected onto the multiple images. A region projected rectangle is assumed to be the player's texture region. The texture of each player is segmented from the background using the background subtraction method.

As shown in Fig. 6, when a target player is captured by multiple cameras, one of them must be chosen. The player-billboard vertically stands at the player's position and faces the viewpoint. The most appropriate texture is selected based on the difference of the observing angle between the viewpoint and a capturing camera.

3.3. Player's view generation

To generate the player views, the 3D position and orientation of the selected player viewpoints are needed. The horizontal position is extracted using the tracking method described in Section 3.1, and the viewpoint's height is fixed at 1.5 m, which is the average human's eye-level. However, stably estimating the view direction is difficult by processing long distant view images, and if a view is presented of another intensely moving person, it may cause a kind of VR sickness. Consequently, we designed the gaze point to trace the 3D position of a soccer ball which players usually gaze at while playing. And our system also provides the optional interface so that users can choose to look at their favorite orientation based on their own discretion. Although users have to operate the gaze direction, it does not obstruct the video appreciation. Because users usually keep their eyes on something in the soccer field, operation is easy with a mouse to control the gaze direction that has two degrees of freedom. A 3D model of the background stadium was generated by the billboard-based method proposed by Nomura et al. [7].

4. EXPERIMENTS

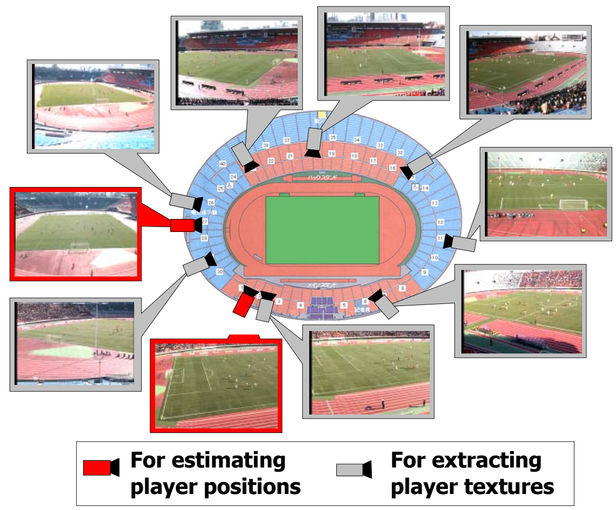

Fig. 7 Layout of cameras in soccer stadium

Table 1 Processing time to generate player's view

Player's trajectory estimation	3D player model generation	Player's view generation
107 ms	7.68 ms	2.13 ms

4.1. Experimental environment

We conducted experiments at Tokyo Olympic Stadium, an actual soccer stadium. As illustrated in Fig. 7, half of the soccer field was captured by ten cameras: two for estimating player trajectories and eight for extracting player textures. The cameras were Sony DXC-9000, which can capture VGA (640 x 480 pixels) progressive images at 30 (fps). The relationship between the camera and the stadium coordinate systems is calibrated using 3D laser-based surveying equipment. We obtained a set of video sequences about 30 seconds from an actual soccer match at the stadium in daytime when the player shadows are strong. In this experiment, all processes were executed on a PC with an Intel Core2Duo2.33 GHz CPUs, 2GB RAM, and a video board with an NVIDIA GeForce 8300 GS GPU.

4.2. Results

Fig. 8 shows the generated player's view video and the field of view (the yellow fan-shaped region). In the sequence, the gaze point is moving to look at a ball. It is confirming that the video provides immersive sight standing on the soccer field is generated. The operation of the virtual viewpoint is very easy to input; one merely chooses the gaze direction.

Table 1 shows the processing time to generate player views. Although the player's trajectory estimation takes time, real-time processing can be realized by optimizing and a pipeline technique. For the future, we aim to construct a system with which multiple users can simultaneously watch player views by connecting all parts.

5. CONCLUSION

This paper proposed a method to generate video of player views using a 3D free-viewpoint technique. Since the 3D positions of players are necessary to generate the video, we developed a player tracking method that robustly works in an actual soccer stadium. Since stably estimating the gazing direction of each player by distant view images of soccer scenes remains difficult, the system left the choice to look at their favorite orientation to user discretion to maintain the quality of the generated video. The 3D models of players were generated using a player-billboard technique. We applied the proposed method to a soccer match held at Tokyo Olympic Stadium. The generated video showed the effectiveness of the proposed method, which gives an immersive sight of standing on the soccer field and can be observed with easy operation compared with ordinary 3D free-viewpoint video systems.

When the virtual viewpoint gets too close to the other players going onto the field, the roughness of player textures obstructs the quality of the generated video. Generating player views with less sense of discomfort by capturing high resolution videos remains future work.

REFERENCES

[1] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes," IEEE Multimedia 1997, Vol.4, No. 1, pp. 34-47.

[2] W. Matusik, C. Buehler, R. Rasker, S. J. Gortler, and L. McMillan, "Image-Based Visual Hulls," ACM SIGGRAPH 2000, pp.369-374.

[3] Joel Carranza, Christian Theobalt, Marcus A. Magnor, and Hans-Peter Seidel, "Free-Viewpoint Video of Human Actors," ACM Transaction on Graphics, vol. 22, No. 3, pp. 569-577, July 2003.

[4] N. Inamoto and H. Saito, "Immersive Observation of Virtualized Soccer Match at Real Stadium Model," ISMAR 2003, pp. 188-197.

[5] Nozomu Kasuya, Itaru Kitahara, Yoshinari Kameda, and Yuichi Ohta, "Robust Trajectory Estimation of Soccer Players by Using Two Cameras," Proceedings of IAPR 19th International Conference on Pattern Recognition, 4 pages, 2008.

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

[7] Naobumi Nomura, Itaru Kitahara, Yoshinari Kameda, and Yuichi Ohta, "A Background Modeling Method with Simple Operations for 3D Video," Proceedings of 3DTV Conference 2007, 4 pages, 2007.

Fig. 8 Examples of player's view (left): corresponding position and field-of-view of virtual camera (right)

ACKNOWLEDGMENT

We express an acknowledgment to Dr.Katsuyoshi Shirai (JISS: Japan Institute of Sports Sciences) and Dr.Fumito Yoshikawa (AIST: National Institute of Advanced Industrial Science and Technology) who cooperated on pushing forward the experiments in Tokyo Olympic Stadium,.