MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
8-25

Real-Time Transmission of 3D Video to Multiple Users via Network

Takayoshi KOYAMA, Yasuhiro MUKAIGAWA†, Yoshinari KAMEDA , Yuichi OHTA

Department of Intelligent Interaction Technologies,
Graduate School of Systems and Information Engineering, University of Tsukuba, Japan

{koyama, mukaigaw, kameda, ohta}@image.esys.tsukuba.ac.jp

Abstract

This paper presents a method to transmit live 3D videos for a soccer stadium and display them at remote PCs for users who watch soccer games. All processes are done in real time, including soccer scene capture with multiple cameras, image analysis, data transmission, and 3D video display. To generate and transmit 3D video via network in real time, each player is represented by "player-billboard" which is composed of one rectangle and associated live texture. Because a player-billboard does not require precise 3D shape reconstruction of each player, computation cost and data size required to generate live 3D videos are very low.

In order to show the validity of our method, we captured multiple videos in a real soccer stadium and transmitted 3D video to remote multiple users simultaneously. The result proved our system can transmit 3D video at 15-24 [frames/sec] through the Internet.

1. Introduction

Many cameras are installed to capture an event in large-scale spaces like a soccer game in a stadium. In traditional image media such as TV, however, viewers watch the event only from the viewpoint given by the director of the TV program. Hence the viewers can not control the viewpoint by themselves.

Recently, some advanced systems which allow viewers to control the viewpoint have been proposed. Virtualized Reality[1] is one of the early systems, which installed more than 50 cameras on a hemispherical dome with a radius of 5 meters. By analyzing all the captured images, the system reconstructs precise 3D shapes of the target objects and can render an image of the objects from any viewpoint. Würmlin has proposed a 3D video system with point-base 3D shape representation that has hierarchical structure[2]. This system can adjust trade-off between 3D video quality and computation cost. Matusik has proposed a method to render 3D video by Image Based Visual Hull[3]. The system can render 3D video about 8 fps. However, these systems can not perform all the processes in real time, including capturing multiple videos, processing images so as to reconstruct view models of the objects, transmitting appropriate data to remote PCs, and displaying 3D video at remote PCs on which users control their viewpoint on line.

In addition, the space in which target objects move is relatively small, e.g. room size, because their researches focus on precise 3D shape reconstruction of the target objects and therefore they limit the space size.

In contrast to these systems, we have proposed a novel visualization method which does not require 3D shape reconstruction[4][5]. Since our goal is to generate 3D video in a large scale space such as a soccer stadium and to transmit it to remote users in real time, our system is designed to reduce computation cost and data size so as to provide 3D video. In our method, each soccer player is represented simply by one rectangle and associated texture. We have developed a system which captures multiple videos in a soccer stadium, sets up the player-billboards, transmits minimal customized data to a remote user so as to render 3D video at a viewpoint specified by the remote user, and renders the 3D video on the remote user's PC. All processes are done in real time. We have presented the player-billboard generation and transmission framework[4][5] and showed the throughput of player-billboard generation with our previous system. In this paper, we examine the network performance of our system over the Internet.

2. Live 3D video system with player billboard

We have already proposed the method to generate live 3D video by utilizing player-billboard[4][5]. The method is suitable for generating 3D video in real time and transmitting it to remote users. In this section, we briefly describe the method. See [4][5] for detail.

2.1 System overview

We have developed a system to realize live 3D video with player-billboard. The overview of our system is shown in Figure 1. Our system consists of player position estimation part, multi-video capture part, player-billboard management part, and visualization part.

The player position estimation part has two cameras

† Now he is at Osaka University
mukaigaw@am.sanken.osaka-u.ac.jp

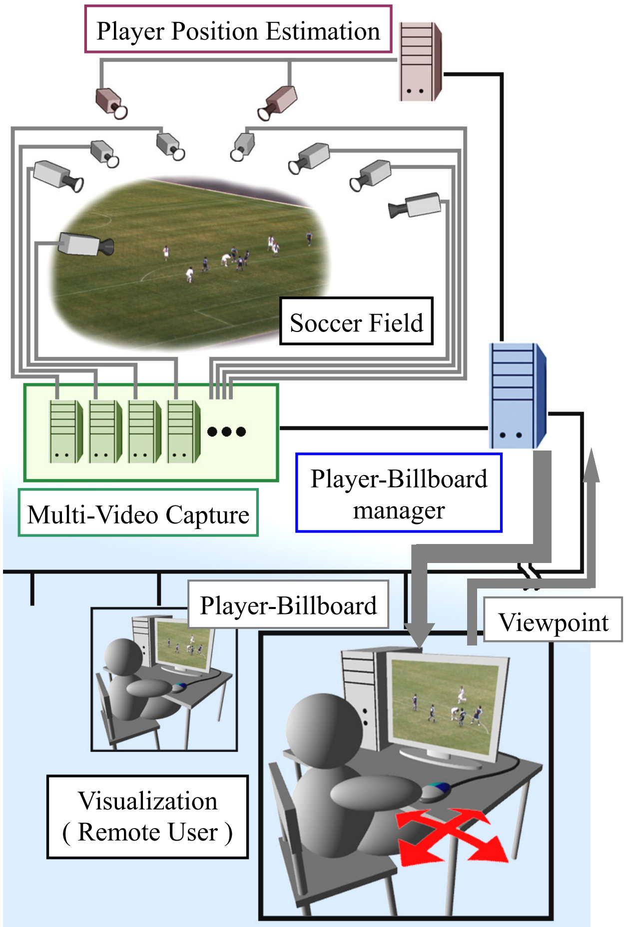

Figure 1. System overview.

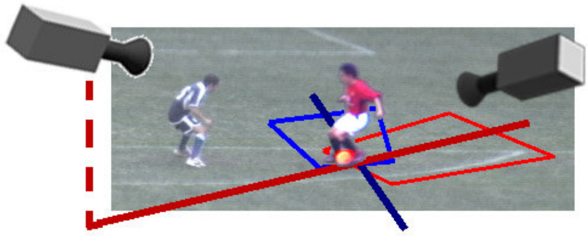

Figure 2. Estimating position of the player.

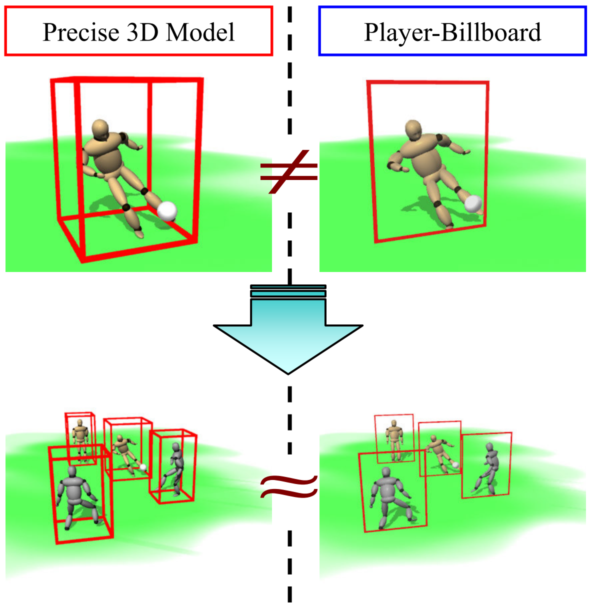

Figure 3. Comparison between close-view and distant-view.

installed at higher place of the stadium. These cameras are angled to cover the soccer field and their optical axes are set to be almost orthogonal each other. 3D positions of the players are estimated by finding overlaps of the projected silhouettes of the players obtained by each camera onto the soccer field. The position of each player is estimated as the center of the overlap (Figure 2). Multi-video capture part extracts the textures of players based on the estimated player positions, and they are sent to player-billboard manager.

Remote users send their viewpoint information to the player-billboard manager via network, and obtain appropriate data set of player-billboards. The users freely control their viewpoint by their mouse operation individually, and they can watch live 3D video from their own viewpoint.

2.2 Player billboard

When users watch 3D video of soccer scene in which many players exist, they sometimes set their viewpoint at some distance away from individual players so as to recognize spatial relationship of the players. As shown in Figure 3, reconstructing precise 3D shape of individual players would be less important as opposed to the importance of reproducing spatial relationship of the players. We cannot distinguish these appearances at distant-view. Therefore, we focus on reproducing 3D positions of the players and their textures.

We represent each player by one rectangle and associated texture. We call the representation "player-billboard". The rectangle is vertically placed at the position of the foot of the player. As shown in Figure 4, the rectangle is always set to face user's viewpoint. The most appropriate texture is selected based on user's viewpoint and it is mapped on the rectangle. Pixels determined as background on the texture are rendered as transparent. Thus, users can recognize spatial relationship of the players naturally.

Because precise 3D shape reconstruction is not required to generate 3D video, the computation cost is reduced drastically. It is one of the advantages of our approach. Moreover, data size to describe 3D video is small because each player is represented simply by one rectangle and one texture. Therefore, this method is considered to be suitable for transmitting 3D video via network.

2.3 Transmitting live 3D video

Transmitting live 3D video means on-line transmission of 3D positions of the players and their textures, and the data to be sent is changed according to the 3D position of user's viewpoint.

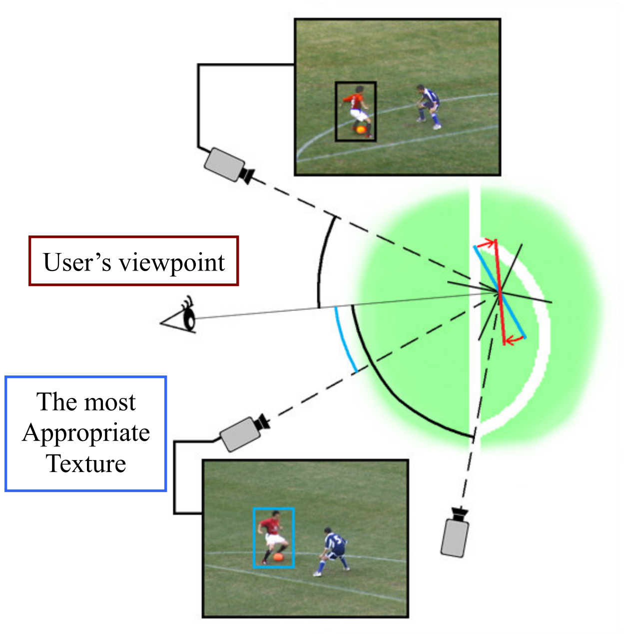

Figure 4. Representation of player-billboard.

A remote user sends the viewpoint at which at he/she wants to watch the soccer scene to the player-billboard manager. The player-billboard manager selects one appropriate texture based on the viewpoint for one player, and returns 3D positions of the player-billboard with their textures to the user. The player-billboard manager does not send background data like soccer stadium at each frame. The background data is sent as 3D model in advance. Then, 3D video is rendered by arranging data of the players transmitted by the player-billboard manager in the 3D stadium model at each frame.

On the user's PC, 3D video visualization process runs concurrently with the process that communicates with the player-billboard manager. As a result, 3D video can be rendered independently from the speed of communication cycle. 3D video at user's PC is refreshed not only when the user's PC receives newer player-billboard data from the manager, but also when the user changes the viewpoint by his/her mouse operation. 3D video is always rendered from the current viewpoint with recently received player-billboard data. Thus, the textures used to render 3D video may not be the most appropriately angled ones for the user's viewpoint until the user receives the appropriate textures for the viewpoint. However, as the user moves his/her viewpoint, all the visualization except for the texture update can be correctly run within a frame. Therefore the speed of 3D video rendering can catch up user's mouse operation within a frame, and it is not affected by jitter or delay of the network.

The player-billboard manager only selects appropriate texture for the current viewpoint of each user. 3D videos are not rendered on player-billboard manager but on the user's PC. Thus, our system can transmit 3D video different viewpoints of individual users simultaneously.

3. Experiments

We conducted experiments at National Kasumigaoka Stadium in Japan to show the performance of our system for generating and transmitting 3D video to remote users in real time.

Figure 5. Snapshots from live 3D video.

3.1 Capturing multiple videos in real soccer stadium

We installed one camera at the top of the roof of the audience seats and the other at the top of the scoreboard in the stadium. These cameras were connected to the PCs of the player position estimation part. In addition, we installed eight cameras at the middle of the audience areas. They surround the soccer field and cover half of the soccer field. These cameras were connected to the PCs of the multi-video capturing part.

We captured real soccer game with these ten cameras and stored them to HDD of PCs. In the following experiments, the stored videos were used as input. The system loaded them in advance onto the memory of each PC. However, all processes to generate 3D video such as estimating positions of the players or obtaining the textures of all the players were performed in real time.

3.2 3D video of soccer game

The 3D video as shown in Figure 5 was generated in video-rate with our system. Although each player was represented by a billboard, the players' movement and their spatial relationships were very well perceived as if they were represented by 3D models.

In the stored videos, about 18 people showed up at each frame on the average. We compressed each texture by JPEG format, and the texture size of each player was about 960 [Byte]. Thus, the total of the transmitted data size was about 17 [KB] per frame, and less than 4

Table 1. Experimental results of transmitting 3D video to remote users.

	SingleSimultaneous			
	ThroughputLatencyThroughputLatency			
ADSL 8Mpbs (5)24 [fps]128 [ms]13229				
CATV 6 (3.5)1521312259				
ADSL 12 (3)2114524124				
CATV 12 (2.5)1718312267				
LAN 100 (30)2910429104				

[Mbps] on transmitting the 3D video in video rate (30 [fps]).

3.3 Transmitting 3D video to remote users

We performed experiments to transmit 3D video to remote users through the Internet. We prepared four types of broadband network as shown in Table 1. We also used a LAN as a reference.

First, single remote user connected to the player-billboard manager via the network, and received the data for rendering 3D video (noted "Single" in Table 1). The result proves that our system can transmit 3D video to a remote user at high frame rate. Moreover, the latency which affects the delay of receiving the most appropriate texture is sufficiently low.

Second, five users connected to the manager and received the data simultaneously (noted "Simultaneous" in Table 1). The result shows that our system can transmit 3D video to five users with little decrease of the throughput. Increase of the latency is also small. We confirmed that our system is capable of transmitting 3D video via various types of network simultaneously.

3.4 Transmitting 3D video to multiple users

We examined the capability of our system to transmit 3D video to users simultaneously. We transmitted 3D video to users via LAN, and measured the throughput and the latency. The result is shown in Figure 6. According to the result, our system is capable of transmitting 3D video to up to fourteen users simultaneously.

According to this result, our system could be easily expanded to transmit 3D video to a larger number of users. In order to deal with more users, we just have to mirror the player-billboard manager and prepare adequate number of mirrored servers. The servers have a copy of dataset of player-billboards for rendering 3D video from the original player-billboard manager. Users connect to one of the mirrored servers, and receive positions and textures of the players. These servers individually select appropriate textures of players for each user, and transmit them so as to visualize the soccer game at user sides in real time.

4. Conclusions

We have presented the live 3D video providing system for large-scale spaces such as a soccer stadium. With our system, all processes from capturing multiple videos to display 3D video at remote users' PC can be done in real time.

We have performed experiments of transmitting 3D videos to remote users via broadband network, and proven the capability to transmit live 3D videos by the presented method. We succeeded in transmitting 3D videos up to fourteen users simultaneously, and the throughput and the latency of transmitted 3D video were kept constant as the number of users was increased.

Currently, we focus on visualizing only players, and we do not track a soccer ball in captured videos. It is necessary to render a soccer ball in 3D video to describe the soccer game fully. Future works also include robust detection of players against illumination changes. Because we capture videos at a soccer stadium, illumination changes due to sun light are quite often and drastic.

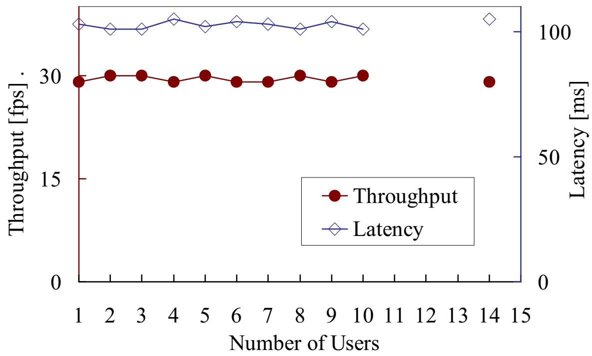

Figure 6. Experimental results of transmitting 3D video to multiple users.

References

[1] T.Kanade, P.W.Rander, and P.J.Narayanan, "Virtualized Reality: Constructing Virtual World from Real Scenes", IEEE Multimedia, Vol.4, No.1, pp.33-47, 1997

[2] S.Würmlin, E.Lamboray, O.G.Staadt, and M.H.Gross, "3D Video Recorder", Proc. of Pacific Graphics '02, pp.325-334, 2002.

[3] W.Matusik, C.Buehler, R.Rasker, S.J.Gortler, and L.McMillan, "Image-Based Visual Hulls", Proc. of ACM SIGGRAPH2000, pp.369-374, 2000

[4] T.Koyama, I.Kitahara, Y.Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium", Proc. of ISMAR2003, pp.178-187, 2003

[5] Y.Kameda, T.Koyama, Y.Mukaigawa, F.Yoshikawa, Y.Ohta, "Free Viewpoint Browsing of Live Soccer Games", Proc. of ICME2004, 2004