Free Viewpoint Browsing of Live Soccer Games

Yoshinari KAMEDA[1], Takayoshi KOYAMA[2], Yasuhiro MUKAIGAWA[3], Fumito YOSHIKAWA[4] and Yuichi OHTA[5],
University of Tsukuba, Department of Intelligent Interaction Technologies[1-3,5]
Japanese Institute of Sports Sciences[4]
kameda@ieee.org[1], {koyama,mukaigaw}@image.esys.tsukuba.ac.jp[2,3]
yoshikawa.fumito@jiss.naash.go.jp[4], ohta@computer.org[5]

Abstract

We present a new video browsing method for multiple videos that are taken at large-scale space for live 3D events such as soccer games. By our method, multiple viewers over computer network can browse a live 3D event from any viewpoint and each viewer can move his/her viewpoint freely. Our algorithm consists of five steps. Our system first captures videos from multiple cameras, then extracts texture segments from the videos, selects appropriate segments according to a viewpoint which is given by user dynamically, transmits them to users, and layouts the segments in virtual space so that each viewer can see the segments in a virtual environment as if the viewers were in the event. Our 3D video display system requires 10Mbps at most to browse a soccer game. We conducted experiments at two real soccer stadiums and succeeded in realizing live realistic visualization with free viewpoint at about 26 fps.

1. Introduction

Events held in a large-scale space like sports, plays, and games are so valuable that many cameras are installed to film them. However, it is always a serious problem how to film the events held in large-scale space and browse videos so that users are satisfied, especially when users have different preference to see the events. The perfect solution is to provide viewpoint-free browsing method to every user.

One of the advanced approaches for filming an event in large-scale space is EyeVision[1] system which was invented by Kanade and PVI Inc. This is ideal for broadcasting purpose because it can give any view of a focused object with arbitrary view direction. However, since it can focus on only one object at a time, it can not satisfy different viewers who want to see the event from different view angle simultaneously.

Constructing a whole 3D scene virtually from multiple videos is considerably a good solution to realize viewpoint-free access to 3D events. In computer vision literature, 3D visualization systems that utilize multiple cameras have been proposed[2]. Although they succeeded in reconstructing a 3D view of object(s), their approaches are not feasible for real-time use in large-scale space because processing cost is not small in the case objects are sparsely located in a large-scale space.

In this paper, we propose a new visualization method that realizes 1) viewpoint-free visualization in a large-scale space, 2) simultaneous access from multiple viewers, and 3) live browsing that is realized by our sophisticated algorithm. We present billboard-based visualization method that enables us to reduce both computation and data transmission cost.

In section 2, we explain billboard-based visualization. Billboard extraction method is mentioned in section 3. Experimental results of soccer games are shown in section 4. At last, we conclude this paper in section 5.

2. Billboard-based visualization

Visualization of human activities like soccer games in a large-scale space needs different property from the one in a small-scale space. Figure 1 shows the overview of both visualization approaches. In small-scale space, people usually pay attention to detailed action of target player (figure 1 left). On the other hand, it is considered that people are mainly interested in watching spatial relationship of players together with their actions when they watch human activities in large-scale space (figure 1 right). Therefore, a 3D model to represent each player can be simplified so far as it expresses his/her actions to viewers who watch the activities in middle or far range.

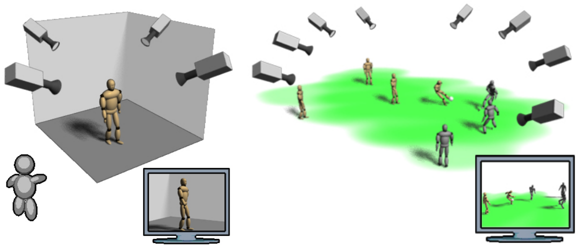

Figure 1. Visualization in small-scale space and large-scale space

We use billboard representation[3] to simplify the 3D model. A player billboard is a small rectangle standing perpendicular to the ground, and 2D texture of one player is shown on it. Figure 2 shows the appearance difference between full 3D shape reconstruction method (figure 2 left) and our method (figure 2 right). Although visual difference is obvious and distorted texture in the billboard does not look like a 3D object at close viewpoint (figure 2 top right), the difference becomes very small at a distant viewpoint (figure 2 bottom) if player billboards are displayed at appropriate location and direction. In this sense, it is important to place player billboards at a right place at right direction. We will discuss how to extract these data in the next section.

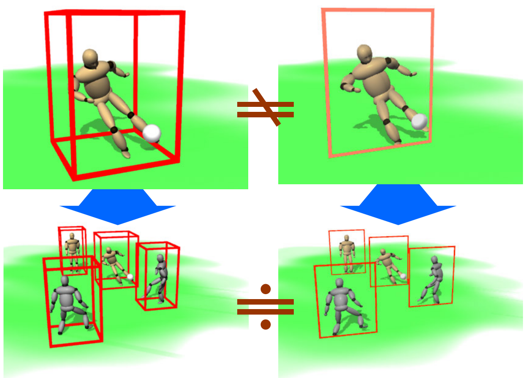

Figure 2. Appearance similarity in close-view and distant-view

3. Producing player billboards

3.1. System overview

In our system, the process cycle to realize free viewpoint browsing consists of five steps. First our system captures videos at multiple cameras, then extracts texture segments from the videos, selects appropriate textures according to viewpoint which is given by user dynamically, transmits them to users, and finally layouts the player billboards in virtual space (figure 3). Note that any viewer can control viewpoint by his/her mouse freely in virtual world at viewer-side computer.

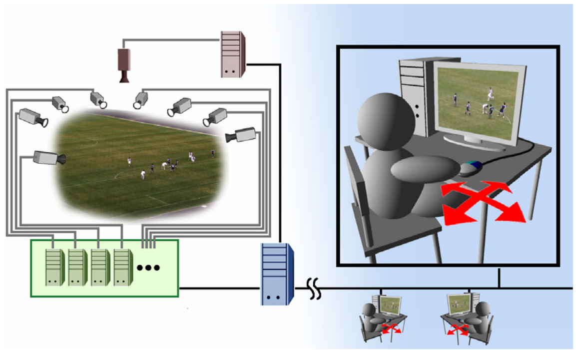

Figure 3. System overview

3.2. Texture extraction

Suppose there are N cameras to obtain billboard textures. They are calibrated in advance. The system obtains 3D location of each player $\boldsymbol{L_p}$ by using one or two ceiling cameras, and extracts corresponding texture segments at a certain size in every video image by projecting $\boldsymbol{L_p}$ onto the image plane. Background region in the texture is removed at this step by video capturing PC. Figure 4 shows the case where two player billboards are obtained at one camera.

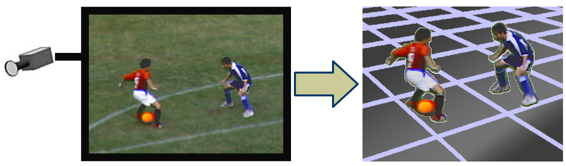

Figure 4. Produced billboards

3.3. Texture selection

The number of obtained textures for one player should be from 0 to N. If there are M players, total number of textures will be N x M at most. Billboard server collects all the texture segments and it selects a set of textures to be sent to each viewer based on his/her viewpoint $\boldsymbol{L_v}$. Suppose there are N=3 cameras in figure 5. There are three textures for player p. When $\boldsymbol{L_v}$ is given, find camera n so that it minimizes the angle $\boldsymbol{\theta_n}$ between the line from the viewpoint $\boldsymbol{L_v}$ to $\boldsymbol{L_p}$ and the line from camera C_n to $\boldsymbol{L_p}$ for all n. In the case of figure 5, a texture segment obtained by C_2 is selected.

The selected textures are going to be sent to a viewer-side computer and the computer carefully places them so that the textures face the viewpoint $\boldsymbol{L_v}$ squarely in the virtual space (figure 5). Therefore, M player billboards are sent to one user at maximum.

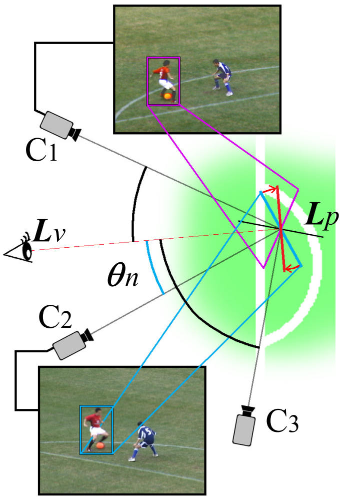

Figure 5. Billboard selection and display

3.3. Texture extraction

Extracting meaningful segments in video images is a key for effective multimedia browsing. Although background region in billboard texture is removed initially at video capturing step, it may include extra player when players are overlapped each other at a certain camera. Figure 6 shows such situation. If two textures are extracted for these two players on this camera, each texture contains extra foreground regions. As a result, it is hard to recognize players in live 3D display.

To eliminate overlapped extra player region, we use stereo based method [4]. For example, in figure 6, blue player should be removed to extract texture of red player (displayed as red square).

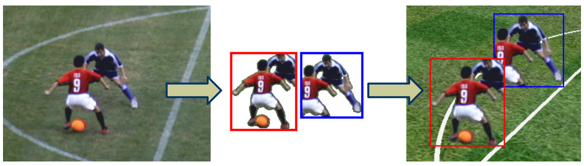

Figure 6. Overlapped texture

We prepare K parallel projective grid planes perpendicular to camera axis within $\boldsymbol{L_p}$ plus or minus B_{size}. Note that constant B_{size} implies the body size of a player. Initial texture image is mapped on each grid plane and the unit size of the grid corresponds to one pixel in the initial texture image. Then, video image from adjacent camera is projected onto each grid and overlapped regions are extracted. Finally, mask area is calculated by mapping all the overlapped regions onto image plane. Figure 7 shows the process of eliminating background player with $K = 3$.

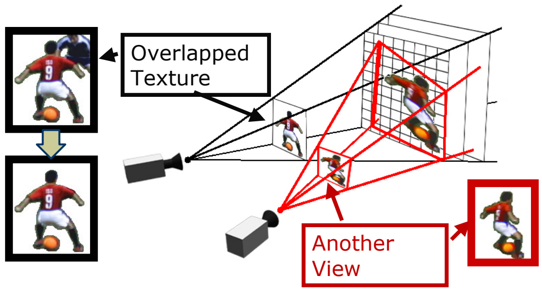

Figure 7. Eliminating background object (blue player) using adjacent camera image (red frame)

4. Experiments

We have implemented the proposed method and conducted two experiments. As our method is designed for large-scale space, our experiment needs big space like stadiums. The first experiment (experiment-A) was done on Oita stadium, Japan and the second experiment (experiment-B) was done on National Kasumigaoka Stadium, which is one of the most authoritative stadiums in Japan.

We installed 8 cameras to obtain textures. Video images were captured at size of 640 x 480 pixels.

In figure 8, precise layout of 8 cameras in the experiment-B is presented. View angle of the cameras is arranged so that each camera can cover the left half of the field.

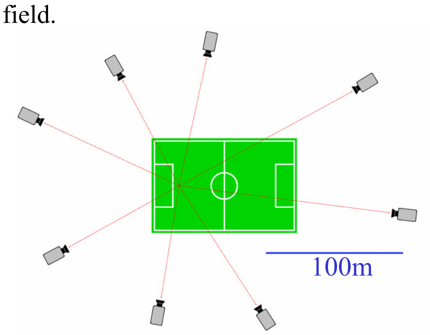

Figure 8. Camera layout in experiment-B

We set billboard size on the field to 1 meter wide and 2 meter high. The size of corresponding billboard texture image depends on the location relationship between cameras and players. In the experiments, billboard size was about 20 x 50 pixels in average and 40 x 100 pixels at maximum. K is set to 10 to eliminate overlapped textures. Table 1 shows data size of transmitting player billboards which includes both texture images and location information. Note that data to be transmitted was dramatically reduced in contrast with transmitting original videos while our prototype system can provide 3D free view of the field. If the data is compressed, the data size becomes small enough to transmit over broadband network.

Table 1 Data size

	1 player	22 players
1 [frame]	2 [KB]	43 [KB]
30 [frame]	0.5 [Mbps]	10 [Mbps]

Figure 9 shows 6 snapshots of experiment-A and figure 10 shows 6 snapshots of experiment-B. Our prototype system achieved about 26 fps in average at viewer-side computer. Stadium structures are installed in advance on the viewer-side computer. Note that viewers perceive as if they control their own flying cameras on the stadium simultaneously. We have also implemented replay browsing system on which users can control their viewpoint freely over archived games.

5. Conclusion

We have proposed a new browsing method for live 3D events held at large-scale space. We have implemented a prototype system that proved its visualization quality.

Acknowledgement

We thank National Kasumigaoka Stadium for their cooperation on experiment-B. This research is partially supported by Grant-in-Aid Scientific Research (A)(2) 14208034 of JSPS.

References

[1] http://www.ri.cmu.edu/events/sb35/tksuperbowl.html

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

[3] T. Akenine-Moller, E. Haines, "Real-Time Rendering", *AK Peters Ltd,* ISBN 1568811829, 2002.

[4] T. Koyama, I. Kitahara and Y. Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium", *Proc. of 2nd IEEE and ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR 2003)*, pp.178-187, 2003.

Figure 9. Results of experiment-A

Figure 10. Results of experiment-B