Light Field Capturing with Lensless Cameras

*Cha Zhang**
Microsoft Research
One Microsoft Way, Redmond, WA 98052
chazhang@microsoft.com

Tsuhan Chen
ECE, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213
tsuhan@cmu.edu

ABSTRACT

We introduce a novel approach to capturing light field with lensless cameras. By moving the cameras back and forth, we capture a set of images. We show that it is possible to reconstruct the light field from these blurry images. The problem is formulized in a way similar to computer tomography, so that the light field can be reconstructed using existing algorithms. The light field can then be used to render 3D scenes. Synthetic examples are presented to show the effectiveness of the proposed method.

1. INTRODUCTION

Image-based rendering (IBR) has attracted a lot of research interest recently [1]. By capturing a set of images or light rays for a 3D scene, IBR can reproduce the scene correctly at arbitrary viewpoint without too much knowledge of the scene geometry. Compared with the traditional model-based rendering, IBR provides the benefits that images are easier to obtain, simpler to handle and more realistic to render.

In practice, in order to capture an IBR scene, one needs to point a camera (or a camera array) to the scene and take many images from different positions. Most existing IBR approaches assume the capturing cameras are ideal, i.e., they are pin-hole cameras. However, practical cameras are often equipped with lenses. Although there have been many different kinds of lenses invented for different purposes, one limitation of having lenses in a camera system is that the camera will have a limited depth of field (DOF). This can cause problem for IBR when the scene has large depth variations, because not all the objects can be in focus in a single image. One way to leverage this problem is to take multiple images at different focal lengths while the camera stays at each of the capturing positions. Work has been done to render all in-focus images from such captured image sets [3][4]. Nevertheless, moving the cameras while adjusting the focal lengths at each camera position is tedious and time-consuming, not to say that the focal length of each image has to be measured accurately for the rendering.

In this paper, we propose a novel approach to capturing the scene with lensless cameras. We remove the lenses in the cameras, thus each pixel in the captured images is the sum of the light rays entering that pixel from different directions. As shown in Figure 1, such a lensless sensor is placed in front of the object. We move the camera forward and backward and capture many images of the scene. We show that although the captured images are blurry, using technology similar to computer tomography (CT), we may reconstruct the individual light rays via existing methods. Other than saving the cost of complex lenses, the major advantage of the proposed scheme is the ability to produce images with infinite DOF. This is achieved without extra assumption about the nature of the light sources and illuminations.

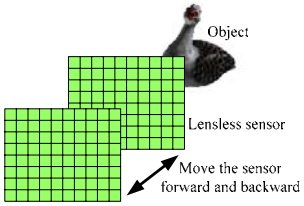

Figure 1 Use lensless cameras/sensors to capture a 3D scene.

The paper is organized as follows. Section 2 gives a brief introduction on light field, which is a representative IBR representation. The proposed capturing scheme is detailed in Section 3. Related work is described in Section 4. Some experimental results are shown in Section 5. Conclusions and future work are given in Section 6.

2. THE LIGHT FIELD

Consider a scene as shown in Figure 2. For illustration purpose, we use 2D examples throughout this paper, although they can be easily extended to 3D. If the light rays emitted/reflected from the objects in the scene do not change their intensity along their path in the free space, we may parameterize them by their crossing points with two horizontal lines, namely the *v* axis and the *t* axis. That is, any light ray in the space can be represented as $l(t,v)$. Such a representation is named the light field [2] of the scene. Notice that we adopt a local *v* axis whose origin aligns with the *t* coordinate. Let *f* be the distance between the *v* axis and the *t* axis; θ be the angle between the light ray and the vertical axis *z*. The local *v* axis setup means:

$$v = f \tan\theta . \quad (1)$$

If $l(t,v)$ is available for any pair of *t* and *v* values, a novel view at arbitrary viewpoints can be easily synthesized by finding the (t,v) pair for each rendered light ray. In practice, we may only store $l(t,v)$ at discrete (t,v) coordinates. The rendering can still be performed by interpolation, which has been shown in [2]. It is also shown that if the light field is sampled dense enough, no scene geometry is needed for 3D photorealistic rendering. We refer the reader to [1] for more details about light field rendering or image-based rendering techniques.

A simple approach to capture such a light field is to use many cameras placed on the *t* axis and captures images. The focal lengths of the cameras are set to be *f*, such that the imaging planes of the cameras overlap with the *v* axis. Such scheme works well if the scene has small depth variations and the cameras have large DOF. The capturing cameras can be considered as pin-hole ones. Unfortunately, when the depth variation is large, common cameras cannot focus on all the objects in the scene due to limited DOFs.

* This work was performed while Dr. Zhang was at Carnegie Mellon Univ. and was supported by NSF Career Award 9984858.

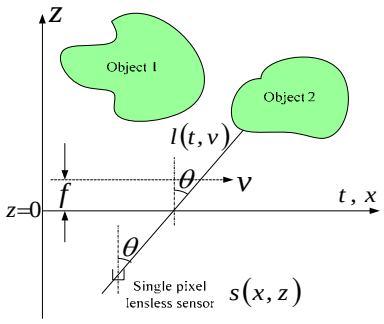

Figure 2 Capture a 3D scene with lensless camera.

3. THE PROPOSED METHOD

Consider a single pixel on the lensless sensor, as shown in Figure 2. In the 2D world, its position is (x,z), where x is the horizontal position (it is the same axis as the t axis in the light field representation) and z is the vertical position (depth). The sensor can move freely in the 2D world. The intensity measured at each position is represented as $s(x,z)$. In practice, a lensless camera (film or CCD sensor) can be placed horizontally to record $s(x,z)$ at a certain depth z. Moving the camera along the depth axis (refer to Figure 1) will record the whole $s(x,z)$ signal. The goal is to reconstruct the light field $l(t,v)$ based on the recorded signal $s(x,z)$.

A. The lensless sensor model

In practice, any sensor responds differently to light rays entering from different directions. Therefore, we may write the acquired data $s(x,z)$ as:

$$s(x,z)=\int_{-\pi/2}^{\pi/2} l_{(x,z)}(\theta)g(\theta)d\theta \tag{2}$$

where θ is the angle between the entered light ray and the z axis. $l_{(x,z)}(\theta)$ is the light ray passing through the sensor at (x,z), and has an angle θ. $g(\theta)$ is the sensor gain for light rays coming along direction θ. Without loss of generality, we assume that this gain function does not change at different places on the sensor (independent of x and z). An example of such a gain function is the well-known foreshortening effect in computer vision [5], where

$$g(\theta)=\cos(\theta) \tag{3}$$

In reality, this gain function can be measured by recording the sensor response to point light sources at different θ directions. Hereafter we assume the gain function is known.

According to Equation (1), there is a one-to-one mapping between θ and v. Therefore, through a variable transform, we can easily obtain the relationship between the captured data $s(x,z)$ and the light field $l(t,v)$:

$$s(x,z)=\int_{-\infty}^{\infty} l\left(x-\frac{v}{f}z,v\right)g(\theta)\frac{d\theta}{dv}dv \tag{4}$$

Notice the item $g(\theta)\frac{d\theta}{dv}$ in the above integration is a function of v. We can merge it with $l\left(x-\frac{v}{f}z,v\right)$ by defining:

$$\tilde{l}(t,v)=l(t,v)g(\theta)\frac{d\theta}{dv} \tag{5}$$

Consequently we have:

$$s(x,z)=\int_{-\infty}^{\infty}\tilde{l}\left(x-\frac{v}{f}z,v\right)dv \tag{6}$$

If the function $\tilde{l}(t,v)$ can be recovered from $s(x,z)$, we may effortlessly obtain the light field $l(t,v)$ through Equation (5).

B. Explanation in the light ray space

The light ray space is the space defined by the axis t and v. It is also called the epipolar image of the light field (EPI). As shown in Figure 3, when x and z are fixed, the integration of Equation (6) is effectively along a line which can be represented as:

$$v=-\frac{f}{z}(t-x) \tag{7}$$

Notice such integration resembles that of the standard computer tomography (CT) [6]. Equation (6) is the standard definition of Radon transform [7]. Therefore, there are many existing algorithms that can be used to solve $\tilde{l}(t,v)$ from $s(x,z)$ captured with the lensless cameras.

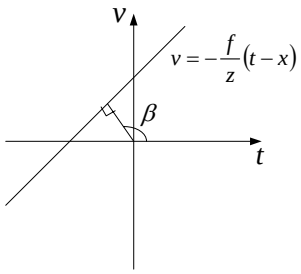

Figure 3 The explanation of our lensless camera approach in the light ray space.

An implication from Figure 3 is that we should not place our lensless films uniformly along z. The projection direction β and depth z has the following relationship:

$$\beta=-\tan^{-1}\frac{f}{z}+\frac{\pi}{2} \tag{8}$$

Therefore, the depth z of the lensless sensors should be arranged such that we have a uniformly spaced projection direction β. This results in a larger distance between sensors when z is large, and vise versa.

C. The reconstruction algorithm

While in CT the sensors can be placed along arbitrary directions due to the free rotation of the emitter/sensor pair, in our approach we cannot place the lensless sensor anywhere in the space. For example, if the scene objects occupy a depth range from $z_{\min}$ to $z_{\max}$, our sensor cannot be there due to physical occupancy confliction. In the CT literature, such situation is referred as the limited-view problem and has been widely studied [9][10][11].

In this paper, we adopted an algebraic reconstruction technique (ART) algorithm [8][11] to recover the light field, as explained below.

We first discretize the light field into sample light rays, as was done in [2] (Figure 4). Denote the samples on the t axis as t_p, p = 1,2,…,P and the samples on the v axis as v_q, q = 1,2,…,Q.

Only the light rays passing through the discrete sample points will be reconstructed. Similarly, the *x* and *z* axis of the captured

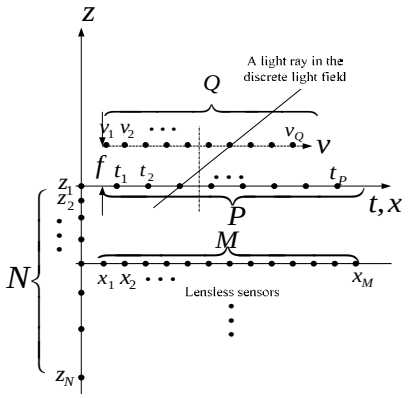

Figure 4 Discretization of the problem.

signal can also be discretized into x_m, m = 1,2,…,M and z_n, n = 1,2,…,N. Notice that discretizing the z axis simply means we will place the lensless sensors at these discrete depth values. In Figure 4, the z axis is non-uniformly discretized. This is because we want to have a uniformly discretized projection direction β, as was shown in Equation (8). The light field can therefore be written as $l(t_p, v_q)$ or l_{pq}, and the captured data can be written as $s(x_m, z_n)$ or s_{mn}. For a given captured data sample s_{mn}, we may rewrite Equation (2) in the following discrete form:

$$s_{mn} = \sum_i l_{x_m z_n}(\theta_i) g(\theta_i) \Delta\theta_i \tag{9}$$

where the input light ray directions are discretized into θ_i's, and the integration in Equation (2) becomes summation. $l_{x_m z_n}(\theta_i)$ is the intensity of the light ray passing through x_m and z_n, and has an angle θ_i. It is easy to find that such a light ray corresponds to:

$$l_{x_m z_n}(\theta_i) \Leftrightarrow l(x_m - z_n \tan\theta_i, f \tan\theta_i) \tag{10}$$

Unfortunately, $x_m - z_n \tan\theta_i$ and $f \tan\theta_i$ may not be integers. Let t_{p_1} and t_{p_2} be the two sample positions on the t axis closest to $x_m - z_n \tan\theta_i$, and v_{q_1} and v_{q_2} be the samples closest to $f \tan\theta_i$. We may derive $l(x_m - z_n \tan\theta_i, f \tan\theta_i)$ via the bilinear interpolation of $l_{p_1 q_1}$, $l_{p_1 q_2}$, $l_{p_2 q_1}$ and $l_{p_2 q_2}$.

In short, each captured data sample can always be written as a linear combination of the discrete light rays in the light field, i.e.,

$$s_{mn} = \sum_{p,q} w_{mn,pq} l_{pq} \tag{11}$$

where $w_{mn,pq}$ is the weights of the light ray l_{pq} when obtaining s_{mn}, which can be calculated based on the item $g(\theta_i)\Delta\theta_i$ in Equation (9) and the weights used during the bilinear interpolation mentioned above.

In a matrix form, If we cascade the light field samples into a $PQ \times 1$ vector $\mathbf{L}$, and the captured samples into a $MN \times 1$ vector $\mathbf{S}$. Based on Equation (11), we have the following relationship between $\mathbf{S}$ and $\mathbf{L}$:

$$\mathbf{S}_{MN\times 1} = \mathbf{W}_{MN\times PQ} \mathbf{L}_{PQ\times 1} \tag{12}$$

where $\mathbf{W}$ is the matrix associates $\mathbf{S}$ and $\mathbf{L}$. The elements in $\mathbf{W}$ represent the contributions of light field samples to the captured pixels, which are known.

Solving Equation (12) directly gives:

$$\mathbf{L} = \mathbf{W}^T \left(\mathbf{W}\mathbf{W}^T\right)^{-1} \mathbf{S}. \tag{13}$$

In practice, the dimension of the matrix $\mathbf{W}$ is often huge. Equation (13) is therefore inefficient to compute because the $\mathbf{W}$ matrix needs to be stored in memory and the matrix $\mathbf{W}\mathbf{W}^T$ need to be inverted. An iterative method is more popular, i.e., by performing:

$$\hat{l}_{pq}^{(k+1)} = \hat{l}_{pq}^{(k)} + \lambda^{(k)} w_{mn,pq} \left[\frac{s_{mn} - \sum_{p,q} w_{mn,pq} \hat{l}_{pq}^{(k)}}{\sum_{p,q} w_{mn,pq}^2} \right] \tag{14}$$

where $\hat{l}_{pq}^{(k)}$ is the k^{th} iteration value of estimated l_{pq}. $\lambda^{(k)}$ is a relaxation factor and may be chosen in the range from 0.0 to 2.0. For more details on the ART method, the readers are referred to [6][11][12][13] etc.

4. RELATED WORK

Three-dimensional imaging has important applications in machine vision, radiometry, modeling, microscopy, etc. There are two major categories of techniques widely studied, namely vision-based techniques and physical optics techniques. The former includes stereo vision [14], depth from defocus [15], image-based modeling [16] and all kinds of range sensors [17], etc. The goal of such algorithms is to reconstruct the 3D geometry of the scene object, typically with regular cameras or special devices such as laser range sensors. Examples of the physical optics techniques are confocal microscopy [18], coherence imaging [19], coherence tomography [20], etc. Usually, physical optics methods rely on the interference of lights, which cannot be performed under normal lighting conditions.

Instead of reconstructing the scene geometry, the proposed method tries to reconstruct the light field of the scene, as it has been shown that a densely sampled light field can be used to render 3D scenes without scene geometry [2]. As we show that the relationship between the captured data and the light field resembles CT, many existing algorithms can be applied to solve the reconstruction problem. In fact, tomography methods has be widely used in many other applications, such as synthetic aperture radar [21], microscopy [18], geometry reconstruction from images [22], etc.

From the point of view of signal processing, what we are capturing with a lensless camera is a filtered version of the light field. In this sense, our proposed method also resembles image restoration from blurred images [7], or the literature in super-resolution [23]. We also want to point out that the capturing method we use resembles that of coded aperture imaging (CAI) [24]. In fact, the aperture designed in CAI can be directly used in our lensless camera. The derivation in Section 3 is still valid as long as we apply the coded pattern for $g(\theta)$. The difference is that CAI reconstructs one single image about an object at a constant depth, while our method reconstructs the light field that can be used for 3D rendering.

5. EXPERIMENTAL RESULTS

Due to page limits, we show a very simple example to demonstrate the ability of the proposed scheme to reconstruct the light field. We use a 3D scene named *Duck*. Figure 5 (a) shows the setup of our experiment. The target light field is one that records all the light rays passing through the *t* axis. Compared with the

regular 4D light field in [2], our light field is 3D, which is equivalent to a set of images captured along a line (the *t* axis). The light field is divided into 64 EPIs, each of which corresponds to a plane with a certain tilt angle α. The EPIs are captured and reconstructed separately with 1D lensless cameras along their corresponding planes.

The EPIs have resolution 64×64. For each EPI, we use 32 1D lensless cameras (at 32 different depth) each with 192-pixel resolution to capture the scene and perform the reconstruction using Equation (13). The gain function of the lensless camera is assumed to be:

$$g(\theta)=\begin{cases}\cos(\theta), & |\theta|\le\dfrac{\pi}{4}\\ 0, & \text{otherwise}\end{cases}. \qquad (15)$$

Figure 5 (b) and (d) show two captured datasets along two different tilt angles. Figure 5 (c) and (e) are their reconstructed EPIs. Figure 5 (f) and (g) show two images rendered at different *t* coordinate using the reconstructed light field. The 3D effect between the two images is very obvious.

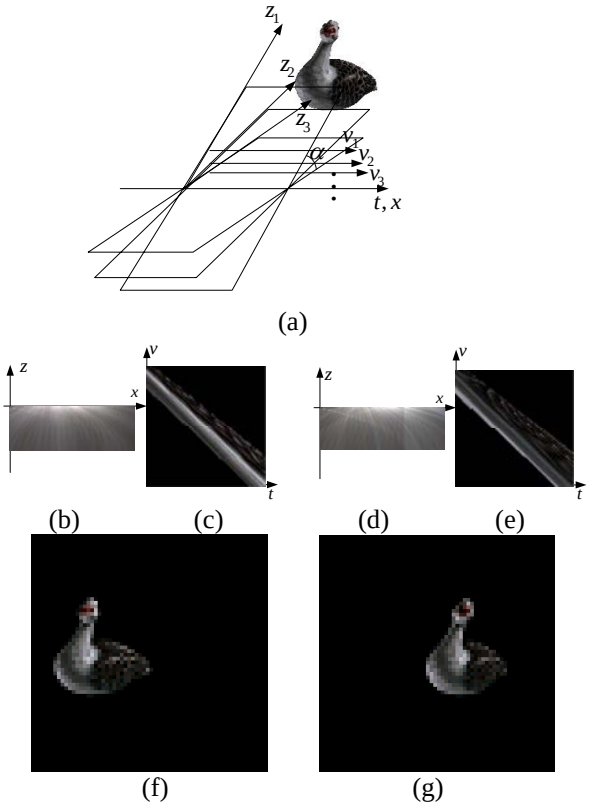

Figure 5 The test scene *duck*.

6. CONCLUSIONS AND FUTURE WORK

The major contribution of this paper was to formulize the problem of capturing 3D scenes with lensless cameras. Instead of reconstructing the scene geometry, we recovered the light field of the scene that can also be used for 3D rendering. We showed that in principle such reconstruction resembles that in CT etc, thus many existing algorithms can be applied directly in our scenario. The major advantage of using lensless cameras to capture scenes is the simplicity and the infinite depth of field.

One constraint of the proposed scheme is that due to the computational cost, we have difficulty in recovering high-resolution light fields. We are exploring the extension of some efficient reconstruction algorithms (such as filtered back projection [6]) to our problem. Using real sensors to capture scenes is also our future work.

REFERENCES

[1] C. Zhang and T. Chen, "A Survey on Image-Based Rendering - Representation, Sampling and Compression", *EURASIP Signal Processing: Image Communication*, pp. 1-28, Jan 2004.

[2] M. Levoy and P. Hanrahan, "Light field rendering", *Computer Graphics (SIGGRAPH'96)*, pp. 31, Aug. 1996.

[3] A. Isaksen, L. McMillan and S. J. Gortler, "Dynamically reparameterized light fields", *Computer Graphics (SIGGRAPH'00)*, pp. 297, July 2000.

[4] A. Kubota and K. Aizawa, "A novel image-based rendering method by linear filtering of multiple focused images acquired by a camera array", ICIP'03, Barcelona, Spain, Sep. 2003.

[5] D. A. Forsyth, J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2002.

[6] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, *IEEE Press*, 1988.

[7] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[8] R. Gordon, R. Bender and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography", *J. Theoretical Biology*, Vol. 29, pp. 471-481, 1970.

[9] B. E. Oppenheim, "Reconstruction tomography from incomplete projections", in *Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine*, M. M. Ter-Pogossian et al. Eds., Baltimore, MD, University Press, 1977.

[10] M. Nassi, W. R. Brody, B. P. Medoff and A. Macovski, "Iterative reconstruction-reprojection: an algorithm for limited data cardiac computed tomography", IEEE Trans. Biomed. Eng., Vol. BME-29, pp.331-341, May 1982.

[11] A. H. Andersen, "Algebraic reconstruction in CT from limited views", IEEE Trans. Medical Imaging, pp. 50-55, Mar. 1989.

[12] G. Herman, A. Lent, P. Lutz, "Relaxation methods for image reconstruction", *Comm. ACM*, Vol. 21, No. 2, pp.152-158, 1978.

[13] A. H. Andersen and A. C. Kak, "Simultaneous Algebraic Reconstruction Technique (SART): a superior implementation of the ART algorithm", *Ultranson. Img.*, Vol. 6, pp. 81-94, 1984.

[14] U. R. Dhond and J. K. Aggarwal, "Structure from Stereo – A Review", *IEEE Trans. Systems, Man and Cybern.*, Vol. 19, No. 6, pp. 1489-1510, 1989.

[15] J. Ens and P. Lawrence. "An investigation of methods for determining depth from focus", *IEEE Trans. on PAMI*, Vol. 15, No. 2, pp. 97-108, Feb. 1993.

[16] Z.-Y. Zhang, "Image-based geometrically-correct photorealistic scene / object modeling (IBPhM): a review", *Asian Conference on Computer Vision (ACCV'98)*, Hong Kong, Jan. 8-11, 1998.

[17] P. Besl, "Active Optical Range Imaging Sensors", *Machine Vision and Applications*, Vol. 1, pp. 127-152, 1988.

[18] T. Wilson, ed., *Confocal Microscopy*, Academic Press, 1990.

[19] D. L. Marks, R. A. Stack, D. J. Brady, D. C. Munson Jr., R. B. Brady, "Visible Cone-Beam Tomography with a lensless interferometric camera", *Science*, pp. 2164-2166, June 1999.

[20] B. E. Bouma, G. J. Tearney, *Handbook of Optical Coherence Tomography*, New York, NY, Marcel Dekker, 2002.

[21] J. P. Fitch, *Synthetic Aperture Radar*, Springer-Verlag New York Inc., 1988.

[22] D. L. Marks, R. Stack, A. J. Johnson, D. J. Brady and D. C. Munson, Jr, "Cone-beam tomography with a digital camera", *Appl. Opt.* Vol. 40, pp. 1795-1805, 2001.

[23] T. S. Huang and R. Tsai, "Multi-frame image restoration and registration", *Advances in Computer Vision and Image Processing*, Vol. 1, pp. 317-339, 1984.

[24] G.K. Skinner, "X-Ray Imaging with Coded Masks", *Scientific American*, pp. 84-89, Aug. 1988.