International Conference on Multimedia & Expo (ICME2007).
2007/07/02-05, Beijing, China.

VIEWPOINT-DEPENDENT QUALITY CONTROL ON MICROFACET BILLBOARDING MODEL FOR SPORTS VIDEO

Hitoshi Furuya, Itaru Kitahara, Yoshinari Kameda, and Yuichi Ohta

Graduate School of Systems and Information Engineering, University of Tsukuba, Japan.
{kitahara,kameda,ohta}@iit.tsukuba.ac.jp

ABSTRACT

We propose a new on-line modeling and rendering method for visualizing players in sports. This method targets the intricate geometry of players in sports such as wrestling. Our system can maintain modeling and rendering quality on producing free-viewpoint 3D video of the players by utilizing the virtual viewpoint information of the viewer. Viewers can move a virtual camera freely over the scene while the system manages the processing cost by changing the size of the voxels and microfacets used for spatial modeling of the players. The system first estimates the rough 3D shape of the players in voxel format, and then assigns a microfacet to each voxel. The texture of the microfacet is obtained from the video cameras surrounding the players. Our preliminary system was evaluated on the wrestling data taken in the Yoyogi National Gymnasium, Tokyo, Japan. We also conducted subjective evaluation of the free-viewpoint video thus produced and found a good balance of rendering quality and high frame-rate performance.

1. INTRODUCTION

Free-viewpoint video [1] for sports is a new method of multimedia entertainment. Viewers can control the virtual camera to allow them to view sports events from their preferred positions, at which a real camera sometimes cannot be set.

On producing free-viewpoint video of sports, care is required to maintain quality as well as achieving a high frame-rate because viewers may wish to see the subtle but fast motion of the players.

Image-based rendering (IBR) is the most promising approach to realize the best rendering quality of the players in a scene [2]. However, the virtual viewpoint of the viewers should be located at the interpolated position of the real cameras and the camera placement should be very dense.

On the other hand, image-based modeling and rendering (IBMR) is a good means to realize arbitrary camera control on free-viewpoint video.

Various IBMR methods have been proposed to generate free-viewpoint video. Pioneering work involved reconstruction of the 3D shape of the objects in a scene utilizing multi-baseline stereo, visual hull, voxel carving, *etc.* [3][4]. Recent studies have achieved real-time modeling and rendering of objects [5][6]. However, in these approaches the reconstructed shape may involve some artifacts, which result in unsophisticated texture rendering because no colors can be matched to the unreal shape.

We have developed a free-viewpoint video system for soccer games and proposed a billboarding-based method based on the advice of processional sport analysts [7][8][9]. In these studies, we found that false shape modeling (including artifacts) of the players should be avoided to fulfill high-quality free-viewpoint video in sports. Real-time performance of processing high frame rates is also required.

Therefore, we propose a new modeling and rendering method that can fulfill these two requirements simultaneously. We adopted the microfacet billboarding method as the base technique because it can cope with intricate geometry of objects without producing false shapes. We have introduced a model quality control method over the conservative microfacet billboarding approach.

The rest of the paper is organized as follows. Section 2 describes the image quality control method on executing the microfacet billboarding method. We explain our preliminary system in section 3 and present the results in section 4. Concluding remarks are presented in section 5.

2. QUALITY CONTROL IN MICROFACET BILLBOARDING

Microfacet billboarding [10] utilizes a set of very small facets to render the intricate geometry of objects. Each facet is placed to express a small portion of the objects, and is aligned perpendicular to the viewing direction.

There are 3 processing steps to render players in sports for free-viewpoint video based on the microfacet billboarding method:

1. Voxel size determination (image quality control)
2. Voxel carving (rough shape recovery of the players)
3. Microfacet setup and texture mapping

As we apply a conservative voxel carving method in step 2, we explain steps 1 and 3 below.

2.1. Voxel size determination

The image quality that viewers perceive depends on the model precision on the image screen. Therefore, insofar as the microfacet size is apparently the same on the screen, the image quality can be said to be maintained to the same level. This means that the system can change the voxel size (which corresponds to the microfacet size in the real world) if only it retains the same size on the screen. This can reduce the calculation cost markedly when the virtual viewpoint is far from the players, as the size of the voxel can be increased. The estimated calculation cost of shape recovery in the second step is $O(cpk^3)$, where c is the number of cameras, p is the projection calculation cost, and k is the number of voxels. If we can decrease the voxel size by half, the cost will be reduced to 1/8.

This approach is effective because viewers sometimes move their viewpoint very far from the players on watching sports games. In Fig. 1, the green frames in the produced image indicate microfacets.

As the overall image quality of the free-viewpoint video is highly dependent on the size of the facet on the image screen, we conducted a subject evaluation experiment with regard to facet size. Ten subjects viewed 5 free-viewpoint video clips of wrestling changing the facet size from 5 to 30 pixels (Fig. 2). The video size was set to 640×480 pixels.

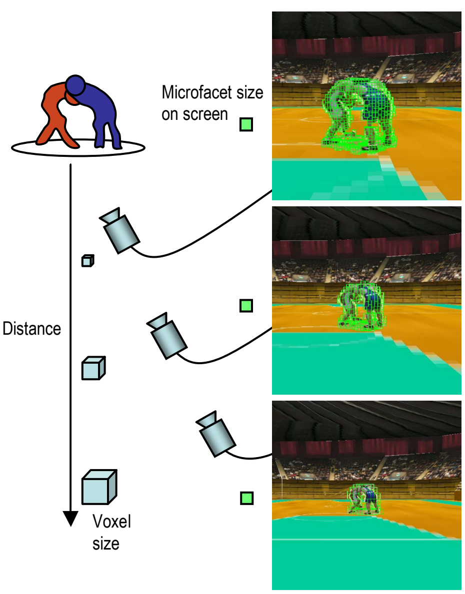

Fig. 1. Microfacet size and voxel size in accordance with the distance of virtual viewpoint.

The score was selected from 1 (worst) to 5 (best). As shown in Fig. 3, the size should be less than 10 pixels, with smaller size yielding better results.

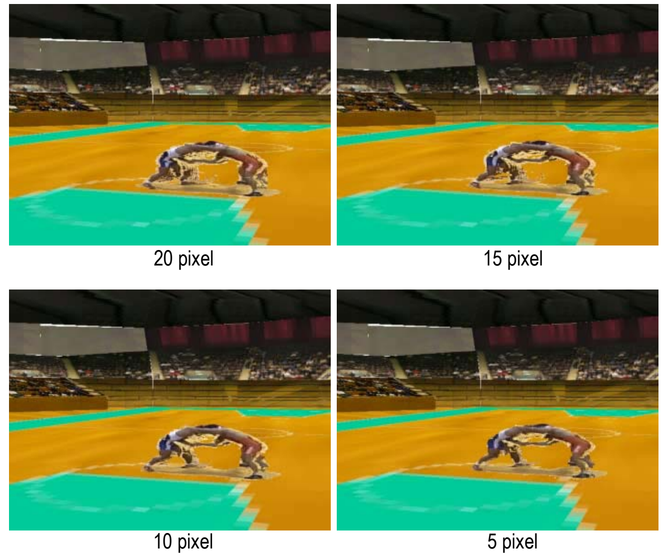

Fig. 2. Sample videos for evaluating different microfacet sizes.

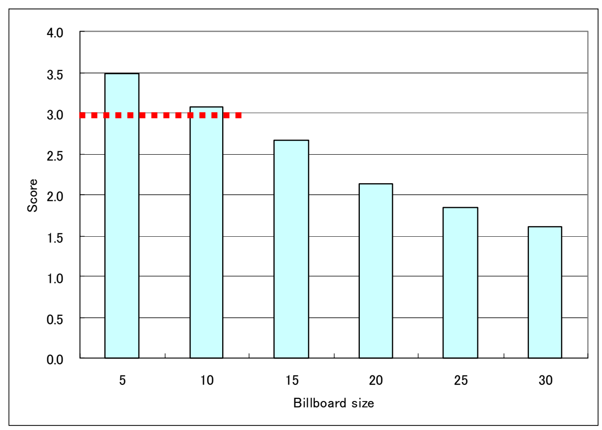

Fig. 3. Quality evaluation by billboard size.

2.2. Microfacet setup and texture mapping

Once the voxel representation is obtained, facets are set up at voxel locations. The texture of the facet is taken from the camera nearest the virtual viewpoint (Fig. 4). However, if the nearest camera is occluded by other voxels, a false texture will be mapped onto the facet. Although a number of methods (such as that described in [11]) have been proposed to handle this problem, our method simply switches to the next available camera to prevent high computation cost.

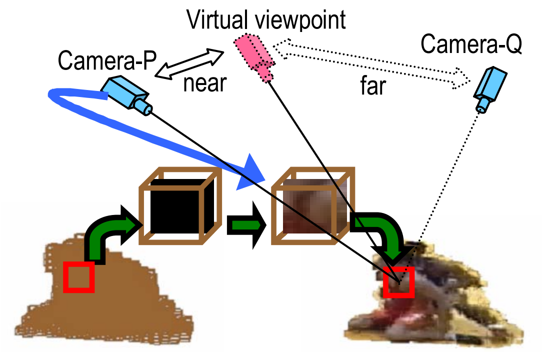

Fig. 4. Texture mapping from the nearest camera image.

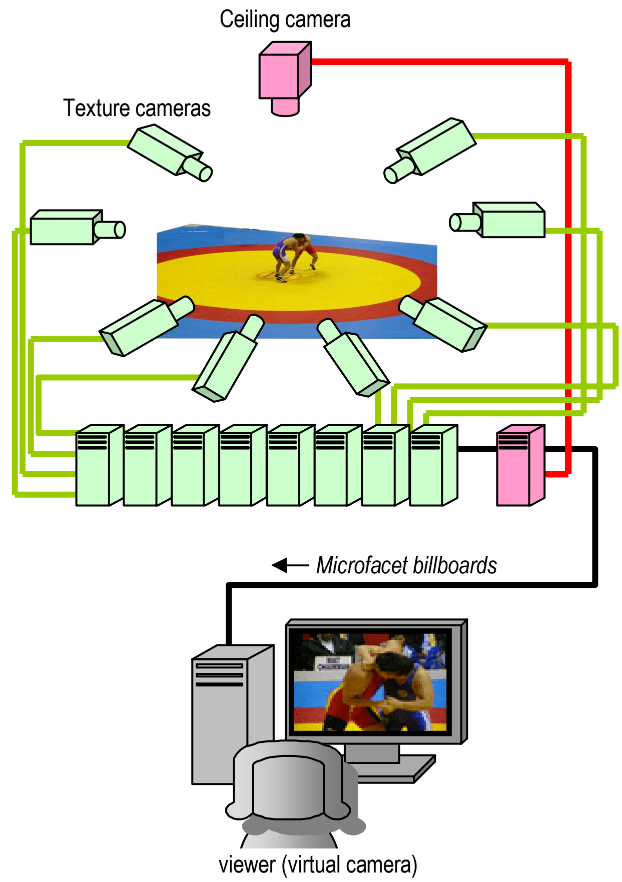

Fig. 5. System overview.

3. FREE-VIEWPOINT VIDEO GENERATION

We have implemented a preliminary video system with 9 SONY BRC-300 NTSC cameras and 10 Intel Pentium4/3.2GHz PCs. Eight synchronized texture cameras with 8 PCs are used to capture the video used both for voxel carving and texture mapping. The video size is 640×480 pixels. One camera is set on the ceiling to obtain the top-down view image, which is used only for voxel carving because it is helpful to estimate occluding voxels on texture mapping. The browsing PC sends virtual viewpoint information to the system and receives the produced microfacets *via* Gigabit ethernet. Fig. 5 shows the system overview and Fig. 6. shows the camera layout in the National Yoyogi Gymnasium, Tokyo, Japan.

We took videos of official wrestling matches using the present system. A set of snapshot images are also shown in Fig. 6.

To fulfill the image quality requirement, the facet should be less than 10 pixels. To achieve a high frame rate, it is necessary to consider the calculation cost. Fig. 7 shows the estimated times of projection calculation in the voxel carving step for the wrestling videos. Although the results are highly dependent on the object geometry, the cost function shown by the red curve almost follows the cubic order ($E \approx 1.03 \cdot 10^5 \cdot s^{-3.07}$; where s is the facet size and E is the number of projection times). With our system, we finally set the facet size to 8 pixels to achieve good balance of image quality and real-time frame rate.

Fig. 8 shows a number of snapshots. The resultant image quality was satisfactory and the system could run almost in real-time (depending on the situation and the location of the virtual camera). As players tackled and went to the ground with each other, the geometry of the players became quite intricately shaped. It is not easy to realize this level of free-viewpoint videos except with our proposed method. The judge also sometimes cut into the scene, and he was also rendered in good quality.

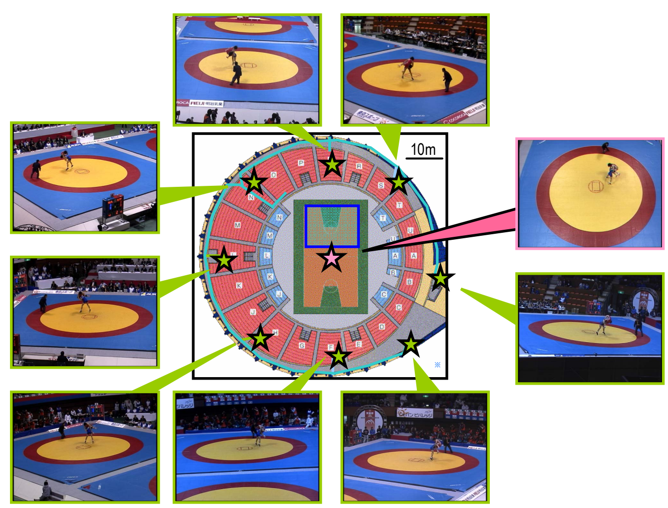

Fig. 6. Camera positions and input images.

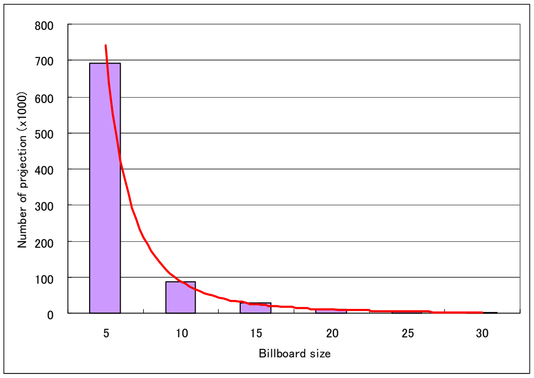

Fig. 7. Projection cost by billboard size.

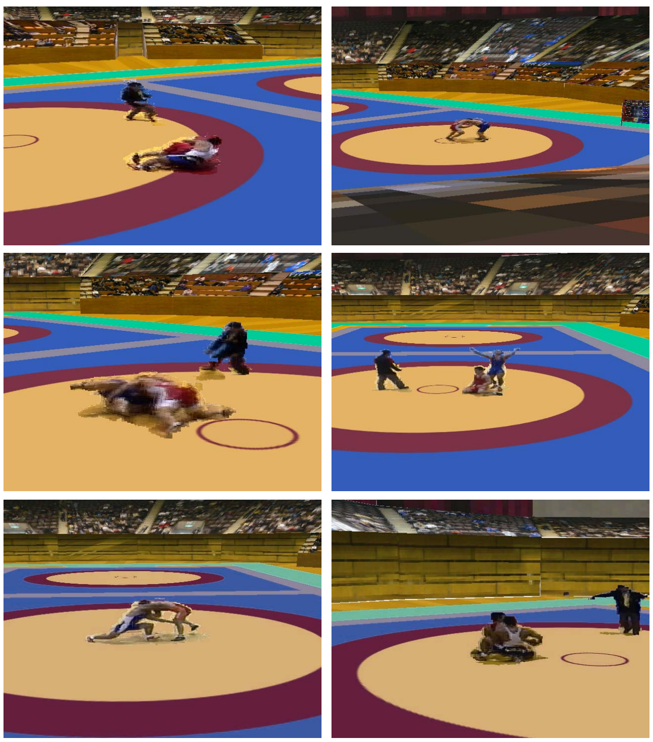

Fig. 8. Six snapshots of free-viewpoint video.

Note that the background structure model was inserted on rendering the free-viewpoint video. The CG model of the Yoyogi National Gymnasium was built in advance, and it did not affect the performance of the system.

9. CONCLUSIONS

We proposed a new method of visualizing sports players based on microfacet billboarding. As both the image quality and the frame-rate performance are crucial to produce free-viewpoint video of sports, our proposed method satisfies both requirements by controlling the voxel size depending on the viewpoint location and by finding the best microfacet size for sports video.

An emerging problem is color consistency among the texture cameras. Experiments on other sports are also expected to be performed with our method.

ACKNOWLEGMENTS

The research work was supported in part by the Strategic Information and Communications R&D Promotion Program (SCOPE) by Ministry of Internal Affairs and Communications, Japan. We are grateful to the Japan Institute of Sports Science (JISS) for their cooperation and professional advice.

REFERENCES

[1] S. Moezzi, L. Tai, and P. Gerard, "Virtual View Generation for 3D Digital Video," IEEE Multimedia, Vol. 4, No.1, pp.18-26, 1997.

[2] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM SIGGRAPH2004, Vol.23, No.3, pp.600-608, 2004.

[3] 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.

[4] H. Saito, S. Baba and T. Kanade, "Appearance-Based Virtual View Generation from Multicamera Videos Captured in the 3-D Room," IEEE Multimedia, Vol.5, No.3, pp.303-316, 2003.

[5] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan, "Image-Based Visual Hulls," ACM SIGGRAPH2000, pp.369-374, 2000.

[6] X. Wu and T. Matsuyama, "Real-Time Active 3D Shape Reconstruction for 3D Video," 3rd Int. Sympo. on Image and Signal Processing and Analysis, pp.186-191, 2003.

[7] I. Kitahara and Y. Ohta, "Scalable 3D Representation for 3D Video Display in a Large-scale Space," IEEE Virtual Reality 2003, pp.45-52, 2003.

[8] Y. Kameda, T. Koyama, Y. Mukaigawa, F. Yoshikawa, and Y. Ohta, "Free Viewpoint Browsing of Live Soccer Games," ICME, 4 pages, 2004.

[9] T. Koyama, Y. Mukaigawa, Y. Kameda, and Y. Ohta, "Real-Time Transmission of 3D Video to Multiple Users via Network," IAPR Conf. on Machine Vision Applications (MVA2005), pp.328-331, 2005.

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

[11] R. Collins, "A Space-Sweep Approach to True Multi-Image Matching," IEEE CVPR, pp.358-363, 1996.