An Optical Analysis of Light Field Rendering

Tao Feng and Heung-Yeung Shum
Microsoft Research, China
{tfeng, hshum}@microsoft.com

Abstract

Light field rendering has been proposed to render novel images from a collection of captured images without explicit geometry information. In this report, we present an optical analysis of light field rendering. In particular, a light field rendering system can be considered as a virtual imaging system with a discrete synthetic aperture (e.g., [10]. We define the key elements of the virtual imaging system including the aperture, the circle of confusion, the depth of field, the hyperfocal distance and the imaging law of the constant depth rendering, with analogy to a conventional optical imaging system.

Based on the optical analysis of the virtual imaging system, we describe the relationship among the depth variation of the scene (depth of field), the constant depth(perfectly focused plane), the spacing of cameras (aperture) and the rendering resolution (circle of confusion). Specifically, the hyperfocal distance of the virtual optical system, as a key parameter for light field rendering, determines the relationship between the spacing of cameras and rendering resolution. Given the minimum and maximum depths of the scene, the optimal constant depth and the hyperfocal distance are derived to achieve the best rendering quality or minimum number of images. The minimum number of images required for anti-aliasing rendering (i.e., the rendering error is smaller than the circle of confusion) can be further reduced by segmenting the depth into multiple depth layers. A quantitative relationship between hyperfocal distance, number of layers, and depth variation of the scene is described.

1 Introduction

Based on the plenoptic function [1], many image-based rendering techniques have been developed to create novel images from a large number of images. Indeed, an image can be considered as a sampling of the plenoptic function [2]. By assuming continuity of the sampled ray space, novel ray through arbitrary viewpoints can be determined by the nearby rays in the captured ray database even in the absence of any geometry information of the scene. Assuming a free space and the transparency of the route of the ray, for example, light field [3] and lumigraph [4] reduce a 5D plenoptic function to a 4D light field. Concentric Mosaics system [5] is a 3D sampling of the plenoptic function along a circle path.

A continuous representation of a ray database would be sufficient to synthesize any desired view from any viewpoint in the feasible space. In practice, however, we must sample the ray space with a finite sampling density and a finite capture resolution of cameras. The success of many existing image-based rendering systems depends on a high sampling rate (over-sampling) and a large amount of image data. Special compression and decompression techniques need to be applied to make these systems useful. The bottom line is that, we have to understand how densely these images should be sampled in order to achieve an acceptable rendering quality.

Much work has been done recently on light field sampling, or the minimum number of images needed for anti-aliased light field rendering. Through a geometric analysis, for example, Lin and Shum [6] obtain the lower bound of the number of images of a light field rendering system using bilinear interpolation a constant-depth assumption. Chai *et al.* [7] analyze the relationship among scene complexity, number of image samples, and output resolution. Furthermore, based on a spectral analysis of sampled light field. The minimum sampling curve in the joint image and geometry space is also proposed by using multiple constant depth layers for the light field.

This paper takes a different approach. We argue that the light field rendering can be considered as a virtual optical imaging system. Analogous to a real optical system, we can define the optical parameters of the light field rendering system, such as the depth of field, the aperture, the circle of confusion and the hyperfocal distance of the virtual imaging system. We present an optical analysis to determine the minimum sampling rate of the light field rendering system.

Treating light field rendering system as a synthetic optical system is not entirely new. The original light field system regards the sampling interval as the aperture [3]. Aperture filtering has also been proposed. Isaksen *et al.* [10] presented a dynamically reparameterized light field by changing the synthetic aperture (i.e., the number of cameras) and the focal planes for environment. Kunita *et al.* [11] use "equivalent depth of field" to characterize the maximum acceptable depth variation in a standard constant depth light field rendering system by measuring the fidelity of the synthesized images.

In our analysis, we quantitatively describe the relationship among three key elements in light field rendering: the scene complexity, the number of images, and the output resolution. Not only can we compute the required number of images for anti-aliased light field rendering given output

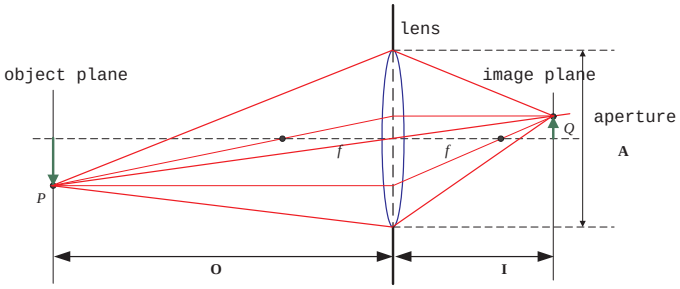


Figure 1. A conventional thin lens imaging system with focal length f and aperture A. An object P at distance O away is filmed at point Q on the image plane.

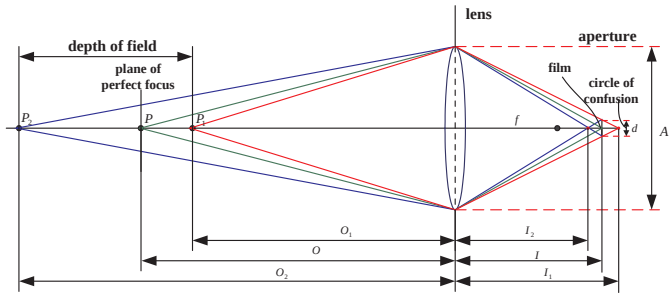


Figure 2. The depth of field of a conventional optical system is defined by the distance between points P_1 and P_2, which are filmed on the image plane with the diameter of circle of confusion d. The point P is perfectly focused.

resolution, but we also can deduce how much geometrical information is necessary if the input image number is insufficient. Therefore, the optical analysis can be applied to guide the design of many image-based rendering system.

The remainder of this paper is organized as follows. In Section 2, we discuss an ideal thin lens imaging system and introduce conventional image formation model commonly used in computer graphics and computer vision. In Section 3, We formulate the light field rendering system as a virtual optical system, and define its optical parameters. Furthermore we describe the imaging law of the constant depth rendering of the light field system. Then, we present our optical analysis and study the relationship among the elements of light field rendering system. Optimal constant depth and the minimal sampling rate are deduced in Section 4. Moreover, the optimal depth segmentation to extend the depth of field and the trade-off between the amount of geometrical information and the number of images needed is studied. We conclude our paper in Section 5.

2 Thin Lens Optical System

2.1 Ideal Thin Lens Model

Fig. 1 shows the basic image formation geometry of a conventional thin lens optical system. With the perfectly converging thin lens and the aperture diameter A, all light rays radiated from an object point P (on the object plane) that pass through the aperture are refracted by the lens to converge at the point Q on the image plane. One can view the lens imaging system as transforming each point in the scene to a single focused point behind the lens. For the thin lens, the relationship between the object distance O, focal length of the lens f, and the image distance I is described by the Gaussian lens law:

$$\frac{1}{f} = \frac{1}{O} + \frac{1}{I} \tag{1}$$

Each point on the object plane, which is not occluded, is projected onto a single point on the image plane, causing a focused image to be formed. The film (or sensor) plane must coincide with the image plane to record a sharp image. Points in front of object plane and behind object plane are not focused perfectly and therefore are distributed over a patch (blurred image) on the film plane.

2.2 Depth of field and hyperfocal distance

As shown in Fig. 2, the point displace from the object plane image as a circle of confusion ($C.o.C$) on the film. The diameter of the circle of confusion C is determined by the congruent trigles formed by the rays passing through the aperture, i.e.,

$$\frac{A}{I'} = \frac{C}{\|I - I'\|} \tag{2}$$

$$\frac{1}{f} = \frac{1}{I} + \frac{1}{O} \tag{3}$$

$$\frac{1}{f} = \frac{1}{I'} + \frac{1}{O'} \tag{4}$$

where O' and I' are the object distance and image distance of the point(In Fig. 2, either O_1 or O_2 can be O', so does I'.).

The C in Eq. 4 can be computed as

$$C = \frac{\|O' - O\|}{O'} \frac{f}{O - f} A. \tag{5}$$

In practice, films have a finite resolution. Films can't resolve details smaller than the minimum grain separation of the film emulsion (or pixel size of CCD). Points with the circle of confusion smaller than the resolution of film are "in focus".

The depth of field (DOF) is defined as the total range of in focus zone,

$$C = \frac{Af}{O - f} \frac{\|O' - O\|}{O'} \leq d \tag{6}$$

where d is the maximum acceptable circle of confusion.

The nearest and farthest points that are acceptable are found at the distances,

$$O_f = \frac{d_H O}{d_H - (O - f)} \tag{7}$$

$$O_n = \frac{d_H O}{d_H + (O - f)} \tag{8}$$

respectively, where

$$d_H = \frac{A}{d/f} \tag{9}$$

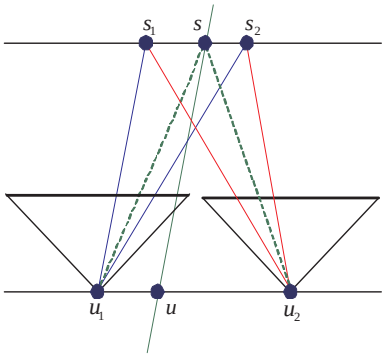


Figure 3. The ray interpolation of the light field rendering: The desired ray (u, s) is interpolated by the nearest neighboring rays (u_1, s_1),(u_1, s_2), (u_2, s_1), and (u_2, s_2).

is so-called "hyperfocal distance"([12]), or the ratio of diameter of aperture to maximum acceptable angular blur. The depth of field is then computed as

$$DOF = O_f - O_n = \frac{2d_H O(O-f)}{{d_H}^2 - (O-f)^2} \tag{10}$$

3 Light Field Rendering: An Optical Analysis

3.1 An Overview of Light Field Rendering

The sampling of the 4D plenoptic function in light field ([3]) / lumigraph ([4]) can be represented by a two-plane parameterization. A ray passing through a light slab, as specified by a line connecting a point on the UV plane and another point on the ST plane, can be uniquely determined by a quadruple (u, v, s, t). A pinhole camera is adopted to capture the light field with the center of projection located on the UV plane (camera plane).

To render a given ray, the line parameters (u, v, s, t) is computed and then the light slab is resampled and interpolated by a certain bandpass filter to reconstruct the desired ray. The 2D case of interpolation is illustrated in Fig. 3. We refer the read to the original light field paper and Lumigraph paper for more details.

For sake of simplicity, we now discuss light field rendering in 2D space. As shown in Fig. 4, cameras are aligned so that their centers of projection are located on the U axis with the same intervals D. To render novel view images, a virtual rendering camera C with infinite resolution is placed on the desired viewpoint behind the U axis at a distance I. Assume that there is only one ideal *object point* Q ([12], [6]) in the scene, which is placed Z units in front of the U axis.

3.2 Imaging law of light field rendering with constant depth

A constant depth plane that is closer to the object is selected to improve the rendering image quality. With the constant-depth assumption, all the ray captured by cameras are hypothesized emitted from points located on the constant depth plane, which is parallel to the camera plane. Fig. 4 illustrates the 2D case of light field rendering process. The

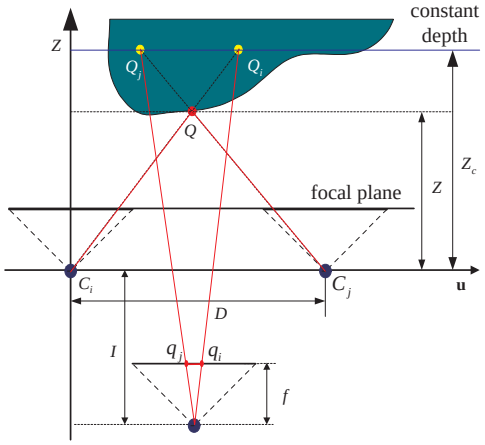


Figure 4. Light field rendering with constant depth assumption. The rendering error $|q_i - q_j|$ incurred by the constant depth assumption (Z_c) is projected by $|Q_i - Q_j|$. C_i and C_j are neighboring capturing cameras, while C at the bottom is the rendering camera.

constant depth line is defined to be parallel to the U axis. Let us denote the distance between the constant depth line and U axis by Z_c. The virtual rendering camera is located behind the U axis with distance I. The rays captured by the cameras C_i through C_j are employed in rendering. Recall that, for constant depth assumption, all rays captured by the camera is hypothesized as being emitted from a point lying on the constant depth plane. To camera C_i, the ray emitted from Q is hypothesized as being emitted from a virtual point Q_i, which is the intersection of constant depth plane and the ray (C_i, Q). Thus, on the novel view image, the ray (C_i, Q) is rendered as image point q_i, which is the image of virtual point Q_i. Similarly for the camera C_j, Q is rendered as q_j.

Accordingly, the light field rendering system with constant depth correction can be considered as a virtual optical system, whose imaging law is described above. The point q_i and q_j are called the *image point* of the object point Q. Only the points exactly on the constant plane are perfectly in focus.

In Fig. 4, from the congruent trigles $\triangle QQ_iQ_j$ and $\triangle QC_iC_j$, we have the following relationship,

$$\frac{|Q_i - Q_j|}{|C_i - C_j|} = \frac{|Z_c - Z|}{Z} \tag{11}$$

where Z_c is the constant depth, Z is the depth of the object point Q. From the congruent trigles $\triangle CQ_iQ_j$ $\triangle Cq_iq_j$, we have,

$$\frac{|Q_i - Q_j|}{|q_i - q_j|} = \frac{Z_c + I}{f} \tag{12}$$

where q_i and q_j are the virtual points on the constant depth line corresponding to cameras C_i and C_j, respectively. And f is the focal length of the virtual rendering camera C.

Substitute Eq. 12 for Eq. 11, we have,

$$|q_i - q_j| = \frac{|Z_c - Z|}{Z} \frac{f}{Z_c + I} |C_i - C_k| \tag{13}$$

Equation 13 is the basic equation of the light field rendering that describes the quantitative relationship among the key

elements of light field system and rendering image quality. From the above analysis, we can consider a light field rendering system as a virtual optical imaging system, whose imaging process depends on the constant depth assumption and rendering algorithm. Levoy and Hanrahan [3] suggested that the light field rendering system be considered as a *discrete aperture* imaging system, with the diameter of the aperture being equal to the spacing between camera locations. Similarly, discrete aperture camera has also been discussed in [11]. In this section, however, we will for the first time derive important qualitative properties of the virtual optical system of light field rendering, much like the conventional thin lens optical system.

3.3 The depth of field

As shown in previous section, in the light field rendering system, images of an ideal object point could be more than one point on the desired image plane. To alleviate the aliasing, pre-filtering and post-filtering is applied by Levoy and Hanrahan. In our analysis, it is equal to blurring the image points to avoid the double image [6]. In this subsection, we will determine the depth variation range of scene, where the rendered image is spread less than a given length for all object points. For Eq. 13, we define the acceptable rendering quality as the largest range of $|q_i - q_j|$. That is

$$|q_i - q_j| = \frac{|Z_c - Z|}{Z_c + I} f |C_i - C_j| \leq d \tag{14}$$

where d is a predefined acceptable rendering quality (rendering resolution).

We define the aperture of the virtual optical system as $A' = |C_i - C_j|$, or the distance between two consecutive cameras. The nearest and farthest depth whose diameter of blurred circle are within d are specified as two solutions of Eq. 14,

$$Z_{min} = \frac{D_H Z_c}{D_H + (I + Z_c)} \tag{15}$$

$$Z_{max} = \frac{D_H Z_c}{D_H - (I + Z_c)} \tag{16}$$

where

$$D_H = \frac{A'}{d/f} \tag{17}$$

We can easily verify that, for all Z ($Z_{min} \leq Z \leq Z_{max}$), $|q_i - q_j| \leq d$ is satisfied. Therefore, the *depth of field* (DOF) of the virtual optical system is defined as acceptable range of focus. That is,

$$Z_{DOF} = Z_{max} - Z_{min} = \frac{2 D_H Z_c (I + Z_c)}{D_H^2 - (I + Z_c)^2} \tag{18}$$

The geometrical interpretation of Eq. 18 is illustrated in Fig 5. Z_c is distance from the constant depth plane to U axis, Z_{min} and Z_{max} are the minimum and the maximum depth range for acceptable rendering quality. Note that the object points lie on Z_{min} and Z_{max} plane causes the same circle of confusion on the desired image. Fig. 6 shows the depth of field variation due to changes in D_H and Z_c.

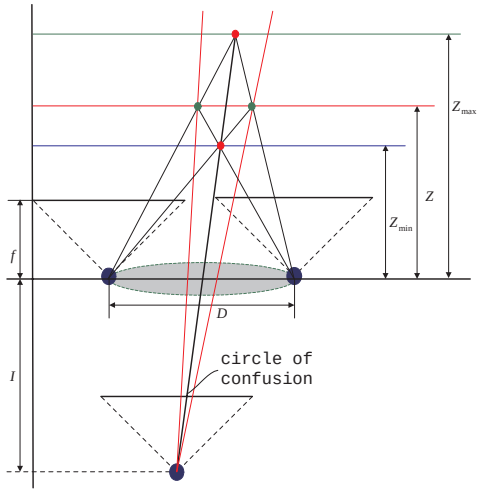
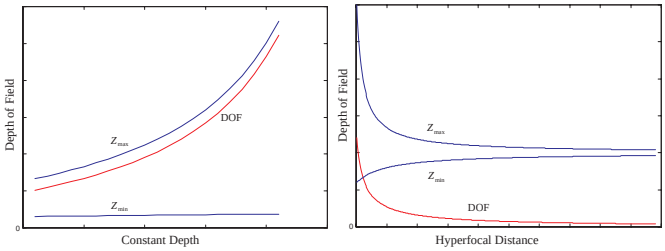


Figure 5. The acceptable range (depth of field) of light field rendering, given the tolerable rendering error (circle of confusion).



(a) (b)

Figure 6. (a) The relationship between depth of field and constant depth with given hyperfocal distance; (b) The relationship between the hyperfocal distance and the depth of field with a fixed constant depth.

3.4 When the rendering camera is on the UV plane

A special configuration of light field rendering is of particular interest, when the rendering camera is located on the same plane with the capturing cameras (UV plane), as shown in Fig. 7. Substituting $I = 0$ for Eq. 13, we have,

$$|q_i - q_j| = \frac{|Z_c - Z|}{Z} \frac{f}{Z_c} |C_i - C_k| \tag{19}$$

As illustrated in Fig. 7, we assign the image distance is the focus length f of the rendering camera C, the object distance is the constant depth Z_c. Then an virtual focus length f' of the optical system is obtained by the Gaussian law (Eq. 1),

$$f' = \frac{Z_c f}{Z_c + f} \tag{20}$$

Substituting Eq. 20 for Eq. 19, we have

$$|q_i - q_j| = \frac{f'}{Z_c - f'} \frac{|Z_c - Z|}{Z} |C_i - C_j| \tag{21}$$

Notice that Eq. 6 has exactly the same form as that of Eq. 21. Therefore, *when the rendering camera is located on the UV plane, a light field rendering system with constant depth correction can be regarded as an ideal thin lens imaging system.*

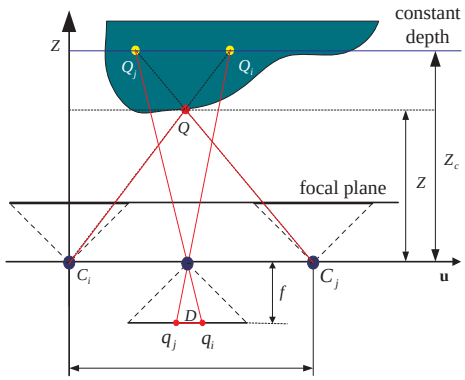


Figure 7. When the rendering camera is located on the capturing camera plane, the light fielding rendering system becomes an ideal thin lens optical system. Objects lying on the constant plane are perfectly focused on the image plan. The focal length of the virtual lens is $f' = fz_c/(f + z_c)$.

We can also verify that the depth of field has the same form as that of Eq. 10.

$$Z'_{DOF} = \frac{2D_H Z_c^2}{D_H^2 - Z_c^2} = \frac{2z_c^2 A'f/d}{(A'f/d)^2 - z_c^2} = \frac{2d'_H z_c(z_c - f')}{d'^2_H - (z_c - f')^2} \tag{22}$$

where $d'_H = A'f'/d$.

In summary, analog to the conventional optical system, the key elements of light field can be defined in term of a virtual optical system:

- Aperture. $A' = |C_i - C_j|$
- Focal length. $f' = \frac{fz_c}{f+z_c}$
- Hyperfocal distance. $d'_H = A'f'/d$
- Depth of field. $z'_{DOF} = \frac{2d'_H z_c(z_c - f')}{d'^2_H - (z_c - f')^2}$

In the virtual optical system, the aperture is defined as the intervals of cameras. Object distance is the distance between object point and UV plane. Image distance is the distance between rendering image plane and UV plane. Focal length is defined as a virtual value decided by the object distance and the image distance.

The hyperfocal distance plays an important role in minimum sampling of light field rendering because it describes the relationship between spacing between cameras (the number of cameras) needed for capturing and the rendering quality (rendering resolution). Given rendering resolution, the higher D_H means fewer raw images needed for anti-aliasing rendering. For a given hyperfocal distance, the relationship between the spacing of camera and rendering resolution d is illustrated in Fig. 8.

4 Minimum Sampling of Light Field

4.1 The optimal constant depth

Given the minimum depth Z_{min} and maximum depth Z_{max} of a scene, we want to determine the optimal constant depth Z_{opt} for minimum sampling rate or best rendering quality. By the definition of hyperfocal distance, it is equivalent to

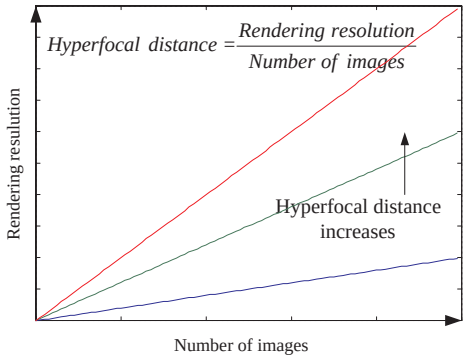


Figure 8. The hyperfocal distance is proportional to the rendering resolution, and inversely proportional to the number of images used. Given a fixed hyperfocal distance, the relationship between rendering resolution and the number of images is linear.

the maximizing the hyperfocal distance d'_H or D_H. The optimal constant depth Z_{opt} satisfies:

$$\begin{aligned} Z_{opt} &= arg\max_z\{D_H\} \\ s.t. \quad & Z_{min} \geq \tfrac{D_H Z}{D_H + Z} \\ & Z_{max} \leq \tfrac{D_H Z}{D_H - Z} \end{aligned} \tag{23}$$

The maximum D_H arrives when

$$\frac{1}{Z_{opt}} = \frac{1}{2}\left[\frac{1}{Z_{min}} + \frac{1}{Z_{max}}\right] \tag{24}$$

$$\frac{1}{D_H} = \frac{1}{2}\left[\frac{1}{Z_{min}} - \frac{1}{Z_{max}}\right] \tag{25}$$

The geometrical interpretation of Z_{opt} is that the optimal constant depth is the harmonic mean of the minimum and maximum depths. The maximum D_H is decided by Eq. 25, which means the relationship between the spacing of the cameras and output quality is determined by the scene depth variation. In other words, this equation completely determines the minimum sampling of light field rendering.

4.2 Multiple depth layers segmentation

From the above analysis, we know the hyperfocal distance is specified by the maximum and minimum depth of the scene. We understand that, the larger the depth of field, the more input images needed for anti-aliased light field rendering. More images mean more storage and less efficient rendering. Therefore, segmenting the depth of the scene into multiple depth layers effectively decreases the depth of field and decreases the number of cameras needed for rendering. The number of layers is determined by the tradeoff between rendering efficiency and the storage of image data.

Fig. 9 illustrates how the scene is recursively segmented into multiple layers. From Eq. 16, we get the following relationship between the layers

$$\begin{cases} \frac{1}{Z_{max}(n)} = \frac{D_H - I}{D_H}\frac{1}{Z_{opt}(n)} - \frac{1}{D_H} \\ \frac{1}{Z_{min}(n)} = \frac{D_H + I}{D_H}\frac{1}{Z_{opt}(n)} + \frac{1}{D_H} \\ Z_{max}(n) = Z_{min}(n+1) \end{cases} \tag{26}$$

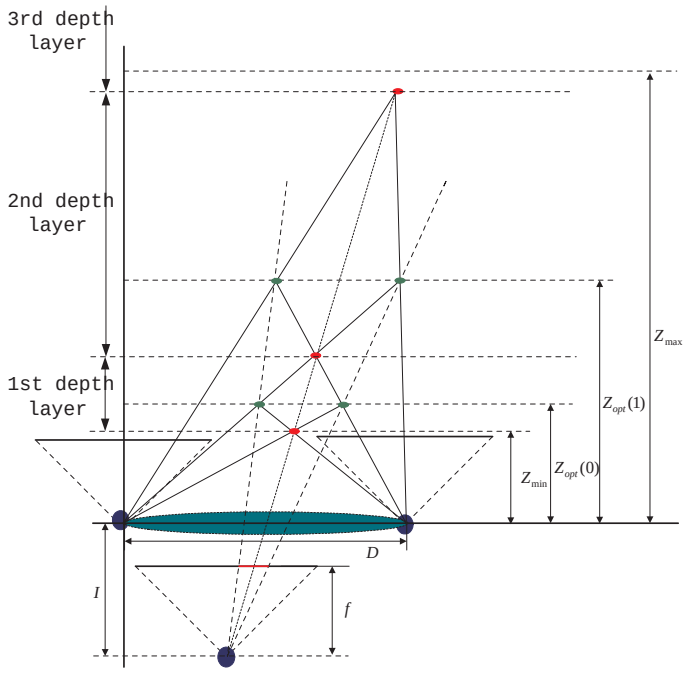


Figure 9. Multiple layers of constant depth are recursively constructed for minimum sampling of light field rendering. Given the minimum and maximum depths of a scene, and the rendering resolution, the number of depth layers needed for anti-aliased light field rendering is computed recursively.

where $Z_{opt}(n)$, $Z_{min}(n)$, and $Z_{max}(n)$ are the optimal constant depth, the minimum depth and the maximum depth of the n^{th} layer, respectively.

From Eq. 26 we have,

$$\frac{1}{Z_{opt}(n)} = (\frac{D_H - I}{D_H + I})^n (\frac{1}{Z_{opt}(0)} + \frac{1}{I}) \tag{27}$$

A special case of Eq. 27 is that $I = 0$. When the rendering camera is located on the U axis, we get the following equations.

$$\frac{1}{Z_{opt}(n)} = \frac{n}{2}[\frac{1}{Z_{min}} + \frac{1}{Z_{max}}] \tag{28}$$

$$\frac{1}{D_H} = \frac{1}{N+1}[\frac{1}{Z_{min}} - \frac{1}{Z_{max}}] \tag{29}$$

The above equations determine the minimum sampling in the joint image and geometry space. Specifically, the minimum sampling problem in the joint image and geometry space is described by the relationship among the number of images, the output resolution and the number of constant depth layers. Depth layer segmentation effectively increases the hyperfocal distance.

5 Discussion and Conclusion

In this paper, we have presented an optical analysis of the field rendering system. We regard the light field rendering as a synthetic aperture optical system with a constant depth assumption for the scene. We define the aperture, the circle of confusion, the depth of field and the hyperfocal distance of the light field rendering system. From an optical analysis of light field sampling, we obtain the relationship among the depth variation of the scene (depth of field), the number of images (aperture) and the rendering resolution (circle of confusion). Specifically, this relationship is completely described by the hyperfocal distance of the virtual optical system. The optical analysis is applied to estimate the optimal constant depth for a given scene for the best rendering quality. The minimum sampling rate is then derived. To extend our method to cover much larger depth variation, we give the optimal depth segmentation using much fewer number of images, without loss of rendering quality. It is worth to mention that we get similar but more general results compared to the geometric analysis of [6], and the spectral analysis of [7], in particular, when the rendering camera is not on the capturing camera plane. In the current optical analysis, the occlusion problem is not discussed. Occlusion is an interesting but difficult problem for our future study in plenoptic sampling. Another interesting research direction is to study the depth recovery problem by the virtual synthetic aperture camera.

References

[1] E. H. Adelson and J. Bergen. “The plenoptic function and the elements of early vision”, In Computational Models of Visual Processing, pages 3-20. MIT Press, Cambridge, MA, 1991

[2] L. McMillan and G. Bishop. “Plenoptic Modeling: An Image-Based Rendering System”. In Proceedings of SIGGRAPH 95, pp. 39-46, 1995.

[3] M. Levoy and P. Hanrahan. “Light Field Rendering”, In Proceedings of SIGGRAPH 96, pp. 31-42.

[4] Gortler, Steven J. , Radek Grzeszczuk,Richard Szeliski and Michael F. Cohen. “The Lumigraph”, Computer Graphics, Proceedings of SIGGRAPH 96, August 4-9, 1996, pp. 43-54.

[5] Heung-Yeung Shum and Li-Wei He. “Rendering with Concentric Mosaics”. In Proceedings of SIGGRAPH 1999.

[6] Z. Lin and Hueng-Yeung Shum. “On The Number of Samples Needed In Light Field Rendering with Constant Depth Assumption”, Accepted by CVPR’2000.

[7] J. Chai, X. Tong, Shing-Chow Chan and Heung-Yeung Shum. “Plenoptic Sampling”, Accepted by Siggraph’2000, New Orleans, July, 2000.

[8] J. Shade, Steven Gortler, Li Wei He, and Richard Szeliski. “Layered Depth Images”, In Proceedings of SIGGRAPH 1998.

[9] Chun-Fa Chang, Gary Bishop and Anselmo Lastra. “LDI Tree: A Hierarchical Representation for Image-Based Rendering”, In Proceedings of SIGGRAPH 1999.

[10] A. Isaksen, L. McMillan and S. Gortler. “Dynamically reparameterized light fields”, Technical report, Technical Report MIT-LCS-TR-778, May 1999.

[11] Y. Kunita, M. Inami, T. Maeda, and S. Tachi. “Real-time rendering system of moving objects”, In Proceedings of the 1999 IEEE Workshop on Multi-View Modeling & Analysis of Visual Scenes (MVIEW’99), pages 81-88, 1999.

[12] M. Bass, *et.al.*, *Handbook of optics*, 2nd ed. McGraw-Hill. New York, 1995.

[13] W. J. Smith, *et al.*, *Modern Optical Engineering*, McGraw-Hill. New York, 1992.