RENDERING DRIVEN DEPTH RECONSTRUCTION

Yin Li[†,‡], *Xin Tong*[†], *Chi-Keung Tang*[‡], *Heung-Yeung Shum*[†,‡]

[†]Microsoft Research Asia, Beijing, PRC
[‡]Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

ABSTRACT

Previous work on image-based rendering suggests that there is a tradeoff between the number of images and the amount of geometry required for anti-aliased rendering. For instance, plenoptic sampling theory indicates that visually acceptable rendering can be achieved when the input images are undersampled, if sufficient depth information is available for all the pixels. In this paper, we propose a novel vision reconstruction approach, *rendering-driven depth recovery*, to recover the amount of geometry that is necessary for anti-aliased rendering. Our approach contrasts conventional stereo reconstruction in that we do not intend to accurately reconstruct the depth for each and every single pixel, leading to a very efficient reconstruction algorithm. Our algorithm uses a block-based multi-layer depth representation, and searches in the depth space based on the causality criterion, by detecting double images. Experiments show that rendering systems using our rendering driven depth recovery algorithm can synthesize satisfactory novel views efficiently by using 'just enough geometry' recovered from undersampled input images.

1. INTRODUCTION

Recent work on image based rendering reveals a spectrum of image-based representations based on different tradeoffs between the number of images and the amount of geometry information used. At one extreme, a large number of images is acquired in lightfield [1], lumigraph [2], and concentric mosaics [3]. Rendering a novel view is achieved by interpolating a subset of densely sampled images. Hence, no geometry information is needed. However, the high consumption of storage and memory have limited their applications. At the other extreme, the demand for images will decrease drastically when accurate geometry is available. A novel image can be rendered from a nearby view by directly warping the input images, according to their depth [4] [5]. However, accurate geometry information is very difficult to obtain by vision methods from real scenes.

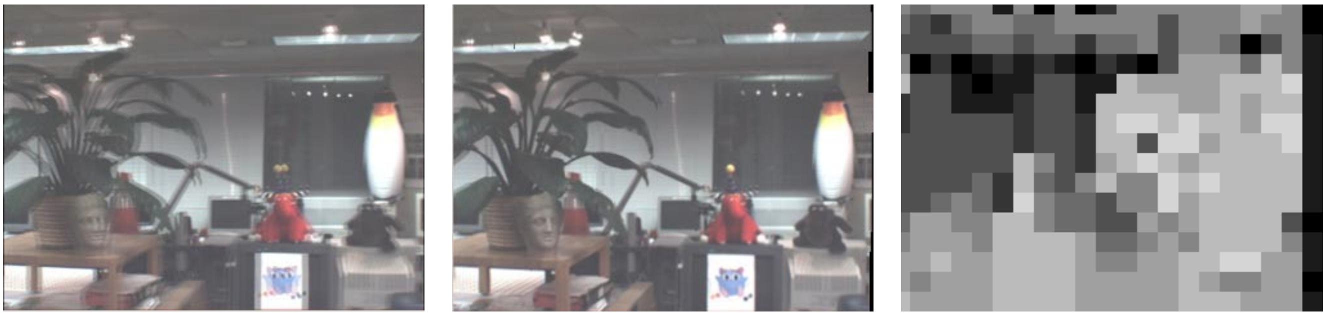
It is observed that *approximate* geometry information can also improve rendering quality for moderately image samples [2] [6]. Recently, the sampling analysis of [7] and [8] indicate that the amount of geometry and the number of images necessary for anti-aliased rendering can be characterized by the *minimal sampling curve* in the joint image and geometry space. This is called the plenoptic sampling theory. In other words, the same rendering result can be obtained from different image and geometry combinations: either 1) a conventional densely sampled lightfield, or 2) an undersampled set of images with the approximate geometry, or 3) a sparse set of images with accurate geometry. It is observed that representations below the sampling rate will lead to the aliasing effect in the rendering result, which are presented as *double images* (Figure 1) in the rendering image [7], or overlapping of the replica in the frequency domain [8].

In this paper, the problem we want to address is: given an undersampled light field, we want to recover the amount of geometry necessary to render an anti-aliased image at a novel viewpoint. To this end, we propose an efficient *rendering driven depth reconstruction* algorithm. Our algorithm uses a block-based multi-layer depth representation, and searches in the depth space based on the causality criterion, by detecting double images.

Our algorithm is different from previous stereo algorithms in several aspects. First, we do not intend to compute an accurate geometric reconstruction. Rather, we are interested in recovering the amount of geometry that is necessary for, or driven by, anti-alised rendering. In conventional stereo reconstruction, due to the ill-posedness of the stereo matching [9], the problem is often formulated into an iterative and time-consuming optimization. Further, since the recovered depth is used for rendering, only local geometries need to be reconstructed for each adjacent pair of images. In many conventional stereo algorithms, these local geometries must be merged into a uniform geometric representation for all images, which is also a difficult problem.

Unlike many vision reconstruction algorithms, which are executed as an independent preprocessing step before rendering, in our algorithm, the rendering and depth recovery processes cooperate with each other. By detecting aliasing effect (i.e. double image) in the rendering result for different candidate depths, our algorithm terminates once a satisfactory depth recovery result for antialiasing image-

0-7803-7663-3/03/$17.00 ©2003 IEEE



(a) (b) (c)

Fig. 1. The LAB scene. (a) Double image artifact when sampling rate is not enough for lightfield rendering. (b) Rendered result at the same viewpoint obtained by our modified lightfield rendering system with depth, after double image elimination. (c) Approximate depth is used to improve the rendering quality.

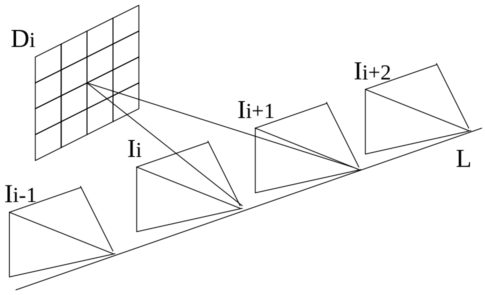


Fig. 2. 1D lightfield configuration.

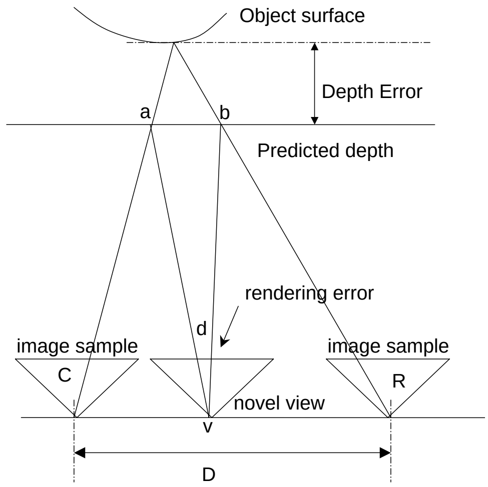


Fig. 3. The stereo problem for each adjacent samples.

based rendering is found.

The remainder of paper is organized as follows. The overview of our algorithm is given in Section 2. Space and time complexities of our algorithm are analyzed in Section 3. We present the experimental results in Section 4. Finally, we draw our conclusion, and propose future work in Section 5.

2. ASSUMPTIONS AND OVERVIEW

In this section, we start by stating our assumptions. The 1D lightfield system for analysis is then introduced. An overview of our algorithm is also given here.

2.1. Assumptions

We assume the scene is Lambertian, since we use local matching in our algorithm. We also assume that occlusion is not significant enough to affect local matching.

We assume the minimal and maximal depths are known, but they need not be very accurate. If the maximal depth is unknown, we can assume it to be infinity, i.e., zero disparity between neighboring views. The minimal depth, in terms of disparity, should be assigned a value larger than the disparity of the nearest object between two neighboring images.

2.2. 1D Lightfield

For simplicity and without losing generality, we use a 1D lightfield system to describe and explain our algorithm. This 1D lightfield system consists of a sequence of images sampled at equal interval along a line L, as shown in Figure 2. Each camera on L has the same pose, and their respective image scanlines are parallel to L. In fact, this is a special case lightfield, with the number of vertical samples equal to one. In this 1D configuration, since the baselines and viewing angles for all cameras are same, we use the disparity between two adjacent images to represent the depth (or inverse depth). The generalization to 2D lightfield is one straightforward subject of future work.

2.3. Algorithm Overview

Given a 1D lightfield, the maximal and minimal depth of the scene, our algorithm recovers sufficient depth information for rendering. The algorithm outline is depicted as follows.

1. For every adjacent view pair, apply the following four steps. Each iteration will generate a matrix of constant depth patches for this view pair. Let $|V|$ be the number of views in the 1D lightfield system. Then, at the end of the algorithm, a total of $|V-1|$ patches are obtained, one for each pair of views.

2. For the selected adjacent pair of views, denote the left view as C and right view as R as shown in Figure 3.

3. Subdivide the depth map on view C into a set of equal-sized blocks or patches. A matrix of patches is thus obtained. Each patch contains a constant depth;

4. For each candidate depth in the depth search space, repeat the next testing step until there is no double image;

5. Hypothesize the current candidate depth for each patch. Warp the image C according to the hypothesized depth onto the view R. Check the existence of any double image;

After the depth for each patch is recovered, the lightfield with multi-layer, constant depth patches can be rendered with our modified lightfield rendering algorithm.

3. SPACE AND TIME COMPLEXITIES

Refer to the algorithm outline in Section 2. Our algorithm only iterates for all pair of adjacent images. Let V be the number of images in 1D lightfield, N be depth range, and M the total number of pixels in one image. The total space complexity is $O(MV)$, and the total time complexity is $O(VNM)$.

4. EXPERIMENTAL RESULTS

Synthetic and real world lightfield data are used to evaluate our algorithm. These lightfield data are actually 2D lightfield ones. We extract one row of the data to generate a 1D lightfield.

Downsampling of the lightfield is done manually, by extracting the sampled images at some constant interval, say 2, or 4. The depth recovery algorithm and the modified rendering only use the downsampled data, assuming that other data has never been captured.

4.1. Experiments on synthetic data

The synthesized lightfield, NETFERT, contains 65 images for the original data. We downsample it and produce a 1D lightfield with only 9 sample images. The double image artifact is easily noticed in Figure 4(a) and (b). Adjusting one constant depth alone cannot improve the condition, because the sampling rate is already below the minimal sampling rate for conventional light field rendering [8]. After the depth recovery from our algorithm, the rendering result (Figure 4(c)) is evidently more satisfactory, which is free of double image. The depth map is shown in Figure 4(d), which shows the recovered matrix of constant depth patches. The depth map obviously does not represent the accurate geometry of the object. However, this depth information is adequate to produce a satisfactory rendering result when cooperating with our modified lightfield system.

4.2. Experiments on real data

Two lightfield data captured from real world scenes: the PLANT and LAB [6] are used in our experiment.

The PLANT data contains a 16×16 lightfield. We extract one row from it and downsample it by a factor of 4, generating a 4×1 lightfield. From the Figure 4(a), we can easily observe the double image artifact caused by downsampling. When we use the depth information recovered from our algorithm, our modified lightfield can render the same scene without double image at the same viewpoint.

The result of another challenging experiment on the LAB data is shown in Figure 1. This time we downsample it by a factor of 4 and generating a 8×1 lightfield. Figure 1(a) shows the rendering result by constant depth lightfield. The window and lamp at the background are clear and sharp, while the screen in front suffers severe double images. In the Figure 1(b), a satisfactory image is generated at the same viewpoint. Thanks for the depth information recovered by our algorithm, Figure 1(c). However, we can still observe some double image are present at severely occluded regions.

5. CONCLUSION AND FUTURE WORK

In this paper, we propose an efficient depth recovery algorithm, which is motivated and driven by the anti-aliased rendering, requiring that no double image should be present. Our algorithm subdivides the depth map for each view pair into a set of multi-layer patches, assigning the depth of each patch to reduce the effect of double image.

In our future work, we will extend this algorithm to the more commonly used 2D lightfields. We are also interested in exploring non-regular partition scheme for the depth map, which may lead to more compact and flexible representation for our ‘just enough geometry’.

(a) (b) (c) (d)

Fig. 4. The NETFERT scene. (a) Rendering with constant depth at nose. Double images at the neck and shoulder are evident. (b) Rendering with constant depth at neck. Now, double images of the nose are evident. (c) Rendering result with our modified lightfield, using the recovered depth patches. (d) The recovered depth patches. The depth does not correspond to accurate geometry; however, it produces satisfactory rendering result.

(a) (b) (c)

Fig. 5. The PLANT scene. (a) Rendering with constant depth at the background. Double images on the flowers are observed. (b) Rendering result with our modified lightfield, using the recovered depth patches. (c) The recovered depth patches are used to improve the rendering quality.

Finally, occlusion remains to be an open issue. Subdividing the depth map may help alleviate the situation. However, if there exists a large depth discontinuity (too large that the sampling rate of images cannot handle), visual artifact will still be noticed. We are now exploring alternatives such as user-assisted segmentation and semi-automatic matting algorithm to address this problem.

6. REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. SIGGRAPH'96*, 1996, pp. 31–42.

[2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proc. SIGGRAPH'96*, 1996, pp. 43–54.

[3] H.-Y. Shum and L.-W. He, "Rendering with concentric mosaics," in *Proc. SIGGRAPH'99*, 1999, pp. 299–306.

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

[5] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proc. SIGGRAPH'95*, 1995, pp. 39–46.

[6] A. Isaksen, L. McMillan, and S. Gortler, "Dynamically reparameterized light fields," in *Computer Graphics (SIGGRAPH 2000) Proceedings*. 2000, pp. 297–306, ACM SIGGRAPH.

[7] Z.-C. Lin and H.-Y. Shum, "On the number of samples needed in light field rendering with constant-depth assumption," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2000(CVPR'2000)*, 2000.

[8] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Computer Graphics (SIGGRAPH'2000)*, 2000, pp. 307–318.

[9] M. Bertero, T.A. Poggio, and V. Torre, "Ill-posed problems in early vision," *PIEEE*, vol. 76, no. 8, pp. 869–889, August 1988.