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

A BACKGROUND MODELING METHOD WITH SIMPLE OPERATIONS FOR 3D VIDEO

Naobumi Nomura, Itaru Kitahara, Yoshinari Kameda, and Yuichi Ohta

Graduate School of Systems and Information Engineering, University of Tsukuba

ABSTRACT

This paper proposes a novel 3D background modeling method that only needs simple manual operations. For generating a background 3D model, the background area of a 3D scene is divided into a “ground region” and a “distant region.” The 3D model of the distant region is represented by vertical rectangular planes with video texture. The ground region is modeled as a simple horizontal plane. These two regions are specified by a few points given manually in each image of multiple cameras. After the points are given by an operator, operation of the 3D background modeling is conducted based on the knowledge of camera poses and the target scene. We generated 3D videos of a background by using multiple 2D videos recorded at the Japan National Kasumigaoka Stadium and the Gymnasium of the Japan Institute of Sports Science to show the validity of our method.

***Index Terms*—** 3D Video, 3D Modeling, Image-based Rendering, Texture Mapping, Sports Scene

1. INTRODUCTION

With recent progress in computer and video technologies, many computer vision-based 3D video systems have been developed [1]-[3]. We also developed a 3D video system that realizes live 3D telecasts of soccer games [4]. These researches, however, tend to concentrate on the generation of the realistic appearance of dynamically changing targets (i.e., such foreground objects as soccer players) more than scene background. In our system [4], for example, the appearance of background is generated by using a 3D CAD model of a soccer stadium with texture mapping. 3D CAD models are often available for modern large-scale structures, such as national stadiums. However, obtaining a CAD model of a small gymnasium in a town is difficult. It is impractical to make a CAD model from scratch simply for generating 3D videos because the human costs are high. 3D background modeling is important because the appearance of background affects the quality of the 3D video. This paper proposes a novel 3D background modeling method that only needs simple manual operations. Our method enables the generation of high quality 3D videos with realistic background of every athletic facility, even if a 3D CAD model is unavailable.

2. 3D BACKGROUND MODELING METHOD

2.1. 3D Shape Approximation of Background

Our method models background by using multiple images captured by fixed and calibrated cameras in an athletic facility [5]. While watching sporting events, the audience mainly looks at foreground objects (e.g., soccer players) but not at the background. Moreover, such a distant region as an auditorium is located much farther from the audience’s viewpoint than foreground objects. Therefore, it is not necessary to correctly represent motion parallax for the background because the 3D video allows the audience’s viewpoint to move around in 3D space. We focus on this point and assume the following three conditions to model a background:

1. A background can be decomposed into ground and distant regions.
2. The ground region should be a horizontal single plane on which all foreground objects are placed.
3. The distant region can be regarded as a set of vertical planes [6].

As shown in Figure 1, by approximating the 3D shape of background as a set of vertical planes, our method can model background by manually indicating several positions on each image captured by multiple cameras.

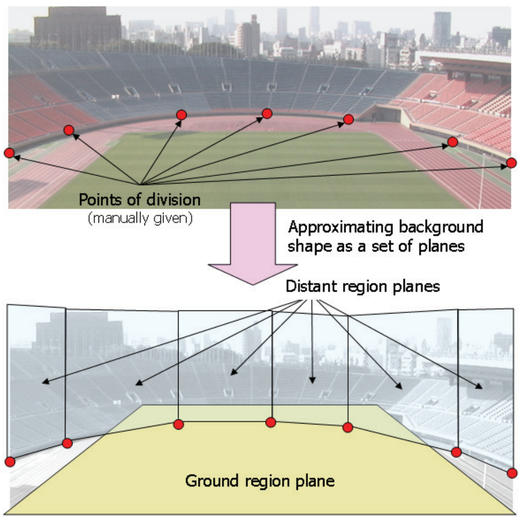

Figure 1: 3D shape approximation of background

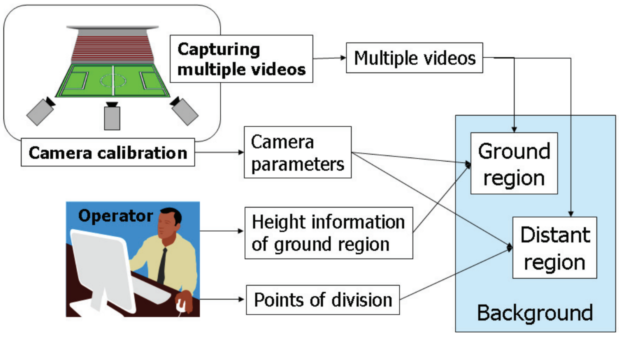

Figure 2: Overview of our method

2.2. Overview of Modeling Method

As illustrated in Figure 2, the input information required for modeling a background is multiple videos for texture mapping, camera parameters of the cameras, the height of the ground region in a 3D scene, and a set of 2D coordinates of the "points of division" specified on each multiple video. Points of division exist on the borderline between the ground and distant regions. By projecting a 2D point onto the ground region (3D plane), one of the vertices on the bottom edge of a vertical plane of the distant region is determined. Since we assume that a distant-region plane is perpendicular to the ground region, it is possible to define a distant-region plane by specifying only two points. When a distant region has a complicated shape, the method can adapt it by increasing the number of points. Finally, multiple videos are mapped onto the 3D planes to render 3D videos.

3. GROUND REGION MODELING

In this section, we describe the ground region modeling process. Suppose *u(u,v)* is a 2D point in an image plane and *X(x,y,z)* represents a 3D point in the captured 3D space. Let us denote their homogeneous coordinates as $\tilde{u}$ *(u,v,1)* and $\tilde{X}$ *(x,y,z,1)*, respectively. The three-dimensional projective transformation projects a 3D point onto a 2D point in the image and can be described as:

$$\lambda\begin{pmatrix}u\\v\\1\end{pmatrix}=\begin{pmatrix}C_{11}&C_{12}&C_{13}&C_{14}\\C_{21}&C_{22}&C_{23}&C_{24}\\C_{31}&C_{32}&C_{33}&C_{34}\end{pmatrix}\begin{pmatrix}x\\y\\z\\1\end{pmatrix}, \quad (1)$$

where λ is a scale factor and C_{3x4} is the projective matrix of the capturing camera determined by the camera parameters given in advance. To simplify the modeling process, assume that the orientation of the z-axis is vertical. When the height of the 3D plane is given as a constant value *h*, Eq. (1) can be converted as

$$\lambda\begin{pmatrix}u\\v\\1\end{pmatrix}=\begin{pmatrix}C_{11}x+C_{12}y+C_{13}h+C_{14}\\C_{21}x+C_{22}y+C_{23}h+C_{24}\\C_{31}x+C_{32}y+C_{33}h+C_{34}\end{pmatrix}. \quad (2)$$

Then two-dimensional projective transformation is derived by reorganizing Eq. (2):

$$\lambda\begin{pmatrix}u\\v\\1\end{pmatrix}=\begin{pmatrix}C_{11}&C_{12}&hC_{13}+C_{14}\\C_{21}&C_{22}&hC_{23}+C_{24}\\C_{31}&C_{32}&hC_{33}+C_{34}\end{pmatrix}\begin{pmatrix}x\\y\\1\end{pmatrix}=B\begin{pmatrix}x\\y\\1\end{pmatrix}. \quad (3)$$

The two-dimensional projective matrix *B* apparently has inverse matrix B^{-1}. This inverse matrix can generate 3D plane texture by projecting a 2D point *(u,v)* in the captured image onto a point on the 3D plane *(x,y,h)* [7][8]. Finally, a 3D model of the ground region with shape and texture information is generated.

4. DISTANT REGION MODELING

In this section, we describe the distant region modeling process (see Figure 3). First, the coordinates of the two points of division, $p=(u_p,v_p)$ and $q=(u_q,v_q)$, which exist on the border between the ground and distant regions, are given to the modeling process. By projecting *p* and *q* onto a ground plane with Eq. (3), the 3D coordinates of the vertices of the bottom edge of distant region $P=(x_P,y_P,h)$ and $Q=(x_Q,y_Q,h)$ are calculated. When the distant region is observed in the captured image as an area with *m* pixels width and *n* pixels height, the 3D coordinates of the vertices of the upper edge of the distant region, $P'=(x_P',y_P',z_P')$ and $Q'=(x_Q',y_Q',z_Q')$, are calculated with a height factor α =n/m:

$$P'=(x_P',y_P',z_P')=(x_P,y_P,h+d\alpha)$$
$$Q'=(x_Q',y_Q',z_Q')=(x_Q,y_Q,h+d\alpha) \quad (4)$$

where *d* is the Euclidean distance between *P* and *Q*. In this way, a vertical 3D plane is reconstructed that is part of the distant region.

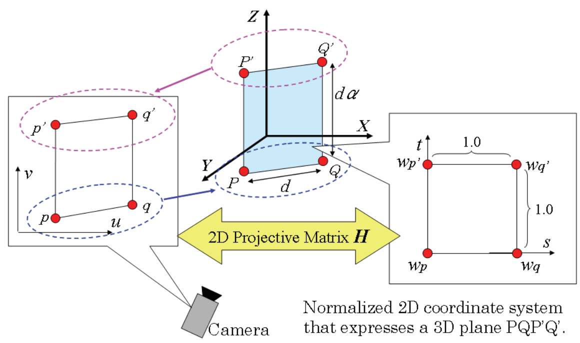

Figure 3: Distant region modeling

We define a new normalized 2D coordinate system that expresses a point *w(s, t)* on a distant region plane. The origin of this system corresponds to *P*, the *s*-axis is parallel to *PQ* and normalized by d=||*PQ*||, and the *t*-axis is parallel to *PP'* and normalized by dα. In this coordinate system, *P, Q, P', Q'* are described as w_p (0, 0), w_q(1, 0), $w_{p'}$(0, α), and $w_{q'}$(1, α), respectively.

P, Q, P', Q' are observed at *p, q, p', q'* in the captured image, respectively. The 2D coordinates of the two points, *p'* and *q'*, are calculated by substituting *P', Q'* into Eq. (1). By using these four corresponding pairs, two-dimensional projective matrix *H* between the image and distant region planes is obtained:

$$\lambda'\tilde{w} = H\tilde{u}\,, \qquad (5)$$

where λ' is a scale factor and $\tilde{w}$ and $\tilde{u}$ are homogeneous coordinates of *w* and *u*.

Two-dimensional projective matrix *H* adds texture to the distant region plane by projecting a 2D point *(u,v)* in the captured image onto a point on the 2D plane *(s,t)*. Finally, a 3D model of the distant region with shape and texture information is generated.

When the distant region's 3D shape is more complicated, it is possible to generate suitable 3D shape by repeating the same process as necessary. In Figure 4, we divided the distant region into six vertical planes. The red points in the upper figure express the points of division. The lower figure shows an example of the generated 3D video of the distant region.

In sport scenes, the appearance of background is sometimes dynamically changed due to audience cheering and movements. We can update texture information of the distant region on line to visualize the change of appearance.

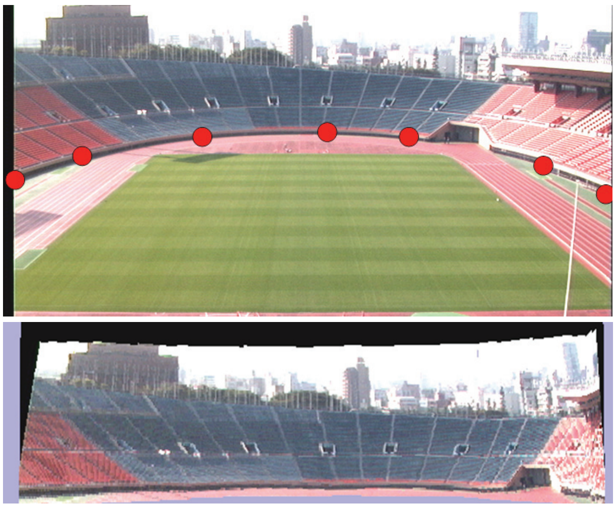

Figure 4: Dividing a distant region into several planes for 3D modeling

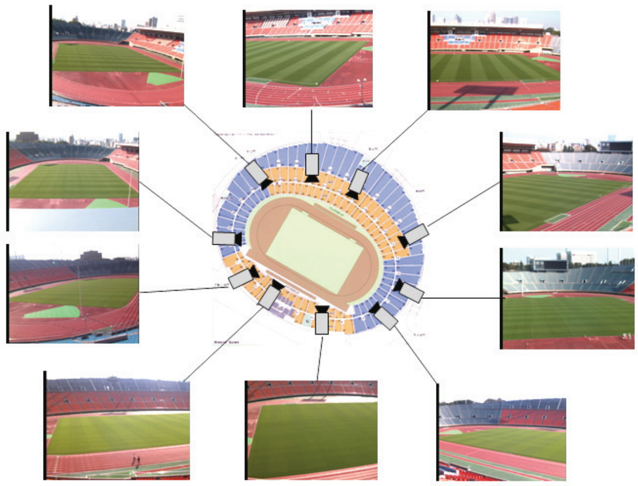

Figure 5: Layout of cameras at Japan National Stadium.

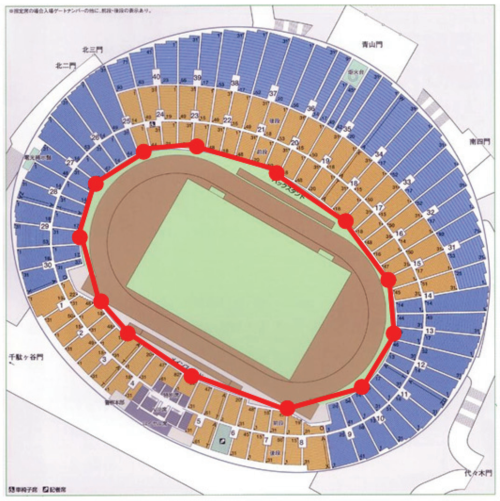

Figure 6: Layout of 13 points of division for the stadium

5. EXPERIMENT

As shown in Figure 5, we conducted an experiment by capturing multiple videos with ten CCD video cameras (SONY, DXC9000) at the Japan National Stadium. Two of the cameras are for tracking soccer players. The other eight cameras are used for obtaining player textures and modeling the background.

As shown in Figure 6, we set 13 points of division to model the distant region by 13 vertical planes. Since it only takes 10 minutes to manually set the points of division, our method requires little effort for background modeling. Although, the dividing number is defined by the complexity of the background shape, our modeling method can optimally define the number for every background by utilizing the adjusting ability of a human operator. Figure 7 shows three snapshots of 3D video generated from the proposed background model. It is possible to generate the appearance of a background stadium from an arbitrary viewing position.

We conducted a similar experiment in the Japan Institute of Sports Science Gymnasium. Figure 8 shows an example of generated 3D video. In this case, the texture of the distant region is updated every frame to express dynamic appearance background changes. You can see the motion of people walking along the wall.

6. CONCLUSION

In this paper, we proposed a novel 3D background modeling method that only needs simple manual operations. The background area is divided into "ground region" and "distant region." The distant region's 3D model is represented by a set of vertical rectangular planes with texture. The ground region is modeled as a simple horizontal plane. These two regions are segmented by a few points given by manual operation. To show the validity of our method, we generated 3D background video by using multiple videos recorded in real large-scale environments.

In the future, color matching and geometric calibration processes need to be improved for better 3D video quality.

REFERENCES

[1] F. Caillette and T. Howard, "Real-time Markerless Human Body Tracking with Multi-view 3-D Voxel Reconstruction," Proc. SIGGRAPH, pp. 43-54, 1996.

[2] S. Yamazaki, R. Sagawa, H. Kawasaki, K. Ikeuchi, and M. Sakauchi, "Microfacet Billboarding," Proc. the 13th Eurographics Workshop on Rendering, pp.175-186, 2002.

[3] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes," IEEE Multimedia, Vol. 4, No.1, pp. 34-37, 1997.

[4] T. Koyama, I. Kitahara, and Y. Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium," Proc. of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR2003), pp.167-178, 2003.

[5] R. Hartley and A. Zisserman, "Multiple View Geometry," Cambridge Univ. Press, 2004.

[6] N. Inamoto and H. Saito, "Free Viewpoint Video Synthesis and Presentation of Sporting Events for Mixed Reality Entertainment," International Conference on Advances in Computer Entertainment Technology (ACE 2004), pp. 42-50, 2004.

[7] R. Szeliski, "Video Mosaics for Virtual Environments", IEEE Computer Graphics and Applications(c), Vol. 16, No. 2, pp. 22-30, 1996.

[8] NaviView: Bird's-Eye View for Highway Drivers Using Roadside Cameras, IEEE International Conference on Multimedia Computing and Systems, Vol. 2, pp. 559-565. 1999.

ACKNOWLEDGEMENT

This research was partially supported by the program of Strategic Information and Communications R&D Promotion Program (SCOPE) under the support of the Ministry of Internal Affairs and Communication, Japan.

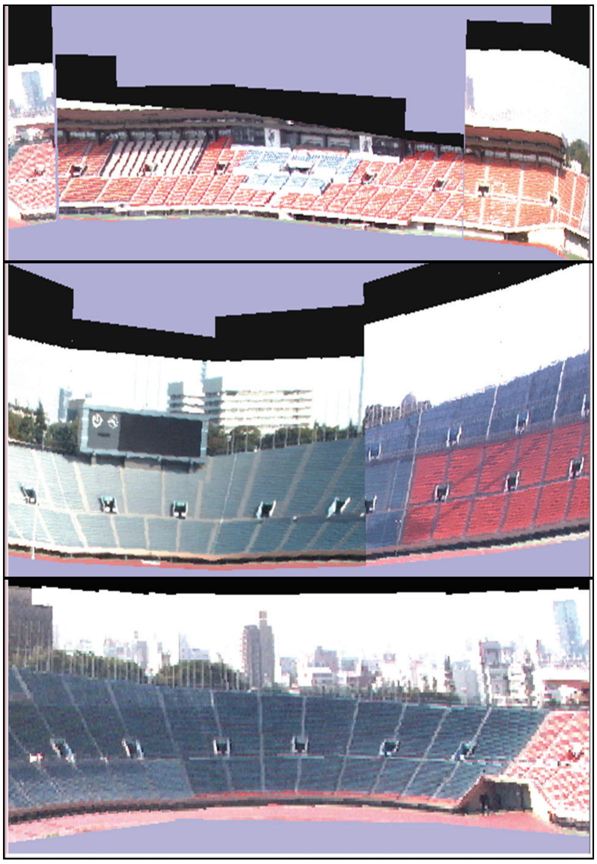

Figure 7: Examples of generated 3D video

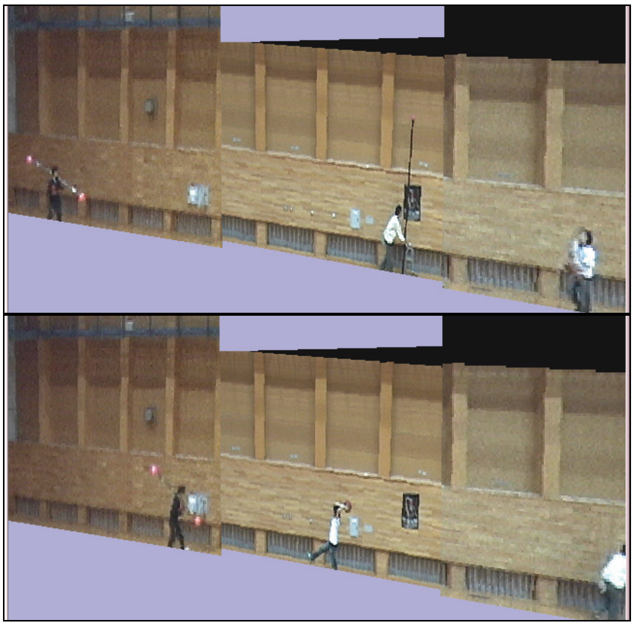

Figure 8: Generated 3D video: people walking along the wall