Mutliple Object Tracking for End-to-End Transmission System of Free-Viewpoint Video

Shogo Miyata, Hideo Saito
Dept. of Information and Computer Science
Keio University
Yokohama, Japan

***Abstract*—** In this paper, we propose a framework for end-to-end transmission system of free-viewpoint video. On soccer game taken multiple cameras, we adapt efficient multiple object tracking method to our system and generate free-viewpoint video represented billboard.

Keywords- *Free-viewpoint video; multiple object tracking;billboard*

I. INTRODUCTION

Free-viewpoint video allows user to interactively choose the arbitrary viewpoint in 3D space. Free-viewpoint video will be new interactive contents that is not provided unilateral by broadcasting companies but viewer has decision of viewpoint selection.

Although a lot of researches on free-viewpoint video [4] [5] [6] have been done, there has not been commercialized yet. It is necessary to establish the pipeline from broadcasting companies to viewers, in other words, we should develop end-to-end transmission system of free-viewpoint video, which converges capturing the video contents, processing to create free-viewpoint video and distributing free-viewpoint video as application to costumer, as shown in Fig 1.

Our purpose is to achieve the part of processing free-viewpoint, which contributes the development of whole system. For free-viewpoint video rendering, we use a billboard to represent 3D model of each object. For making free-viewpoint video in a soccer stadium, it is described in [10] that billboard method is superior to the shape-from-silhouette method because it is more robust to noise and it also better represents smooth motion as the number of cameras is increased. Accordingly, we also generate each player's billboard as output from the input of the multiple synchronized video sequence of soccer games.

To generate billboard, we use OpenMV (Open Multiple View Framework), which is developed for our end-to-end transmission system and provides the framework for sharing various kinds of data of the end-to-end multiple view video processing. For creating billboard, we need to analysis multiple view images and extract player's information.

Our billboard representation consists of the following components: color images which include player, mask image which is representing the silhouette of player, intrinsic and extrinsic parameters of camera, and player's location for every frame and player's ID. Identification of each player is not necessary if we only create free-viewpoint video, but we should often identify players for using in real application, such as creating movies that always takes only a specific player by virtually move the viewpoint.

Hence, we propose a free-viewpoint video rendering system in soccer game based on player detection and tracking for end-to-end transmission system. We have adapted efficient multiple object tracking method of [1] to our system. At first, we extract the silhouettes of players by background subtraction, and with the binary images, we estimate the probabilistic occupancy map (POM), stands for probability of ground plane occupancy in each frame. Given the results of estimating POM, we compute heuristic global optimization of player's trajectories with dynamic programming.

At the stage of free-viewpoint rendering, we create billboard with the results of tracking. To create billboard, we correct color image, mask image, camera parameters, player ID, player's position and temporal synchronizing information. After generating billboard with the results of tracking, these objects are rendered on CG soccer stadium by Unity.

The paper is organized as follows: we first briefly review related works. Then, we describe proposed system of multiple object tracking for free-viewpoint video rendering. Finally, we show experimental results of multiple players tracking and free-viewpoint video.

II. RELATED WORKS

A lot of researches on free-viewpoint rendering using the movie captured from multiple cameras has been done. In particular, using in sports scene is expected for virtual replay or new TV contents that user can choose the point of view. However, it is difficult to adapt classical techniques for free-viewpoint video to sports scene because of larger area and uncontrollable condition. We should also take care of computational cost so that we can distribute the free-viewpoint video.

Guillemaut and Hilton [4] proposed a technique to segment the images of multiple cameras into background and foreground layers and estimate the depth of each pixel. This approach can treat severe environment but it is too expensive because they reconstruct 3D model.

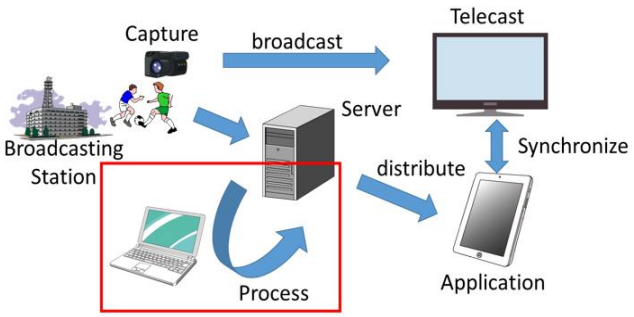

Figure 1 End-to-End Transmission System of free-viewpoint video

Billboard representation [5] [6] is good solution for application of distributing free-viewpoint video. They approximate the foreground players as rectangle and achieve free-viewpoint video rendering keep the cost.

To generate billboard, we have to analyze the images with multiple object tracking and extract the necessary information for representing each object as a billboard. Most of conventional approaches for multiple object tracking rely on the recursive tracking from frame to frame. Temporal filtering technique such as a Kalman filtering [7] and particle filtering [8] is an efficient approach to track multiple object. They are suitable for real-time processing because they are recursive method.

In [2] and [9], they formulate the player's trajectory as graph whose nodes represent all potential locations over time. They are much more robust due to solve optimal global solution with linear programming.

III. PROPOSED SYSTEM

The overview of our system is shown in Fig 2. Our purpose is to obtain free-viewpoint video by representing every player's 3D structure width a billboard moving on pre-captured 3D model of stadium. We should detect and track players to extract player's location for every frame and identify each players. We have adapted efficient multiple object tracking method of [1] to our system for free-viewpoint rendering.

In the pre-process stage of the system, we estimate the intrinsic parameters and camera pose of each color cameras. After the pre-process, we extract the silhouettes of players by background subtraction. With the binary images, we estimate the probabilistic occupancy map (POM) of the players, which provides optimal trajectories of players width dynamic programming.

We discretize the ground of soccer field into grids 50 cm × 50 cm for estimation of player's position.

A. *Player Detection*

To produce the POM, we first generate the foreground blobs of multiple images acquired simultaneously calibrated cameras by background subtraction algorithms [3].

POM algorithm [1] uses a generative model to estimate the most probable positions of players under the obtained foreground images. At every time frame t, and for every location k of the grid, it estimates marginal posterior probability of presence of a player at that position. This algorithm can perform well at the scene of occlusions because of taking account of information observed in each camera.

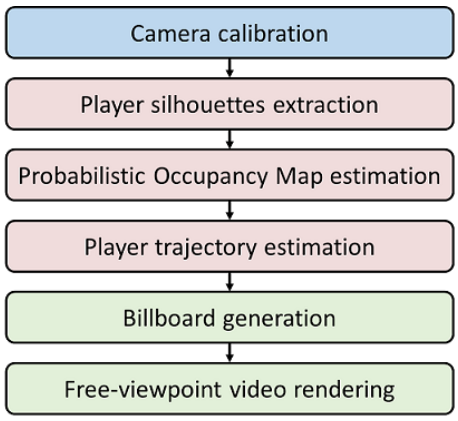

Figure 2 Proposed system overview.

B. *Player Trajectory Estimation*

In this section, we estimate the optimal trajectories of player with the results of POM algorithm. We use multiple object tracking algorithm of Fleuret, Berclaz, Lengagne and Fua [1] with dynamic programming.

If we estimate the trajectories recursively from frame to frame, we may fail to track when detection is unsuccessful in a frame. To avoid such failure of tracking, we process the video sequences by batches that are sufficiently long to compute the most optimal trajectory for each player.

In one batch, we compute the one most likely trajectory after the other. At first, we find the trajectory which minimize the total cost. We define the cost of trajectory at location k at time t as (1).

$$cost_t(k) = w_p c_p + \min\,(w_d c_d + cost_{t-1}(l)), \tag{1}$$

where c_p denotes the cost about the probability of presence of player and defined as (2).

$$c_p = 1 - p^t{}_k, \tag{2}$$

where $p^t{}_k$ represents the probability of presence of player which derived the results of POM. c_d is the cost about the distance of player between location k at current frame and location l at previous one and defined as (3).

$$c_d = \|k - l\| \tag{3}$$

Equation (1) means recursive expression, which we can solve with dynamic programming. After seeking the one trajectory, we repeat the computing of the optimal trajectory except for

the locations which belong to trajectories of the other players, so that we can get trajectories of all players in one Batch. To estimate the trajectories of the next batch, we start computing of the trajectory with reliable player in previous batch by keeping the result of first several frames of previous batch, so that we can prevent from confusing trajectories and we can get reasonable heuristic global solution.

C. *Billboard Generation*

To generate billboard, we use OpenMV (Open Multiple View Framework), which provides the framework for sharing various kinds of data of the end-to-end multiple view video processing.

For creating billboard representation, we need to correct the following data: color images including the player, mask images of player region, camera parameters, temporal synchronizing information, and player's position and ID number. Mask images are obtained by background subtraction at the player detection. Camera parameters are computed at the pre-process stage. Player's position and temporal information and ID are results of tracking. We integrate these elements for creating billboard representation of every player for all frames. These objects are rendered on CG soccer stadium by Unity.

IV. RESULTS

In this section, we introduce our experiment of multiple players tracking and free-viewpoint movie rendering of soccer game. We use the video sequence of outdoor soccer match, they are captured by three Canon XF305 with a resolution of 1920 × 1080 and twelve Canon XHG1 with a resolution of 1440 × 1080 at the gallery of stadium. The computation is done on a PC with Core i7, 2.8 GHz CPU and 16 GB Memory.

A. *Player Tracking*

Fig 3 presents the example of results of multiple players tracking. Although very small player silhouette, severe sunlight, high contrast between light and shadow and ambiguous uniform color against the field, player tracking has successfully be performed. However, there have been some failures, missing player and switching individual.

As shown in Fig 4, red rectangle misses the player. Poor quality of result of background subtraction causes the wrong occupancy map estimation and influences tracking, although it is stable to be unsuccessful in detection of a little frames. To avoid this, we have to use more efficient background subtraction algorithm.

Fig 5 shows the switching the trajectories of individual, left image is after several frames of right image and we can see the number was exchanged. Because we only use the information of location, when used to track people in the case of crowd, it causes the identity switching. A potential solution is to use color information or player's direction of movement.

B. *Free-Viewpoint Video Rendering*

We created the billboards with the results of multiple players tracking and free-viewpoint video. Free-viewpoint video rendering is processed using Unity. Billboards are rendered on CG soccer stadium. As move of user's viewpoint, represented image on billboard is smoothly changed, which is chosen nearest. Users can control camera position and direction anywhere and choose any point of view while the players are moving on the field. The examples of results of the free-viewpoint video rendering are presented in Fig 6 and Fig 7. Since each player's trajectory is tracked by the proposed method, we can easily synthesize the first person's view of each player as shown in Fig.7.

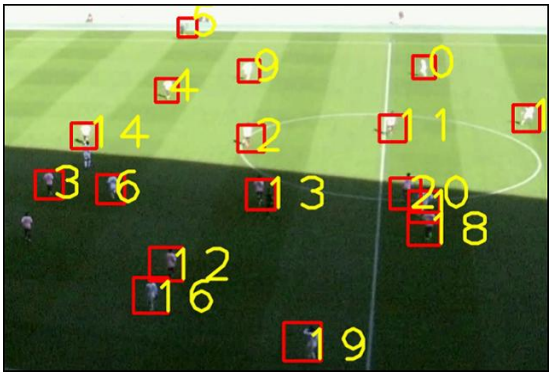

Figure 3. Example of the results of player tracking

We discuss the evaluation of them. These billboards represent the location of player relatively high accuracy without collapsed tracking. The quality of player's silhouette is depend on the results of background subtraction as mask image used to create billboard. In this case, the foreground blobs is not so fine because of the reasons mentioned in player tracking and some of the billboards are not good shape. We also extract the silhouette from the corresponding rectangle area of each cameras which derived from the results of tracking, the error of camera calibration affects the projection of player's location. In addition, we ignored the trajectory of soccer ball and it is necessary other process of ball tracking. Finally, we completely ignored the occlusion of players at rendering and we can notice that some billboards include several players.

V. CONCLUSIONS

We proposed a framework to realize development an end-to-end transmission system of free-viewpoint video. We adapted the efficient multiple object tracking method to our free-viewpoint video rendering system. Multiple player tracking is achieved by POM, which is the probability of ground plane occupancy in each frame, and heuristic global optimization of trajectories of players. Free-viewpoint video rendering is achieved by generating billboard with the results of tracking using OpenMV and Unity.

For future works, we will focus on improvements of accuracy of multiple players tracking. We have to use more efficient background subtraction algorithms. We can take account into colors of player's uniform and direction of player's movement. Furthermore, we will focus on generate high quality billboards. The possibilities are to enhancement of

extraction of player's silhouette, improvement of accuracy of camera calibration and handling with occlusions of players.

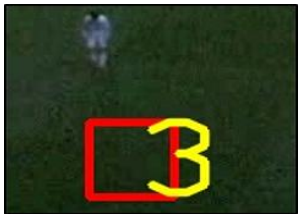

Figure 4. Example of missing the player.

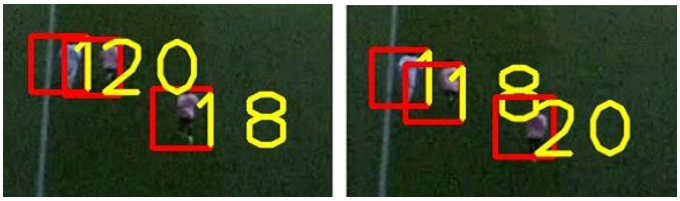

Figure 5. Example of switching the trajectories. Left image is after several frames of right image.

Figure 6 Examples of free-viewpoint video rendering

ACKNOWLEDGMENT

This work has been supported in part by National Institute of Information and Communications Technology (NICT), Japan.

REFERENCES

[1] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera People Tracking with a Probabilistic Occupancy Map," IEEE Transactions on pattern analysis and machine intelligence, vol. 30, no. 2, pp.267-282, 2008.

[2] Jerome Berclaz, Francois Fleuret, Engin Turetken, and Pascal Fua, "Multiple object tracking using k-shortest paths optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 9, pp. 1806-1819, 2011.

[3] Zoran Zivkovic and Ferdinand van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," Pattern Recognition Letters, Vol. 27, No. 7, pp. 773-780, 2006.

[4] Jean-Yves Guillemaut and Adrian Hilton, "Joint multi-layer segmentation and reconstruction for free-viewpoint video applications," International Journal of Computer Vision, Vol. 93, No. 1, pp. 73-100, 2011.

[5] Takayoshi Koyama, Itaru Kitahara, and Yuichi Ohta, "Live mixed-reality 3d video in soccer stadium. In International Symposium on Mixed and Augmented Reality," pp. 178-186. IEEE, 2003.

[6] Kunihiko Hayashi and Hideo Saito, "Synthesizing free-viewpoint images from multiple view videos in soccer stadiumadium," In International Conference on Computer Graphics, Imaging and Visualisation, pp. 220-225. IEEE, 2006.

[7] Sachiko Iwase and Hideo Saito, "Parallel tracking of all soccer players by integrating detected positions in multiple view images," In International Conference on Pattern Recognition, Vol. 4, pp. 751-754. IEEE, 2004.

[8] Kevin Smith, Daniel Gatica-Perez, and Jean-Marc Odobez, "Using particles to track varying numbers of interacting people," In Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 962-969. IEEE, 2005.

[9] Horesh Ben Shitrit, Jerome Berclaz, Francois Fleuret, and Pascal Fua, "Tracking multiple people under global appearance constraints," In International Conference on Computer Vision, pp. 137-144. IEEE, 2011.

[10] Tetsuya Shin, Nozomu Kasuya, Itaru Kitahara, Yoshinari Kameda, and Yuichi Ohta, "A comparison between two 3d free-viewpoint generation methods: Player-billboard and 3d reconstruction," In 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video, pp. 1-4. IEEE, 2010

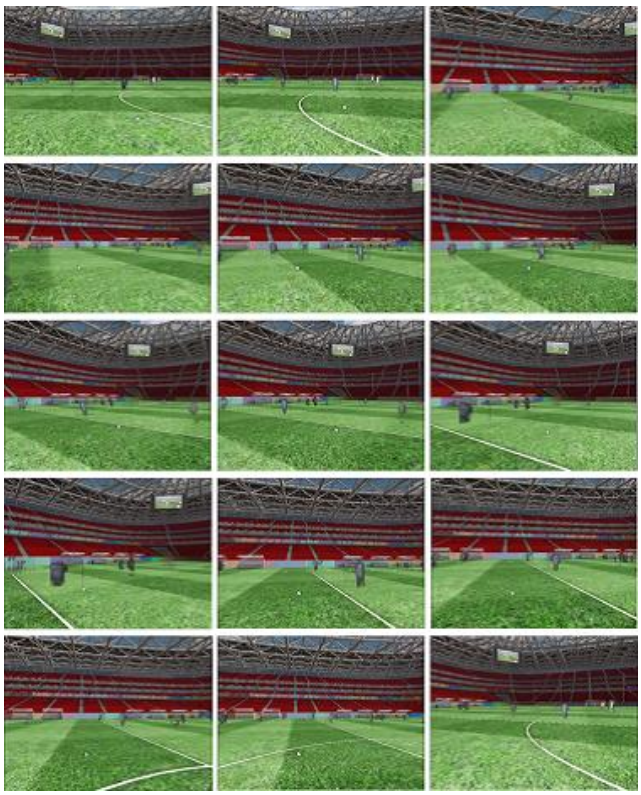

Figure 7 Sequence of first person's view of one player.