3DTV conference, 2007/5/7-9.

A Perceptually Correct 3D Model for Live 3D TV

Yuichi Ohta, Itaru Kitahara, Yoshinari Kameda,
Hiroyuki Ishikawa and Takayoshi Koyama

Department of Intelligent Interaction Technologies
University of Tsukuba

ABSTRACT

3D modeling of a scene is essential for the generation of 3D video from multiple images. In this paper, we present a perceptually correct 3D modeling scheme. "Perceptually correct" means that it may not represent the physically correct object shape but it can generate perceptually almost-correct views of the object. A perceptually correct 3D model is an approximation of the physically correct 3D model, but it can be constructed more easily and stably from images. The quality of 3D video generated using a perceptually correct 3D model could be better than that generated using a physically correct model, when the models are constructed from images. A live 3D TV system for soccer scene has been developed based on the modeling scheme.

Index Terms— 3D TV, 3D video, mixed reality, image-based rendering, 3D model, billboard, live texture

1. INTRODUCTION

3D modeling of a scene is essential for the generation of 3D video from multiple images. Polygon-based 3D shape models used for ordinary computer graphics are often employed as 3D modeling schemes for 3D TV systems. Polygon-based 3D shape models can generate high quality 3D videos as long as it is possible to construct physically correct 3D shape models. In a 3D TV system, however, it is necessary to construct the 3D model from images, and it is very difficult to obtain a physically correct 3D shape model using images. The shapes and the surface textures of polygon-based 3D models constructed from images are often considerably degraded and therefore so is the quality of the resulting 3D videos.

In 1996, Satoh reported pioneering work on a 3D image display system [1]. In the sensing system, a scene is observed by using a camera matrix. A high quality 3D model, such as a stage setting, was constructed by using an occlusion detectable stereo algorithm. In the display system, arbitrary 2D views of the scene are presented in real time on the screen following an observer's head position sensed by a magnetic tracker, as shown in Figure 1. This system can thus present realistic motion parallax.

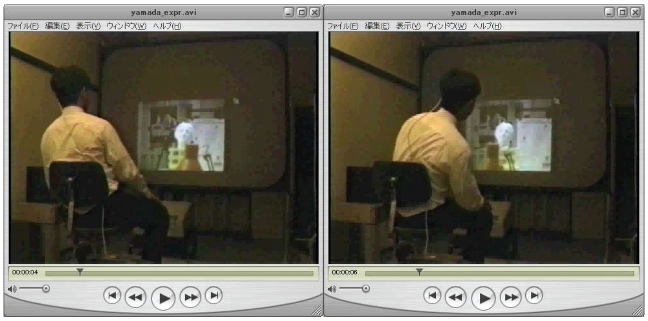
yamada_expr.avi
ファイル(F) 編集(E) 表示(V) ウィンドウ(W) ヘルプ(H)
00:00:04
yamada_expr.avi
ファイル(F) 編集(E) 表示(V) ウィンドウ(W) ヘルプ(H)
00:00:06

Figure 1. 3D image display system with motion parallax[1]. The image projected on the screen changes in real time following the head motion of the observer generating motion parallax. A 3D model similar to a stage setting was constructed from polynocular stereo images.

The quality of images generated from the 3D model was almost equivalent to that of the original stereo images. However, the observer cannot look around the objects because the 3D model is like a stage setting, which is not a physically correct 3D modeling scheme, but is a perceptually correct 3D modeling scheme as long as the viewing position of the observer is limited to in front of the objects.

Here, we introduce a perceptually correct modeling scheme suitable for dynamic sports scenes such as soccer games. Each player is represented by a rectangular 3D plane and a live texture. We call this the "player billboard." The player billboard does not represent the precise 3D shape of the object. It represents the 3D location and the 2D appearance of the object. However, by rotating the orientation of the plane and by switching the dynamic texture following the observer's viewing angle, the player billboard scheme can present a live 3D video with good quality. An observer can select an arbitrary viewing angle, even the player's-eye view, and can feel motion parallax.

2. PLAYER BILLBOARD

The player billboard is similar to the billboard technique. A vertical rectangular plane located at the position of each 3D object represents the 3D shape of each object, as illustrated in Figure 2. A video texture selected from multiple videos is mapped onto the plane. In this scheme, the shape of the 3D object is approximated by a plane and 3D shape reconstruction is not necessary. Thus, the processing cost to generate a 3D model by merging multiple videos can be reduced markedly. At the same time, mapping a real video as a texture on the plane produces a realistic object representation.

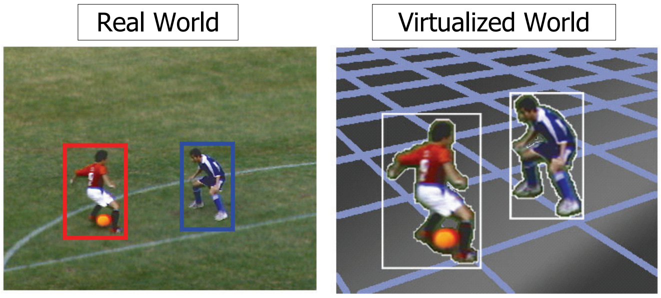

Figure 2. Player billboard [2].
The player billboard is similar to the billboard technique. A vertical rectangular plane located at the position of the 3D object represents the 3D shape of each object.

Each 3D object is approximated by a single plane. Therefore, the plane must be set up with an appropriate orientation. If the orientation is fixed and the observer moves around towards the side of the plane, the visible surface area would decrease. As illustrated in Figure 3, the surface normal of the plane is always aligned with the observer's line of sight, following the billboard technique. The texture mapped on the plane is selected from textures captured by multiple cameras surrounding the object so that the viewing direction of the camera is closest to that of the observer.

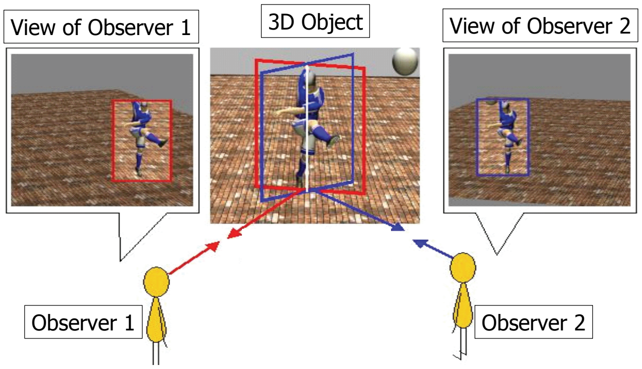

Figure 3. Orientation of billboard.
The surface normal of the plane is aligned with the observer's line of sight.

3. REPRESENTATION OF 3D OBJECTS BY PLAYER BILLBOARD

A player billboard can precisely represent the 3D location of an object. A set of player billboards located on the virtual playing field can represent the mutual occlusion between objects. The motion parallax caused by the movement of observer's viewing position is represented by the relative locations of multiple billboards. The player billboard is a physically correct 3D modeling scheme from the above point of view.

However, each 3D object is represented by a single plane. The changes in appearance of each object caused by changes in the observer's viewing position are represented by adjusting the orientation of the billboard and by switching the video texture from multiple cameras surrounding the object. This raises the question; "Can the 3D video generated by using player billboards give 3D perception to an observer?" The answer is 'yes,' and this is the main issue in this paper. The player billboard is a perceptually correct 3D modeling scheme for a moving object, such as a sports player.

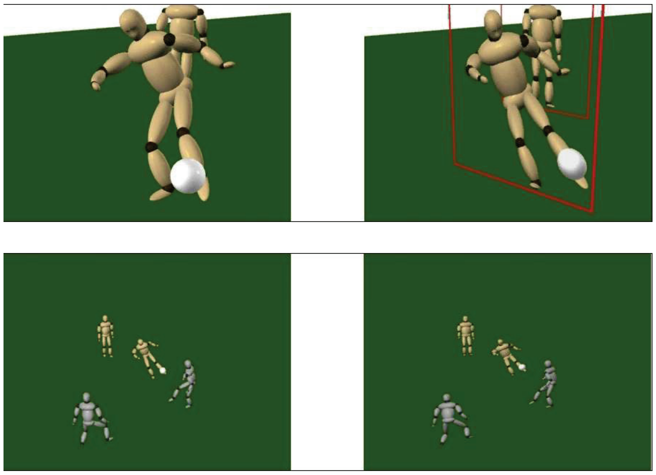

Figure 4. Comparison of 3D videos generated by CG.
The video on the left was generated using a 3D CG model. The video on the right was generated using player billboards with static 2D texture. In the video, the appearance of the scene changes by altering the location of the virtual camera.

Figure 4 shows snapshots of four videos synthesized by computer graphics. The scene is static but the virtual camera moves. The two on the left were rendered using ordinary 3D polygon models, while those on the right were rendered by simulating the player billboard scheme. The upper two close-up views were rendered by identical movement of the virtual camera, as were the lower two long-shot views. Some people cannot see any difference between the upper two videos when presented side-by-side. If we carefully examine the mutual positions of the body and arm, for example, it is not difficult to detect the difference between the two. In the video on the left, the mutual positional relation of body and arm changes as a result of motion parallax. In the video on the right, no motion parallax is presented within a single billboard. In both cases, the motion parallax between the two players, near and far, is presented correctly. It seems that our eyes do not have sufficient acuity to detect subtle changes in appearance within a single 3D object caused by motion parallax. It is almost impossible to detect the difference between the lower two videos when presented side-by-side,

The scene in Figure 4 is static. To examine the ability of the player billboard in a more realistic scene, we built a virtual 3D field and a running virtual soccer player with a 3D CG modeling tool (3D Studio MAX). The motion data of the running player were obtained from a real player using a motion capture technique. Once we have a 3D CG space, we can render images from arbitrary views and can evaluate the correctness of the 3D video generated by the player billboard scheme.

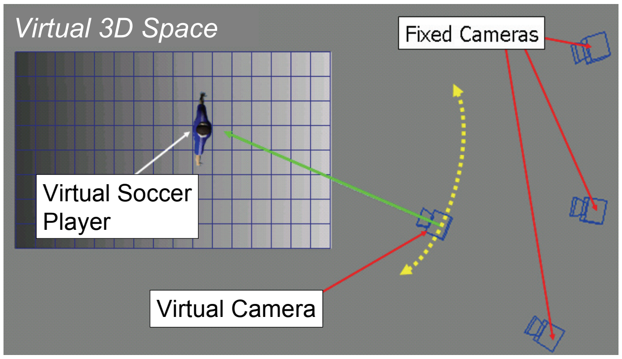

Figure 5. Layout of multiple virtual fixed cameras and a swinging virtual camera in 3D CG space where a soccer player is running.

The videos rendered at the fixed cameras are used as dynamic textures to render a video at the swinging virtual camera.

As illustrated in Figure 5, we set up 3 fixed virtual cameras in the 3D CG space where a virtual soccer player was running. The convergence angle between two adjacent cameras is about 22.5°. Another virtual camera swings between the right and left fixed cameras. We rendered two types of video from the viewpoint of the swinging virtual camera, (1) with direct rendering from the 3D CG model, and (2) with the player billboard using the dynamic textures obtained from the 3 fixed cameras, as shown in Figure 6. In case (2), the texture of the player is switched during swinging of the virtual camera. The two videos were presented to the subjects. Interestingly, none of the subjects perceived the texture switch in the video in case (2). If the player is static in the video, they may perceive texture switching. When the player is moving and the texture is changing dynamically, it is difficult to perceive the texture switch. This illusory perception of flickering is of interest, and there have been some previous studies of this phenomenon in human vision [3]. With this human perception of flickering, it is feasible that our scheme can generate 3D videos in which no human observers would perceive the texture switch caused by the simplified 3D model representation, *i.e.,* player billboard.

The player billboard 3D modeling scheme represents the 3D locations of objects correctly and the motion parallax between multiple objects is represented in a physically correct manner. On the other hand, it does not represent the 3D structure in a single 3D object, and the change in appearance caused by the motion parallax is represented by switching dynamic textures obtained from multiple cameras surrounding the object. As a result of this simplification, the 3D video cannot represent the subtle changes in appearance caused by the 3D structure of the object. It seems that, however, the acuity of the human eye is insufficient to perceive these subtle changes. The player billboard also causes an undesirable side effect, texture switching. The texture switching, however, is not perceptible to a human observer when the texture is changing dynamically. These considerations support our suggestion that the player billboard is an effective perceptually correct 3D modeling scheme.

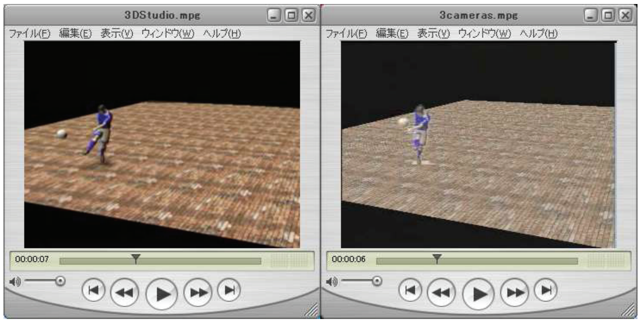
3DStudio.mpg
ファイル(F) 編集(E) 表示(V) ウィンドウ(W) ヘルプ(H)
00:00:07
3cameras.mpg
ファイル(F) 編集(E) 表示(V) ウィンドウ(W) ヘルプ(H)
00:00:06

Figure 6. Snapshots of 3D videos rendered from the swinging virtual camera in Figure 5.

The video on the left was rendered directly from 3D CG model. The video on the right was rendered by the player billboard. The texture switching in the right video is not perceptible.

4. A PROTOTYPE LIVE 3D TV SYSTEM

In this section, we describe a prototype live 3D TV system. Figure 7 illustrates an overview of the system. It should be noted that in this system all processes are computed in real time. The system consists of a capture unit, a server PC, and rendering PCs. The capture unit consists of multiple cameras, a scene analysis PC, and texture PCs. The scene analysis PC extracts all players on the pitch using two cameras observing the pitch from elevated positions. The texture PCs receive information regarding the number of players and their estimated 3D positions *via* the server PC. In the texture PC, the texture region to be mapped on the billboard is extracted using the 3D positions of the players. The texture data of all players are sent to the server PC, which generates a 3D data stream for every player. The rendering PC obtains the viewing position and orientation of the observer's view from a human interface device, which is then sent to the server PC as virtual camera parameters. The rendering PC receives the data stream necessary for synthesis of 3D video from the server PC. Only a single texture is sent for each player billboard. The most appropriate texture is selected depending on the observer's viewpoint. A 3D CG model of the stadium (background object of 3D scene), which includes the entire stadium, pitch, and goals, is supplied to the observer's PC separately before the dynamic event starts.

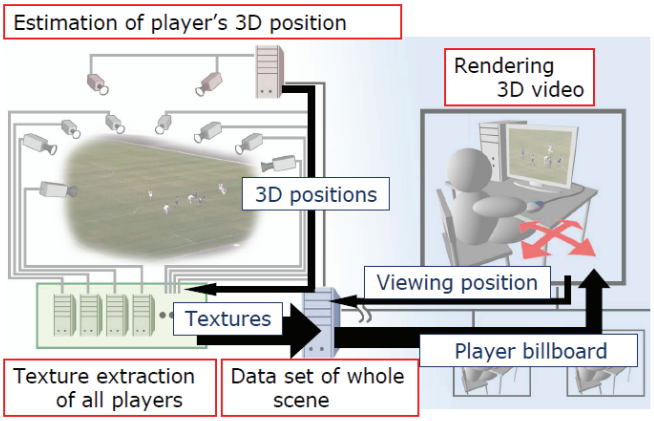

Figure 7. Overview of our live 3D TV system.

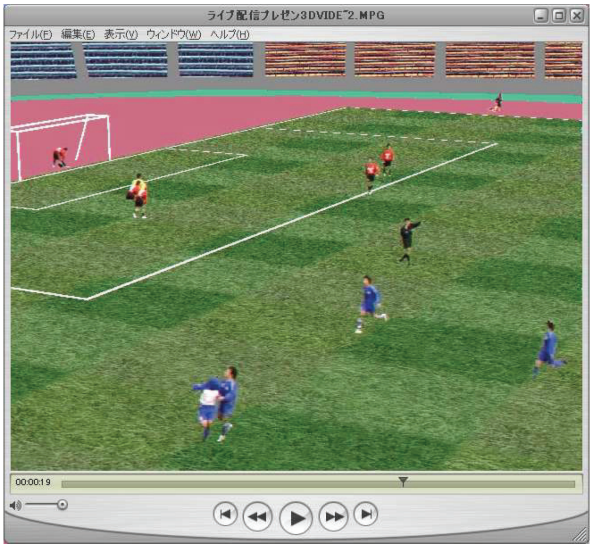
ライブ配信プレゼン3DVIDE~2.MPG
ファイル(F) 編集(E) 表示(V) ウィンドウ(W) ヘルプ(H)
00:00:19

Figure 8. A snapshot from a 3D video of a real soccer game at the National Stadium, Tokyo, Japan.

Finally, the 3D video is synthesized based on the viewing position specified by the observer at the rendering PC. We note that the observer can change his viewpoint smoothly any time during the 3D video presentation.

Figure 8 shows a snapshot of the 3D video of a real soccer game recorded at the National Stadium, Tokyo, Japan. This figure demonstrates the effectiveness of the player billboard modeling scheme for sports scenes.

5. CONCLUSIONS

In this paper, we proposed a concept of perceptually correct 3D modeling scheme and presented the player billboard scheme as an example. The results of experiments using 3D videos generated by computer graphics strongly support the suggestion that the player billboard scheme is perceptually correct. We are planning to perform more precise and systematic experiments to confirm the validity as well as to clarify the limitations of the scheme.

A prototype live 3D TV system based on the player billboard scheme was introduced. This system can capture video data from multiple cameras, reconstruct 3D models, transmit 3D video streams *via* the network, and display them on remote PCs. All processes are done in real time. The simplified 3D modeling scheme, player billboard, makes this live 3D video system possible. We feel that the quality of 3D video images generated by the player billboard will be better than those by other 3D volume-oriented modeling schemes, because the player billboard scheme requires much simpler image processing than others. This issue will be examined in future studies as well.

ACKNOWLEDGMENTS

This research was supported in part by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications. We thank all the members of our research group, including Y.Mukaigawa at Osaka University, and H. Yoshikawa and K.Shirai at the Japan Institute of Sports Sciences, for cooperation in the experiment at the National Kasumigaoka Stadium, Tokyo, Japan.

REFERENCES

[1] K. Satoh, I. Kitahara, and Y. Ohta, "3D Image Display with Motion Parallax by Camera Matrix Stereo," *Proc. IEEE MULTIMEDIA '96*, pp. 349-357, 1996.

[2] T. Koyama, I. Kitahara, and Y. Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium," *Proc. Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR2003)*, pp.178-187, 2003.

[3] T. Takeuchi, and K.D. Valois, "Motion Sharpening in Moving Natural Images," *Journal of Vision,* Vol.2 (7), pp.377, 2002.