ON THE RENDERING AND POST-PROCESSING OF SIMPLIFIED DYNAMIC LIGHT FIELDS WITH DEPTH INFORMATION

Zhi-Feng Gan, Shing-Chow Chan, King-To Ng, Kin-Lok Chan

Heung-Yeung Shum*

Department of Electrical and Electronic Engineering
The University of Hong Kong
{zfgan, scchan, ktng, kinlok}@eee.hku.hk

* Microsoft Research, Asia

hshum@microsoft.com

ABSTRACT

This paper studies the rendering and post-processing of a dynamic image-based representation called the simplified dynamic light fields (SDLF) (or plenoptic videos) with depth information. The user viewpoints are limited to a camera line to simplify the capturing and compression processes. By associating each image pixel with its depth value, methods for improving the rendering quality and detecting occlusions are proposed. Due to the limited sampling at depth discontinuities, adaptive lowpass filtering is applied to the detected occluded regions near object boundaries in order to suppress the aliasing artifacts. Rendering results using computer-generated images show that considerable improvement in rendering quality even for dynamic scenes with large depth variations.

1. INTRODUCTION

Image-based rendering (IBR) is an emerging and promising technology for photo-realistic rendering of scenes and objects from a collection of densely sampled images and videos. Another important advantage of IBR is the superior image quality that they offers over 3D model building, especially for very complicated real world scenes. They also require much less computational power for rendering, regardless of the scene complexity. Thus, they are also very useful in re-rendering complicated synthetic environment, which can be very time consuming. Since the data size of IBR is usually very large, capturing, storage, transmission and efficient rendering are the fundamental problems in IBR research. Interested readers are referred to a recent survey for more information [9].

In this paper, we study the rendering and post-processing of a four-dimensional dynamic IBR representation, called the simplified dynamic light field (SDLF) with depth information. A SDLF is a simplified lightfield of dynamic scenes with viewpoints being constrained along a line instead of a 2D plane and it is a four-dimensional plenoptic function. This can greatly reduces the complexity of real-time capturing systems [8]. One problem with lightfield rendering (IBR) is the number of samples required to render the scene without introducing excessive artifacts. If the scene is free of occlusions, then the concept of plenoptic sampling can be applied to determine the sampling rate in the camera plane. Unfortunately, because of the depth discontinuity problem to be explained later, the sampling rate is usually insufficient around object boundaries. This motivates us to study the performance and the necessary processing required for such a renderer, if a sufficiently accurate depth map is available. We assume that each image pixel in a lightfield (in general any IBRs) has a color as well as depth value. This is reasonable, especially when the ligthfield is obtained from a 3D range scanner. We then examine a matching method for detecting possible occlusions during rendering and find a more consistent value of the rendered pixels. This is similar to some kind of real-time disparity estimation and it is computational intensive. Nevertheless, it is fully justified by its excellent performance. We also found that the matching process, as predicted by plenoptic sampling with depth discontinuities, gives erroneous results around image discontinuities where occlusions occur. To overcome this problem, we propose an adaptive post-processing technique to

handle these "in-determinate" (discontinuous boundary) regions. Rendering results using computer-generated images show that considerable improvement in rendering quality even for dynamic scenes with large depth variations. The paper is organized as follows: a brief introduction to the plenoptic function and the simplified dynamic light field (SDLF) is given in Section 2. The proposed rendering and post-processing algorithms for SDLFs with depth values are described in Section 3. Experimental results are presented in Section 4. Finally, conclusions are given in Section 5.

2. SIMPLIFIED DYNAMIC LIGHT FIELD (SDLF)

Central to IBR is the plenoptic function, which was first used by Adelson and Bergen [1] to describe all the radiant energy that can be perceived by the observer at any point (V_x, V_y, V_z) in space and time τ. At each point in space, we can select any of the viewable rays by choosing an azimuth and elevation angles (θ, ϕ) as well as a band of wavelengths, λ. For dynamic scenes, there is an additional time variable τ. Therefore, the plenoptic function p has seven dimensions. Due to its high dimensional nature, data reduction (compression), transmission and efficient rendering of the plenoptic function are essential to IBR systems. Fig. 1(a) illustrates the principle of an 4D plenoptic function called light field [3] or lumigraph [2] (lumigraph differs from light fields in using additional depth information), where images on a camera plane are taken to obtain the rays for rendering the scene. Fig. 1(b) shows a simplified dynamic lightfield (SDLF) for capturing dynamic scenes where user viewpoints are constrained along a camera line. The compression and capturing of SDLF were recently studied by the authors in [7] and [8], respectively. Because of the close relationship between the SDLF with traditional videos, we also referred it to as the "plenoptic videos" to signify its dynamic nature. Lower dimensional IBR representations can be obtained by elegantly restricting the users' viewpoints as in the two-dimensional panoramas, three-dimensional concentric mosaics, and other variations of these representations.

A. The Depth Discontinuity Problem

One problem with lightfield rendering is the number of pictures we need to render a given scene. This problem was studied by Chai et al under the concept of plenoptic sampling [5]. Assuming that there are no occlusions or depth discontinuities, it was found that the spectral support of a lightfield is dependent on the depth of the objects. This allows us to determine the sampling rate for proper reconstruction of the lightfield and to avoid undesirable aliasing effects. The results not only lead to improved understanding of the basic principle behind IBR, but also simplify the practical implementation of IBR system in terms of capturing, rendering and processing. For example, in the rendering of plenoptic videos [7,8], fairly high rendering quality can be achieved by using a mean depth for scenes where depth variation is relatively small. However, for scenes with large depth variations, rendering artifacts such as ghosting and blurring are very difficult to avoid around depth discontinuities (e.g. in Fig. 3), because extremely high sampling will be required.

In the spirit of plenoptic sampling [5], there are two possible methods to alleviate this problem. First, from a sampling theorem point of view, these aliasings can be reduced by filtering the lightfield with a reconstruction filter having a small bandwidth or support. However, the resolution of the lightfield will suffer at the same time. We are thus forced to live with the tradeoff between aliasing and resolution reduction. Another method, which can potentially yield better rendering results, is to compute a depth map of the scene so that a mean depth or very rough depth map can be used for objects without depth discontinuities [10]. Filtering is then applied around the boundaries to reduce the artifacts generated locally due to insufficient sampling at the discontinuities. As mentioned earlier, we are interested in the performance and the necessary processing required for such a renderer, if a sufficiently accurate depth map is available. Since in a practical multi-camera system, an accurate global depth map is rather difficult to compute, some kind of local depth maps are desirable. Following our previous work on panoramic videos with depth information [12] and the concept of layered depth images [6], we assume that each image pixel in a lightfield (in general any IBRs) has a color as well as depth value. This is reasonable, especially when the ligthfield is obtained from a 3D range scanner. Using a similar concept as in [12], it is possible to shot a ray to determine whether there is a possibility of occlusion during rendering. Using this concept, it is possible to find more consistent values of the rendered pixels. This is similar to some kind of real-time disparity estimation, which is somewhat computational intensive. Nevertheless, it is fully justified by its excellent performance. Even so, we found that the matching process, as predicted by plenoptic sampling with depth discontinuities, gives erroneous results around image discontinuities where occlusions occur. To overcome this problem, we propose an adaptive post-processing technique to handle these "in-determinate" (discontinuous boundary) regions. Instead of rendering their values from the basic principle of lightfield rendering (because the depth information cannot resolve this problem for us anymore), we based our method on the concept of image consistency and limited spatial resolution of the images. These will be further elaborated in the following sections.

3. RENDERING AND POST-PROCESSING OF SDLF

Fig. 1(a) illustrates the principle of lightfield rendering in the standard two-plane ray space parameterization. There is a camera plane, with parameter (s,t), and a focal plane, with parameter (u,v). Each ray in the parameterization is uniquely determined by the quadruple (u,v,s,t). For fixed values of s and t, we obtain an image taken at a location indexed by (u,v). Interested readers are referred to [2,3] for more details. Fig. 1(c) shows an example of a simple 2D light field. Assuming a pin-hole camera model, the pixel value observed is the convolution of the plenoptic function with the point spread function. At the camera position t_1, a ray r_1 of a point P on the object surface is observed, and it is recorded as a pixel at position v_1. Whereas, for camera position t_0, the same point P at depth z_P is observed as ray r_0 and is recorded as pixel v_0. Usually a global depth map is assumed. The disparity, or the displacement of the image pixel of P is $f\Delta t / z_P$, where f is the focal length and Δt is the change in t direction. If the object surface is Lambertian, pixels v_0 and v_1 will be identical to each other. If z_P is known, then the pixels captured in certain camera positions (say regularly along a straight line) can be used to reconstruct the images in between, if the sampling is sufficiently dense and there is no occlusion. If the depth

variation is limited, we can assume that all objects are coplanar with a mean depth as shown in Fig. 1(c). To render a ray r at a camera position t, we calculate its intersection with this plane and shoot back two rays r_0 and r_1 to the camera positions t_0 and t_1, where we obtained two approximations of point Q, which is the point that would be seen by placing a camera at t in direction r. If the surface is Lambertian, we can average the two corresponding pixels recorded at t_0 and t_1 to yield pixel v. Since the depth variations in Fig. 1(c) are relatively small, the artifacts introduced are usually acceptable.

A. Depth Matching and Occlusion Detection

Fig. 1(d) illustrates the depth discontinuity problem, where the point P is near the boundary of an inserted object OBJ, but is not occluded. Suppose that each image pixel has its own depth value $z_0(v_0)$ and $z_1(v_1)$, say obtained from a 3D digitizer. When we render a ray r at the virtual camera position t, we do not know the depth value associated with this ray initially. However, we can assume that the rays are coming from a set of possible depth values say from $z_{\min}$ to $z_{\max}$ as shown in Fig. 1(d). For each potential depth value, say P', we shoot back two rays to t_0 and t_1, (r_0' and r_1'), and obtain the corresponding image pixels, (v_0' and v_1'), together with their depth values ($z_0(v_0')$ and $z_1(v_1')$). By comparing these two depth values with the distance from P' to t_0 and t_1 (i.e. $\| r_0' \|$ and $\| r_1' \|$, respectively), we can determine the most possible depth value $z(v)$ such that $z_i(v_i')$ is as close to $\| r_i' \|$ as possible. By interpolating the corresponding pixels (v_0 and v_1) found from the captured images at t_0 and t_1, better rendering quality can be achieved. When r moves gradually towards OBJ, the rays of t_1 are blocked by OBJ and the matching errors of $z_0(v_0)$ and $z_1(v_1)$ will be very high for all possible depth values. Therefore, an occlusion is detected. The two backward shooting rays will rejoin again, when r further points to the front surface of OBJ and the occlusion disappears. The above discussion suggests that depth information provides a useful means to improve the rendering quality, but it is still limited by depth discontinuities around object boundaries.

B. Adaptive Post Processing

Although it is very difficult to determine the exact depth values at object boundaries, we can still "guess" their intensity values from image consistency. Fig. 4 shows the "in-determinate" regions after depth matching. It can be seen that they come mostly from image edges. Therefore, it is possible to recover them if some kinds of edge models are available. In this paper, we adopt a simple adaptive post processing approach. For each horizontal line in the "in-determinate" regions, we determine the mid point of this line and perform linear interpolation from the two image pixels just before and after the "in-determinate" line. The weights of the interpolation determine how sharp the edges will look like. In this way, the occluded regions are filled by pixels from the rendered image itself. Further, if the resolution of the depth map is limited, then the image consistency can be employed to improve the rendering quality. For example, a matching criterion consisting of the depth and color values can be used. These will be left for future research.

4. EXPERIMENTAL RESULTS

The SDLF used in our experiments is synthetic and it was rendered by using the 3D Studio Max software. Figure 2 shows part of the SDLF used in this study. It consists of 16×1 24-bit

RGB videos with 320×240 pixels and 24 frames per second, Fig. 2(a). The depth value of each pixel (10-bit) is recorded during the rendering, Fig. 2(b). Fig. 3 shows several renderings with different mean depths. Blurring and ghosting artifacts are observed. Fig. 4 shows the renderings obtained by the rendering algorithm in section 3. The occluded regions are detected and are shown in black color. The matching is performed initially on a coarse grid, followed by a finer search around the potential candidates. Most of them are around object boundaries. After post-processing, the aliasing artifacts (isolated impulsive image pixels, not shown here) are effectively suppressed and the visual quality is considerably improved. Fig. 5 shows other renderings of the SDLF with different depth variations. It shows that the proposed algorithm employing depth information gives better rendering quality than traditional method using mean depth. Almost all occlusion has

been resolved, and there are slight artifacts around the image edges due to the high resolution of the graphics images.

5. CONCLUSION

The rendering and post-processing algorithms of the simplified dynamic light fields (SDLF) or plenoptic videos with depth information are presented. By associating each image pixel with its depth value, the rendering quality can be improved and possible occlusion regions can be detected. Adaptive post-processing in form lowpass filtering is applied to the detected occluded regions near object boundaries for aliasing suppression. Rendering results using computer-generated images show that considerable improvement in rendering quality can be achieved even for dynamic scenes with large depth variations. Locally dense depth map [11] is thus a fruitful area of research in dynamic IBR.
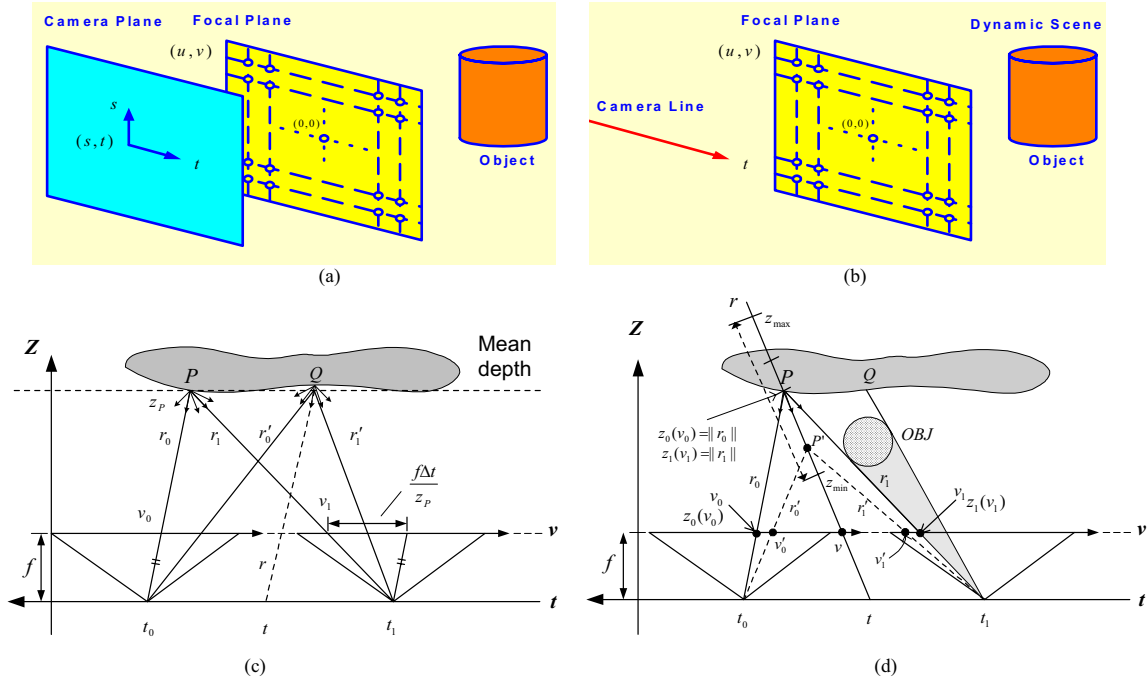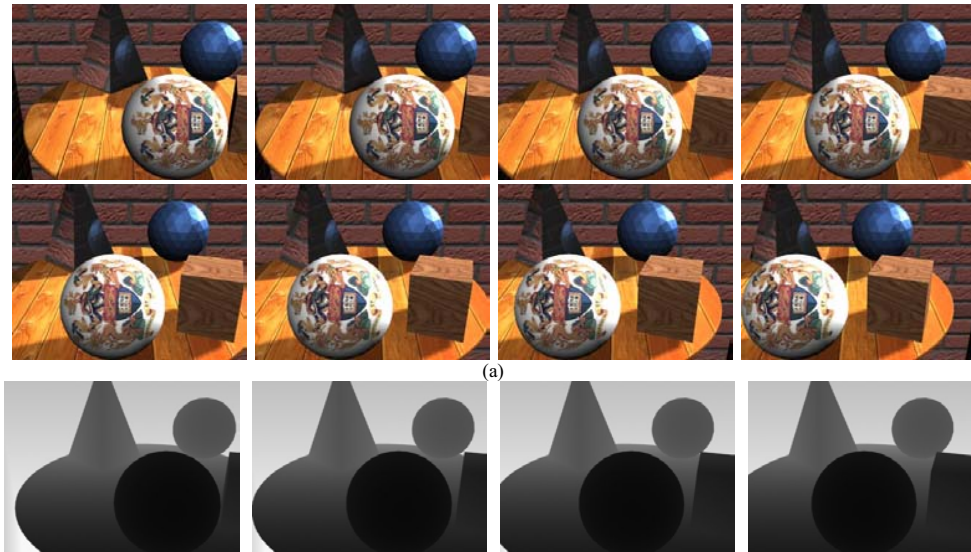

Fig. 1. (a) 4D static light fields: viewpoints constrained on a 2D plane. (b) 4D simplified dynamic light fields (the plenoptic video): viewpoints constrained along a line in a dynamic environment. (c) Rendering of 2D light field using mean depth. (d) Rendering using pixels with depth value. The object surface is assumed to be Lambertian.
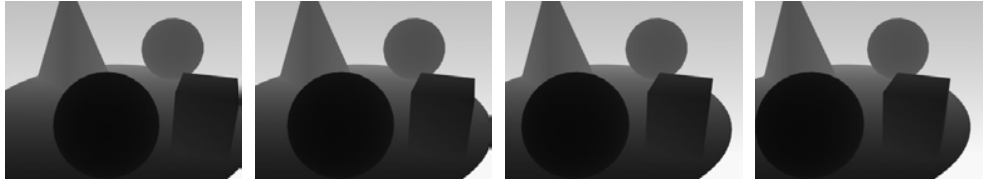

(a)

(b)

Figure 2. (a) Original light field images (b) 10-bit depth maps of a snapshot of the SDLF. Camera 0 to camera7 (from top to bottom, left to right).

Figure 3. From left to right: renderings with mean depths at the back, middle and front objects of the SDLF.
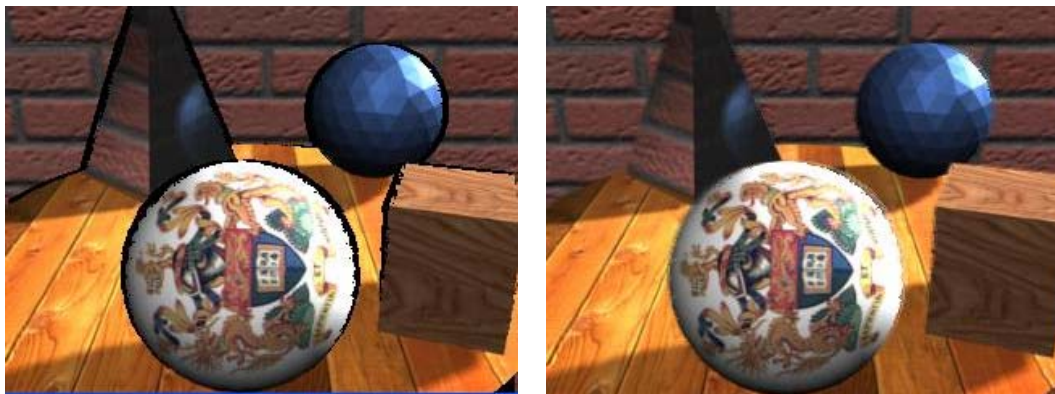

Figure 4. Renderings with depth values: left: occlusion region detected (black color); right: post-processing of occlusion regions.
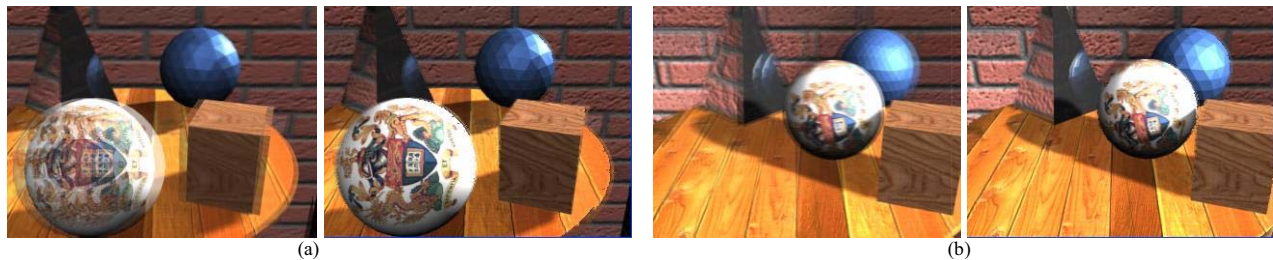

(a)   (b)

Figure 5. Rendering at two time instants (a) and (b). Left : using mean depth. Right : using depth values and post-processing.

REFERENCES

[1]   E. H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, pp. 3-20, MIT Press, Cambridge, MA, 1991.

[2]   S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph," in *Proc. of the annual conference* on *Computer Graphics (SIGGRAPH'96)*, pp. 43-54, Aug. 1996.

[3]   M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'96)*, pp. 31-42, Aug. 1996.

[4]   L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'95)*, pp. 39-46, Aug. 1995.

[5]   J. X. Chai, X. Tong, S. C. Chan and H. Y. Shum, "Plenoptic sampling," in *Proc. of the annual conference on Computer Graphics (SIGGRAPH'00)*, pp. 307-318, July 2000.

[6] J. Shade, S. Gortler, L.-W. He, and R. Szeliski. Layered depth images. In *Computer Graphics (SIGGRAPH'98) Proceedings*, pages 231–242, Orlando, July 1998. ACM SIGGRAPH.

[7] S. C. Chan, K. T Ng, Z. F. Gan, K. L. Chan and H. Y. Shum, "The data compression of simplified dynamic light fields," in Proc. ICASSP'2003.

[8]   S. C. Chan, K. T Ng, Z. F. Gan, K. L. Chan and H. Y. Shum, "The plenoptic videos," submitted to IEEE Trans. on Circuits and Systems for Video Technology, 2003.

[9] Heung-Yeung Shum., Sing Bing Kang, and Shing-Chow Chan, "Survey of Image-based Representations and Compression Techniques," to appear in IEEE Trans. CSVT.

[10] H. Y. Shum, J. Sun, S. Yamazaki, Y. Lin, C. K. Tang, "Pop-Up Light Field: An Interactive Image-Based Modeling and Rendering System," to appear in ACM Trans. on Graphics.

[11] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47(1), pp. 7-42, May 2002.

[12] K. L. Chan and S. C. Chan, "Video object coding and rendering for panoramic video, "in Proc. European Signal Processing Conference (EUSIPCO2002), Vol. II, pp. 506-509, Sept. 3-6, 2002, Toulouse, France.