Plenoptic Sampling

Jin-Xiang Chai* Xin Tong Shing-Chow Chan† Heung-Yeung Shum‡
Microsoft Research, China

Abstract

This paper studies the problem of **plenoptic sampling** in image-based rendering (IBR). From a spectral analysis of light field signals and using the sampling theorem, we mathematically derive the analytical functions to determine the **minimum sampling rate** for light field rendering. The spectral support of a light field signal is bounded by the minimum and maximum depths only, no matter how complicated the spectral support might be because of depth variations in the scene. The minimum sampling rate for light field rendering is obtained by compacting the replicas of the spectral support of the sampled light field within the smallest interval. Given the minimum and maximum depths, a reconstruction filter with an optimal and constant depth can be designed to achieve anti-aliased light field rendering.

Plenoptic sampling goes beyond the minimum number of images needed for anti-aliased light field rendering. More significantly, it utilizes the scene depth information to determine the **minimum sampling curve** in the joint image and geometry space. The minimum sampling curve quantitatively describes the relationship among three key elements in IBR systems: scene complexity (geometrical and textural information), the number of image samples, and the output resolution. Therefore, plenoptic sampling bridges the gap between image-based rendering and traditional geometry-based rendering. Experimental results demonstrate the effectiveness of our approach.

Keywords: sampling, plenoptic sampling, spectral analysis, plenoptic functions, image-based rendering.

1 Introduction

Image-based modeling and rendering techniques have recently received much attention as a powerful alternative to traditional geometry-based techniques for image synthesis. Instead of geometrical primitives, a collection of sample images are used to render novel views. Previous work on image-based rendering (IBR) reveals a continuum of image-based representations [15, 14] based on the tradeoff between how many input images are needed and how much is known about the scene geometry.

At one end, traditional texture mapping relies on very accurate geometrical models but only a few images. In an image-based rendering system with depth maps, such as 3D warping [18], view interpolation [7], view morphing [21] and layered-depth images (LDI) [22], LDI tree [6], etc., the model consists of a set of images of a scene and their associated depth maps. When depth is available for every point in an image, the image can be rendered from any nearby point of view by projecting the pixels of the image to their proper 3D locations and re-projecting them onto a new picture.

At the other end, light field rendering uses many images but does not require any geometrical information. Light field rendering [16] generates a new image of a scene by appropriately filtering and interpolating a pre-acquired set of samples. Lumigraph [10] is similar to light field rendering but it applies approximated geometry to compensate for non-uniform sampling in order to improve rendering performance. Unlike light field and Lumigraph where cameras are placed on a two-dimensional manifold, Concentric Mosaics system [23] reduces the amount of data by only capturing a sequence of images along a circular path. Light field rendering, however, typically relies on oversampling to counter undesirable aliasing effects in output display. Oversampling means more intensive data acquisition, more storage, and more redundancy. To date, little research has been done on determining the lower bound or the minimum number of samples needed for light field rendering.

Sampling analysis in IBR is a difficult problem because it involves the complex relationship among three elements: the depth and texture information of the scene, the number of sample images, and the rendering resolution. The topic of prefiltering a light field has been explored in [16]. Similar filtering process has been previously discussed by Halle [11] in the context of Holographic stereograms. A parameterization for more uniform sampling [4] has also been proposed. From an initially undersampled Lumigraph, new views can be adaptively acquired if the rendering quality can be improved [20]. An opposite approach is to start with an oversampled light field, and to cull an input view if it can be predicted by its neighboring frames [12, 24]. Using a geometrical approach and without considering textural information of the scene, Lin and Shum [17] recently studied the number of samples needed in light field rendering with constant depth assumption and bilinear interpolation. However, a mathematical framework has not been fully developed for studying the sampling problems in IBR.

In this paper, we study *plenoptic sampling*, or how many samples are needed for plenoptic modeling [19, 1]. Plenoptic sampling can be stated as:

> *How many samples of the plenoptic function (e.g., from a 4D light field) and how much geometrical and textural information are needed to generate a continuous representation of the plenoptic function?*

Specifically, our objective in this paper is to tackle the following two problems under plenoptic sampling, with and without geometrical information:

- Minimum sampling rate for light field rendering;
- Minimum sampling curve in joint image and geometry space.

We formulate the sampling analysis as a high dimensional signal processing problem. In our analysis, we assume Lambertian surfaces and uniform sampling geometry or lattice for the light field.

*Currently at Carnegie Mellon University. jchai@cs.cmu.edu
†Visiting from University of Hong Kong. scchan@eee.hku.hk
‡{xtong,hshum}@microsoft.com

SIGGRAPH 2000, New Orleans, LA USA
© ACM 2000 1-58113-208-5/00/07 ...$5.00

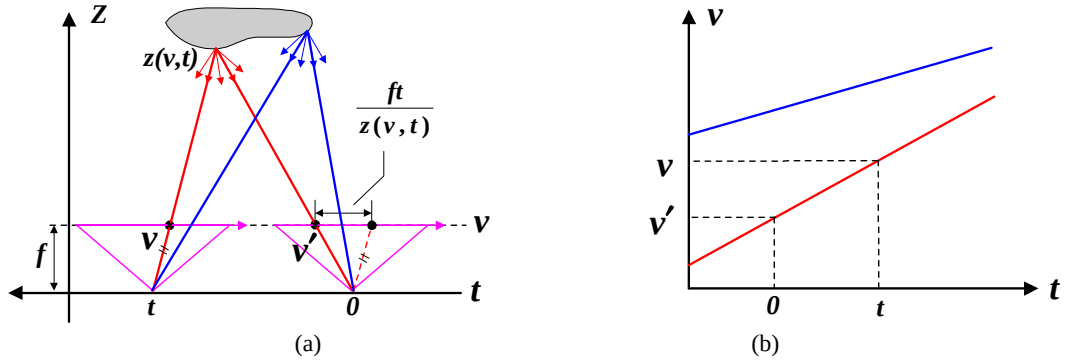


Figure 1: An illustration of 2D light field or EPI: (a) a point is observed by two cameras 0 and t; (b) two lines are formed by stacking pixels captured along the camera path. Each line has a uniform color because of Lambertian assumption on object surfaces.

Rather than attempting to obtain a closed-form general solution to the $4D$ light field spectral analysis, we only analyze the bounds of the spectral support of the light field signals. A key analysis to be presented in this paper is that the spectral support of a light field signal is bounded by only the minimum and maximum depths, irrespective of how complicated the spectral support might be because of depth variations in the scene. Given the minimum and maximum depths, a reconstruction filter with an optimal and constant depth can be designed to achieve anti-aliased light field rendering.

The minimum sampling rate of light field rendering is obtained by compacting the replicas of the spectral support of the sampled light field within the smallest interval without any overlap. Using more depth information, plenoptic sampling in the joint image and geometry space allows us to greatly reduce the number of images needed. In fact, the relationship between the number of images and the geometrical information under a given rendering resolution can be quantitatively described by a minimum sampling curve. This minimal sampling curve serves as the design principles for IBR systems. Furthermore, it bridges the gap between image-based rendering and traditional geometry-based rendering.

Our approach is inspired by the work on motion compensation filter in the area of digital video processing, in which depth information has been incorporated into the design of the optimal motion compensation filter [25, 9]. In digital video processing, global constant depth and arbitrary motion are considered for both static and dynamic scenes, whereas in our work, we analyze static scenes with an arbitrary geometry and with uniformly sampled camera setups.

The remainder of this paper is organized as follows. In Section 2, a spectral analysis of $4D$ light field is introduced and the bounds of its spectral support are determined. From these bounds, the minimum sampling rate for light field rendering can be derived analytically. Plenoptic sampling in the joint image and geometry space is studied in Section 3. The minimum sampling curves are deduced with accurate and approximated depths. Experimental results are presented in Section 4. Finally we conclude our paper in Section 5.

2 Spectral analysis of light field

2.1 Light field representation

We begin by briefly reviewing the properties of light field representation. We will follow the notations in the Lumigraph paper [10]. In the standard two-plane ray database 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). We refer the reader to Figure 2(a) of [10] for more details.

A two dimensional subspace given by fixed values of s and t resembles an image, whereas fixed values of u and v give a hypothetical radiance function. Fixing t and v gives rise to an epipolar image, or EPI [3]. An example of a 2D light field or EPI is shown in Figure 1. Note that in our analysis we define (u,v) in the local coordinates of (s,t), unlike in conventional light field where (u,v,s,t) are defined in a global coordinate system.

Assume the sample intervals along s and t directions be Δs and Δt, respectively, the horizontal and vertical disparities between two grid cameras in the (s,t) plane are determined by $k_1\Delta sf/z$ and $k_2\Delta tf/z$, respectively, where f denotes the focal length of the camera, z is the depth value and $(k_1\Delta s, k_2\Delta t)$ is the sample interval between two grid points (s,t).

Similarly, we assume that the sample intervals along u and v directions be Δu and Δv, respectively. A pinhole camera model is adopted to capture the light field. What a camera sees is a blurred version of the plenoptic function because of finite camera resolution. A pixel value is a weighted integral of the illumination of the light arriving at the camera plane, or the convolution of the plenoptic function with a low-pass filter.

2.2 A framework for light field reconstruction

Let $l(u,v,s,t)$ represent the continuous light field, $p(u,v,s,t)$ the sampling pattern in light field, $r(u,v,s,t)$ the combined filtering and interpolating low-pass filter, and $i(u,v,s,t)$ the output image after reconstruction. Let L, P, R and I represent their corresponding spectra, respectively. In the spatial domain, the light field reconstruction can be computed as

$$i(u,v,s,t) = r(u,v,s,t) * [l(u,v,s,t)p(u,v,s,t)] \quad (1)$$

where $*$ represents the convolution operation.

In the frequency domain, we have

$$I(\Omega_u,\Omega_v,\Omega_s,\Omega_t) = R(\Omega_u,\Omega_v,\Omega_s,\Omega_t)(L(\Omega_u,\Omega_v,\Omega_s,\Omega_t) * P(\Omega_u,\Omega_v,\Omega_s,\Omega_t)) \quad (2)$$

The problem of light field reconstruction is to find a reconstruction filter $r(u,v,s,t)$ for anti-aliased light field rendering, given the sampled light field signals.

2.3 Spectral support of light fields

In this section, we will introduce the spectral supports of continuous light field $L(\Omega_u,\Omega_v,\Omega_s,\Omega_t)$ and sampled light field $L(\Omega_u,\Omega_v,\Omega_s,\Omega_t) * P(\Omega_u,\Omega_v,\Omega_s,\Omega_t)$.

2.3.1 Spectral support of continuous light field

We assume that the depth function of the scene is equal to $z(u,v,s,t)$. As shown in Figure 1(a), the same 3D point is observed at v' and v in the local coordinate systems of cameras 0 and

t, respectively. The disparity between the two image coordinates can be computed easily as $v - v' = ft/z$. Figure 1(b) shows an EPI image where each line represents the radiance observed from different cameras. For simplicity of analysis, the BRDF model of a real scene is assumed to be Lambertian. Therefore, each line in Figure 1(b) has a uniform color.

Therefore, the radiance received at the camera position (s, t) is given by

$$l(u, v, s, t) = l(u - \frac{fs}{z(u, v, s, t)}, v - \frac{ft}{z(u, v, s, t)}, 0, 0)$$

and its Fourier transform is

$$L(\Omega_u, \Omega_v, \Omega_s, \Omega_t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l(u, v, s, t)e^{-j\Omega^T \mathbf{x}}d\mathbf{x} \; e^{-j(\Omega_s s + \Omega_t t)}dsdt \quad (3)$$

where $x^T = [u, v]$ and $\Omega^T = [\Omega_u, \Omega_v]$.

However, computing the Fourier transform (3) is very complicated, and we will not go into the details of its derivation in this paper. Instead, we will analyze the bounds of the spectral support of light fields. Also for simplicity, it is assumed that samples of the light field are taken over the commonly used rectangular sampling lattice.

2.3.2 Spectral support of sampled light field

Using the rectangular sampling lattice, the sampled light field $l_s(u, v, s, t)$ is represented by

$$l_s(u, v, s, t) = l(u, v, s, t) \sum_{n_1, n_2, k_1, k_2 \in Z} \delta(u - n_1\Delta u)\delta(v - n_2\Delta v)\delta(s - k_1\Delta s)\delta(t - k_2\Delta t) \quad (4)$$

and its Fourier transform is

$$L_s(\Omega_u, \Omega_v, \Omega_s, \Omega_t) = \sum_{m_1, m_2, l_1, l_2 \in Z} L(\Omega_u - \frac{2\pi m_1}{\Delta u}, \Omega_v - \frac{2\pi m_2}{\Delta v}, \Omega_s - \frac{2\pi l_1}{\Delta s}, \Omega_t - \frac{2\pi l_2}{\Delta t}) \quad (5)$$

The above equation indicates that $L_s(\Omega_u, \Omega_v, \Omega_s, \Omega_t)$ consists of replicas of $L(\Omega_u, \Omega_v, \Omega_s, \Omega_t)$, shifted to the $4D$ grid points

$$(2\pi m_1/\Delta u, 2\pi m_2/\Delta v, 2\pi l_1/\Delta s, 2\pi l_2/\Delta t),$$

where $m_1, m_2, l_1, l_2 \in Z$, and Z is the set of integers.

These shifted spectra, or replicas, except the original one at $m_1 = m_2 = l_1 = l_2 = 0$, are called the alias components. When L is not bandlimited outside the Nyquist frequencies, some replicas will overlap with the others, creating aliasing artifacts.

In general, there are two ways to combat aliasing effects in output display when we render a novel image. First, we can increase the sampling rate. The higher the sampling rate, the less the aliasing effects. Indeed, uniform oversampling has been consistently employed in many IBR systems to avoid undesirable aliasing effects. However, oversampling means more effort in data acquisition and requires more storage. Though redundancy in the oversampled image database can be partially eliminated by compression, excessive samples are always wasteful.

Second, light field signals can also be made bandlimited by filtering with an appropriate filter kernel. Similar filtering has to be performed to remove the overlapping of alias components during reconstruction or rendering. The design of such a kernel is, however, related to the depth of the scene. Previous work on Lumigraph shows that approximate depth correction can significantly improve the interpolation results. The questions are: is there an optimal filter? Given the number of samples captured, how accurately should the depth be recovered? Similarly, given the depth information one can recover, how many samples can be removed from the original input?

2.4 Analysis of bounds in spectral support

2.4.1 A model of global constant depth

Let us first consider the simplest scene model in which every point is at a constant depth (z_0). The first frame is chosen as the reference frame, and $l(u, v, 0, 0)$ denotes the 2D intensity distribution within the reference frame. The 4D Fourier transform of the light field signal $l(u, v, s, t)$ with constant depth is

$$\begin{aligned} L(\Omega_u, \Omega_v, \Omega_s, \Omega_t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l(u, v, 0, 0)e^{-j(\Omega_u u + \Omega_v v)}dudv \\ &\int_{-\infty}^{\infty} e^{-j(\frac{f}{z_0}\Omega_u + \Omega_s)s}ds \int_{-\infty}^{\infty} e^{-j(\frac{f}{z_0}\Omega_v + \Omega_t)t}dt \\ &= 4\pi^2 L'(\Omega_u, \Omega_v)\delta(\frac{f}{z_0}\Omega_u + \Omega_s)\delta(\frac{f}{z_0}\Omega_v + \Omega_t) \end{aligned}$$

where $L'(\Omega_u, \Omega_v)$ is the 2D Fourier transform of continuous signal $l(u, v, 0, 0)$ and $\delta(\cdot)$ is the 1D Dirac delta function. To keep notation, representations and illustration simple, the following discussion will focus on the projection of the support of $L(\Omega_u, \Omega_v, \Omega_s, \Omega_t)$ onto the (Ω_v, Ω_t) plane, which is denoted by $L(\Omega_v, \Omega_t)$.

Under the constant depth model, the spectral support of the continuous light field signal $L(\Omega_v, \Omega_t)$ is defined by a line $\Omega_v f/z_0 + \Omega_t = 0$, as shown in Figure 2(b). The spectral support of the corresponding sampled light field signals is shown in Figure 2(c). Note that, due to sampling, replicas of $L(\Omega_v, \Omega_t)$ appear at intervals $2\pi m_2/\Delta v$ and $2\pi l_2/\Delta t$ in the Ω_v and Ω_t directions, respectively.

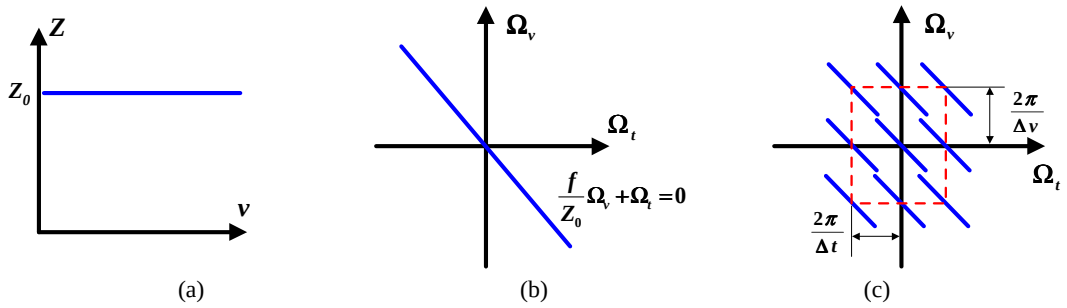
Figure 6(a) shows a constant depth scene (a1), its EPI image (a2), and the Fourier transform of the EPI (a3). As expected, the spectral support is a straight line.[1]

2.4.2 Spatially varying depth model

Now it is straightforward to observe that any scene with a depth between the minimum z_{min} and the maximum z_{max} will have its continuous spectral support bounded in the frequency domain, by two lines $\Omega_v f/z_{min} + \Omega_t = 0$ and $\Omega_v f/z_{max} + \Omega_t = 0$. Figure 6(b3) shows the spectral support when two planes with constant depths are in the scene. Adding another tilted plane in between (Figure 6(c1)) results in no variations in the bounds of the spectral support, even though the resulting spectral support (Figure 6(c3)) differs significantly from that in Figure 6(c2). This is further illustrated when a curved surface is inserted in between two original planes, as shown in Figure 6(d1). Even though the spectral supports differ significantly, Figures 6(b3), (c3) and (d3) all have the same bounds.

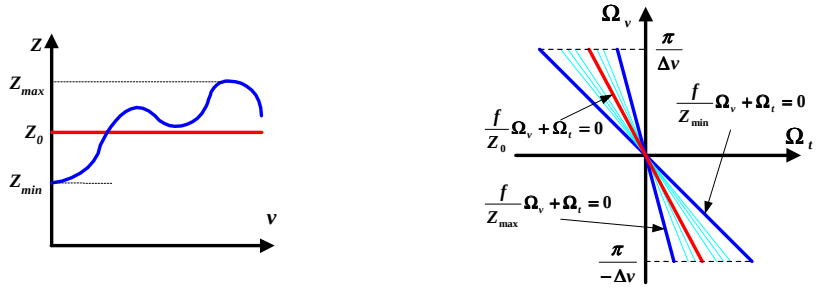
Another important observation is that geometrical information can help to reduce the bounds of the spectral support in the frequency domain. As will be illustrated in the following section, the optimal reconstruction filter is determined precisely by the bounds of the spectral support. And these bounds are functions of the minimum and maximum depths of the scene. If some information on the scene geometry is known, we can decompose the scene geometry into a collection of constant depth models on a block-by-block basis. Each model will have a much tighter bound than the original model. How tight the bound is will depend on the accuracy

[1] The ringing effect in the vicinity of the horizontal and vertical axes is caused by convolving with $\sin(\Omega_v)/\Omega_v$ because of the rectangular image boundary.



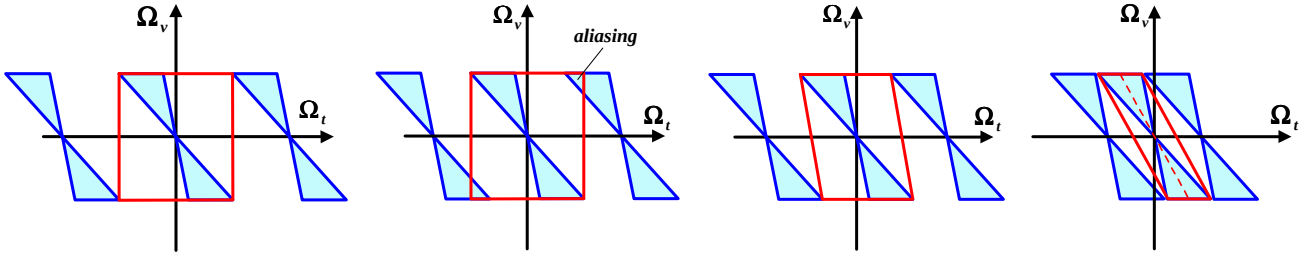
(a) (b) (c)

Figure 2: Spectral support of light field signals with constant depth: (a) a model of constant depth; (b) the spectral support of continuous light field signals; (c) the spectral support of sampled light field signals.



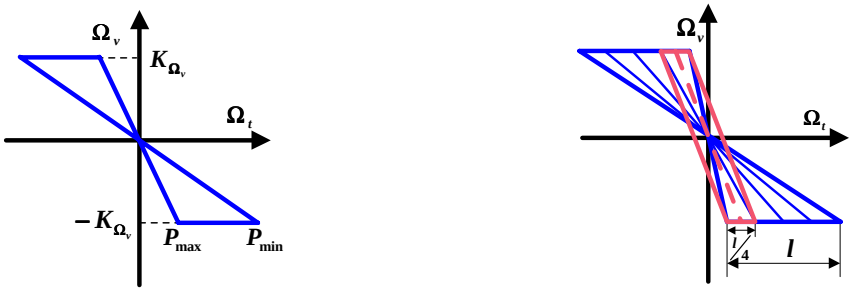
(a) (b)

Figure 3: Spectral support for light field signal with spatially varying depths: (a) a local constant depth model bounded by z_{min} and z_{max} is augmented with another depth value z_0; (b)spectral support is now bounded by two smaller regions, with the introduction of the new line of z_0.



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

Figure 4: Three reconstruction filters with different constant depths: (a) infinite depth; (b) infinite depth (aliasing occurs); (c) maximum depth; (d) optimal depth at z_c.



(a) (b)

Figure 5: (a) The smallest interval that replicas can be packed without any overlap is $P_{max}P_{min}$, determined by the highest frequency K_{Ω_v}. (b) A spectral support decomposed into multiple layers.

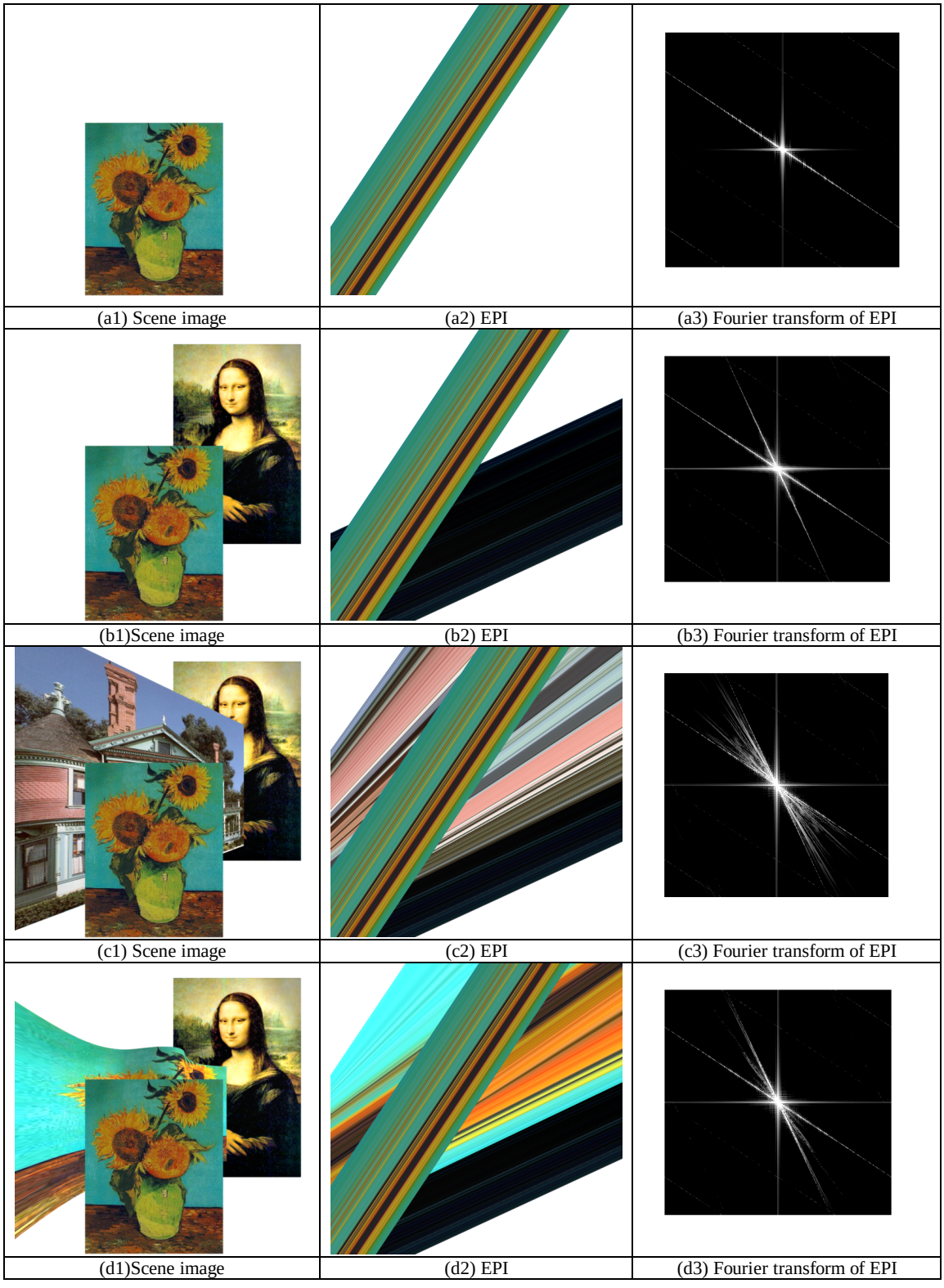


Figure 6: Spectral support of a 2D light field: (a) a single plane; (b) two planes; (c) a third and tilted plane in between; (d) a curved surface in between.

of the geometry. Figure 3 illustrates the reduction in bounds, from $[z_{min}, z_{max}]$ to $\max([z_{min}, z_0], [z_0, z_{max}])$, with the introduction of another layer.

2.4.3 A model with truncating windows

Because of the linearity of the Fourier transform, the spectral support of the EPI image for a scene with two constant planes will be two straight lines. However, this statement is true only if these two planes do not occlude each other. For synthetic environments, we can construct such EPI images on different layers by simply ignoring the occlusion.

In practice, we can represent a complicated environment using a model with truncating windows. For example, we can approximate an environment using truncated and piece-wise constant depth segments. Specifically, suppose the depth can be partitioned as

$$z(v) = z_i, \text{ for } v_i \leq v < v_{i+1}, \ i = 1, \cdots, N_d$$

where v_1 and v_{N_d+1} are the smallest and largest v of interest respectively. Then

$$l(v,t) = l_i(v - ft/z_i, 0), \text{ if } v_i \leq v < v_{i+1}$$

and

$$\begin{aligned} L(\Omega_v, \Omega_t) &= \sum_{i=1}^{N_d} \exp(-j\frac{v_i + v_{i+1}}{2}(\Omega_v + \Omega_t z_i/f)) \\ &\quad \frac{2\sin(\frac{v_{i+1}-v_i}{2}(\Omega_v + \Omega_t z_i/f))}{f\Omega_v/z_i + \Omega_t} L_i(-\Omega_t z_i/f) \\ &\equiv \sum_{i=1}^{N_d} Q_i(\Omega_v, \Omega_t) \end{aligned} \quad (6)$$

where L_i is the $1D$ Fourier transform of l_i.

In (6), because the function $\frac{\sin x}{x}$ decays fast, and $L_i(-\Omega_t z_i/f)$ also decreases fast when $|\Omega_t|$ grows, the spectral support of $Q_i(\Omega_v, \Omega_t)$ will look like a narrow ellipse. Nevertheless, because of high frequency leak, cut-off frequency should be used in the sampling analysis.

An example of two constant planes in an environment is shown in Figures 6(b1) (original image), 6(b2) (EPI) and 6(b3) (spectral support). Note that the shape of each of the two spectral supports, i.e., two approximated lines, is not significantly affected by occlusion because the width of each spectral support is not too large.

2.5 A reconstruction filter using a constant depth

Given a constant depth, a reconstruction filter can be designed. Figure 4 illustrates four different designs of reconstruction filters oriented to different constant depths. Aliasing occurs when replicas overlap with the reconstruction filters in the frequency domain (Ω_t and Ω_v), as shown in Figure 4(a)(b)(d). Anti-aliased light field rendering can be achieved by applying the optimal filter as shown in Figure 4(c), where the optimal constant depth is defined as the inverse of average disparity d_c, i.e.,

$$d_c = \frac{1}{z_c} = (\frac{1}{z_{min}} + \frac{1}{z_{max}})/2.$$

Figure 7 shows the effect of applying reconstruction filters with different constant depths. As we sweep through the object with a constant depth plane, the aliasing effect is the worst at the minimum and maximum depths. The best rendering quality is obtained at the optimal depth (Figure 7(b)), not at the focal plane as has been commonly assumed in light field [16] or Lumigraph [10] rendering. In fact, the optimal depth can be used as a guidance for selecting the focal plane. For comparison, we also show the rendering result using average depth in Figure 7(c).

Similar sweeping effects have also been discussed in the dynamically reparameterized light field [13]. However, an analytical solution using the minimum and maximum depths has never been presented before.

2.6 Minimum sampling rate for light field rendering

With the above theoretical analysis, we are now ready to solve the problem of the minimum sampling rate for light field rendering. Since we are dealing with rectangular sampling lattice, the Nyquist sampling theorem for 1D signal applies to both directions v and t. According to the Nyquist sampling theorem, in order for a signal to be reconstructed without aliasing, the sampling frequency needs to be greater than the Nyquist rate, or two times that of the Nyquist frequency. Without loss of generality, we only study the Nyquist frequency along the Ω_t direction in the frequency domain. However, the Nyquist frequency along the Ω_v direction can be analyzed in a similar way.

The minimum interval, by which the replicas of spectral support can be packed without any overlapping, can be computed as shown in Figure 5(a)

$$|P_{max}P_{min}| = K_{\Omega_v} f h_d = 2\pi K_{f_v} f h_d \quad (7)$$

where

$$h_d = \frac{1}{z_{min}} - \frac{1}{z_{max}},$$

and

$$K_{f_v} f h_d = \min(B_v^s, 1/(2\Delta v), 1/(2\delta v))$$

is the highest frequency for the light field signal, which is determined by the scene texture distribution (represented by the highest frequency B_v^s), the resolution of the sampling camera (Δv), and the resolution of the rendering camera (δv). The frequency B_v^s can be computed from the spectral support of the light field. Our formulation takes the rendering resolution into account because rendering at a resolution higher than the output resolution is wasteful. For simplicity, we assume $\delta v = \Delta v$ from now on.

The minimum sampling rate is equivalent to the maximum camera spacing Δt_{max}, which can be computed as

$$\Delta t_{max} = \frac{1}{K_{f_v} f h_d}. \quad (8)$$

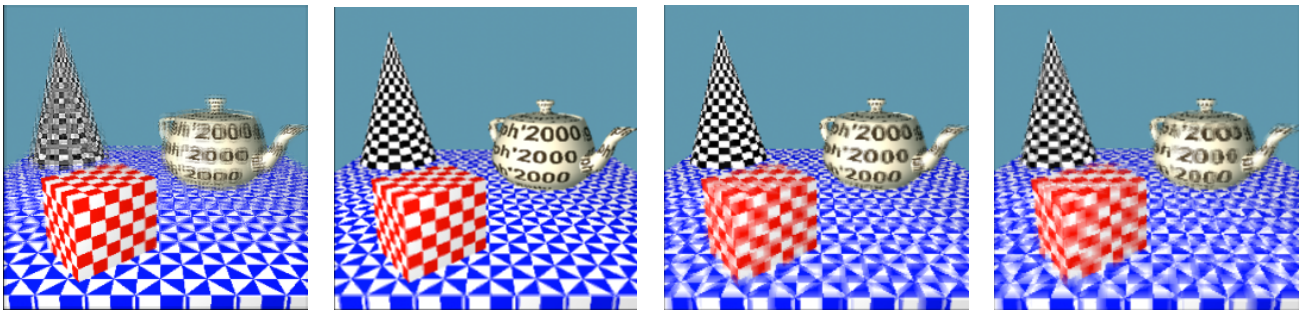
The minimum sampling rate can also be interpreted in terms of the maximum disparity defined as the projection error using the optimal reconstruction filter for rendering. From Equation 8, we have the maximum disparity

$$\Delta t_{max} f h_d / 2 = \frac{1}{2K_{f_v}} = \max(\Delta v, 1/(2B_v^s)). \quad (9)$$

Therefore, the disparity is less than 1 pixel (i.e., the camera resolution) or half cycle of the highest frequency ($1/B_v^s$ is defined as a cycle) presented in the EPI image because of the textural complexity of the observed scene.

If the textural complexity of the scene is not considered, the minimum sampling rate for light field rendering can also be derived in the spatial domain. For example, by considering the light field rendering as a synthetic aperture optical system, we present an optical analysis of light field rendering in Appendix A.

The maximum camera spacing will be larger if the scene texture variation gets more uniform, or if the rendering camera resolution becomes lower. By setting the higher frequency part of the spectrum to zero so that $B_v^s < 1/(2\Delta v)$, we can reduce the minimum sampling rate. One way to reduce B_v^s is to apply a low-pass filter to



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

Figure 7: Sweeping a constant depth plane through an object: (a) at the minimum depth; (b) at the optimal plane; (c) at the average distance between minimum and maximum depths; (d) at the maximum depth. The best rendering quality is achieved in (b).

the input v-t image. This approach is similar to prefiltering a light field (see Figure 7 in [16]).

In particular, the minimum sampling rate is also determined by the relative depth variation $f(z_{min}^{-1} - z_{max}^{-1})$. The closer the object gets to the camera, the smaller the z_{min} is, and the higher the minimum sampling rate will be. As f gets larger, the sampling camera will cover a more detailed scene, but the minimum sampling rate needs to be increased. Therefore, the plenoptic sampling problem should not be considered in the image space alone, but in the joint image and geometry space.

3 Minimum sampling in the joint image and geometry space

In this section, we will study the minimum sampling problem in the joint geometry and image space. Since the CPU speed, memory, storage space, graphics capability and network bandwidth used vary from users to users, it is very important for users to be able to seek the most economical balance between image samples and depth layers for a given rendering quality.

It is interesting to note that the minimum sampling rate for light field rendering represents essentially one point in the joint image and geometry space, in which little amount of depth information has been utilized. As more geometrical information becomes available, fewer images are necessary at any given rendering resolution. Figure 8 illustrates the minimum sampling rate in the image space, the minimum sampling curve in the joint image and geometry space, and minimum sampling curves at different rendering resolutions. Any sampling point above the minimum sampling curve (e.g., Figure 8b) is redundant.

3.1 Minimum sampling with accurate depth

From an initial set of accurate geometrical data, we can decompose a scene into multiple layers of sub-regions. Accordingly, the whole spectral support can be decomposed into multiple layers (see Figure 5b) due to the correspondence between a constant depth and its spectral support. For each decomposed spectral support, an optimal constant depth filter can be designed. Specifically, for each depth layer $i = 1, \ldots, N_d$, the depth of optimal filter is described as follows

$$\frac{1}{z_i} = \lambda_i \frac{1}{z_{min}} + (1 - \lambda_i)\frac{1}{z_{max}} \tag{10}$$

where

$$\lambda_i = \frac{i - 0.5}{N_d}$$

Therefore a depth value can be assigned to one of the depth layers $z = z_i$ if

$$\frac{-h_d}{2N_d} \leq \frac{1}{z} - \frac{1}{z_i} \leq \frac{h_d}{2N_d}. \tag{11}$$

The layers are quantized uniformly in the disparity space. This is because perspective images have been used in the light fields. If we use parallel projection images instead, the quantization should be uniform in the depth space [5].

Similar to Equation 8, the minimum sampling in the joint image and accurate depth space is obtained when

$$\frac{\Delta t}{N_d} = \frac{1}{K_{f_v} f h_d}, \; N_d \geq 1 \tag{12}$$

where N_d and Δt are the number of depth layers and the sampling interval along the t direction, respectively. The interval between replicas is uniformly divided into N_d segments.

The number of depth layers needed for scene representation is a function of the sampling and rendering camera resolution, the scene's texture complexity, the spacing of the sampling cameras and the depth variation relative to the focal length.

3.1.1 Applications

Based on the above quantitative analysis in the joint image and depth space for sufficient rendering, a number of important applications can be explored.

- **Image-based geometry simplification.** Given the appropriate number of image samples an average user can afford, the minimum sampling curve in the joint space determines how much depth information is needed. Thus, it simplifies the original complex geometrical model to the minimum while still guaranteeing the same rendering quality.

- **Geometry-based image database reduction.** In contrast, given the number of depth layers available, the number of image samples needed can also be reduced to the minimum for a given rendering resolution. The reduction of image samples is particularly useful for light field rendering.

- **Level of details (LOD) in joint image and depth space.** The idea of LOD in geometry space can be adopted in our joint image and geometry space. When an object becomes farther away, its relative size on screen space diminishes so that the number of required image samples or the number of required depth layers can be reduced accordingly. Zooming-in onto and zooming-out of objects also demand a dynamic change in the number of image samples or depth layers.

- **Light field with layered depth.** A general data structure for the minimum sampling curve in the joint image and geometry space can be light field with layered depth. With different numbers of images and depth layers used, the trade-off between rendering speed and data storage has to be studied.

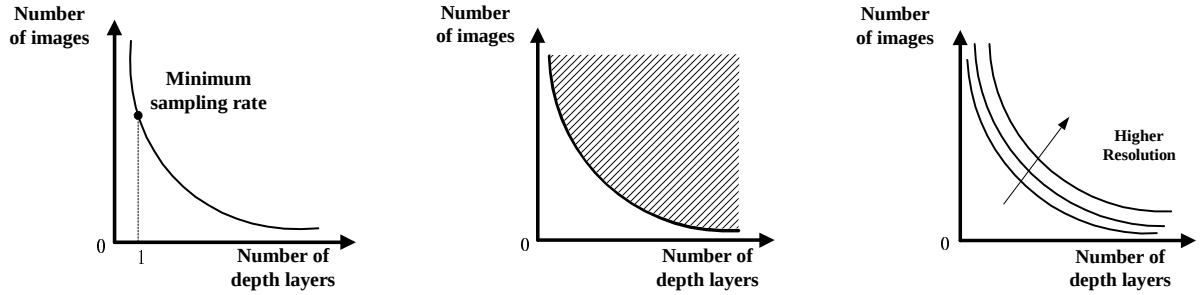


Figure 8: Plenoptic sampling: (a) the minimum sampling rate in image space; (b) the minimum sampling curve in the joint image and geometry space (any sampling point above the curve is redundant); (c) minimum sampling curves at different rendering resolutions.

3.2 Minimum sampling with depth uncertainty

Another aspect of minimum sampling in the joint image and geometry space is related to depth uncertainty. Specifically, minimum sampling with depth uncertainty describes the quantitative relationship between the number of image samples, noisy depth and depth uncertainty. It is important to study this relationship because in general the recovered geometry is noisy as modeling a real environment is difficult. Given an estimated depth z_e and its depth uncertainty $\Delta\eta$, the depth value should be located within the range $(z_e - \Delta\eta, z_e + \Delta\eta)$. The maximum camera spacing can be computed as

$$\Delta t_{max} = \min_{z_e} \frac{(z_e + \Delta\eta)(z_e - \Delta\eta)}{2fK_{f_v}\Delta\eta} = \frac{\min_{z_e} z_e^2 - \Delta\eta^2}{2fK_{f_v}\Delta\eta}. \quad (13)$$

In addition, geometrical uncertainty also exists when an accurate model is simplified. Given the correct depth z_0 and an estimated depth z_e, the maximum camera spacing can be computed as

$$\Delta t_{max} = \min_{z_e} \frac{z_e z_0}{2fK_{f_v}|z_e - z_0|}. \quad (14)$$

3.2.1 Applications

Knowledge about the minimum number of images under noisy depth has many practical applications.

- **Minimum sampling rate.** For a specific light field rendering with no depth maps or with noisy depth maps, we can determine the minimum number of images for antialiased light field rendering. Redundant image samples can then be left out from the sampled database for light field rendering.
- **Rendering-driven vision reconstruction.** This is a very interesting application, considering that general vision algorithms would not recover accurate scene depth. Given the number of image samples, how accurately should the depth be recovered to guarantee the rendering quality? Rendering-driven vision reconstruction is different from classical geometry-driven vision reconstruction in that the former is guided by the depth accuracy that the rendering process can have.

4 Experiments

Table 1 summarizes the parameters of each light field data set used in our experiments. We assume that the output display has the same resolution as the input image. Furthermore, without taking into consideration the actual texture distribution, we assume that the highest frequency in images is bounded by the resolution of the capturing camera.

We have used different settings of focal length for the Head, the Statue and the Table. We put the focal plane slightly in front of the Head. A smaller focal length will reduce the minimum sampling rate. For the Statue, the focal plane is set approximately at its forehead. In fact, we have set the focal length (3000) very close to the optimal (3323). Because the Table scene has significant depth variations, a small camera focal length was used so that each image can cover a large part of the scene.

First, we compare the rendering quality along the minimal sampling curve in the joint image and geometry space, with the best rendering quality we can obtain with all images and accurate depth. According to our theory (Eq (12)), the number of images is inversely proportional to the number of depth layers in use. The rendering results corresponding to five different image and depth combinations along the minimum sampling curve are shown in Figures 11(A)-(E). For example, C(7,8) represents the rendering result using 7 layers of depth and 8×8 images. In contrast, Figure 11(F) shows the best rendering output one can achieve from this set of data: accurate depth and all 32×32 images[2]. The quality of the rendered images along the minimal sampling curve is almost indistinguishable[3] from that of using all images and accurate depth.

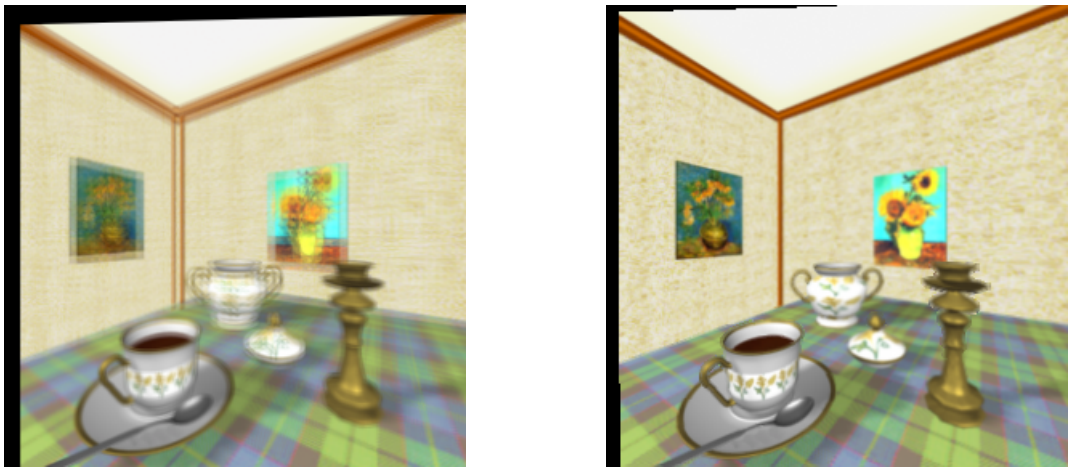
Figure 12(a) compares the rendering quality using different layers of depth and a given number of image samples. With 2×2 image samples of the Head, images (A)-(E) in Figure 12(a) show the rendered images with different layers of depth at 4, 8, 10, 12, and 24. According to Eq (12), the minimum sampling point with 2×2 images of the Head is at approximately 12 layers of depth. Noticeable visual artifacts can be observed when the number of depth is below the minimal sampling point, as shown in images (A)-(C) of Figure 12(a). On the other hand, oversampling layers of depth does not improve the rendering quality, as shown in the images (D) and (E).

With the minimal sampling curve, we can now deduce the minimum number of image samples at any given number of depth layers available. For the Table scene, we find that 3 bits (or 8 layers) of depth information is sufficient for light field rendering when combined with 16×16 image samples (shown in image (D) of Figure 12(b)). When the number of depth layers is below the minimal sampling point, light field rendering produces noticeable artifacts, as shown in images (A)-(C) of Figure 12(b).

Given a single depth layer, our analysis (Eq 12) shows that the number of images for antialiased rendering of the table scene requires 124×124 images. Note that conventional light field may require even a larger number of images without using the optimal depth. This very large set of light field data is due to the signifi-

[2]We were not able to use all 64×64 images with accurate depth because of memory limitations.

[3]There exists little discrepancy because of the fact that we can not apply the optimal reconstruction filter in rendering.



(a) (b)

Figure 9: Comparison between conventional light field with 48×48 images and rendering with 16×16 images and 3 bits of depth: (a) artifacts are visible on the left with conventional rendering, (b) but not present with additional geometrical information because minimum sampling requirement is satisfied.

cant depth variations in the Table scene. This perhaps explains why inside-looking-out light field rendering has not been used often in practice. Also according to our analysis, using 3 bits (8 layers) of depth helps to reduce the number of images needed by a factor of 60, to 16×16 images. For comparison, Figure 9(a) shows conventional light field rendering with 48×48 images and Figure 9(b) shows the rendering result with 16×16 images plus 3 bits of depth. Visual artifacts such as double images at the edge of the wall are clearly visible in Figure 9(a). They are not present in Figure 9(b).

Experiments using depth with uncertainty also demonstrate the effectiveness of our analysis. Due to space limitation, we will not present any results of minimum sampling curve using depth with uncertainty.

5 Conclusion and future work

In this paper we have studied the problem of plenoptic sampling. Specifically, by analyzing the bounds of spectral support of light field signals, we can analytically compute the minimum sampling rate of light field rendering. Our analysis is based on the fact that the spectral support of a light field signal is bounded by only the minimum and maximum depths, irrespective of how complicated the spectral support might be because of depth variations in the scene. Given the minimum and maximum depths, a reconstruction filter with an optimal constant depth can be designed for anti-aliased light field rendering. The minimum sampling rate for light field rendering is obtained by compacting the replicas of the spectral support of the sampled light field within the smallest interval. Our work provides a solution to overcoming the oversampling problem in light field capturing and rendering.

By studying plenoptic sampling in the joint image and geometry space, we have also derived the minimum sampling curve which quantitatively describes the relationship between the number of images and the information on scene geometry, given a specific rendering resolution. Indeed, minimum sampling curves with accurate depth and with noisy depth serve as the design principles for a number of applications. Such interesting applications include image-based geometry simplification, geometry-assisted image dataset reduction, rendering-driven vision reconstruction, in addition to depth-assisted light field compression, or the minimum sampling rate for light field rendering.

While we have studied minimum sampling using light fields in this paper, the very idea of plenoptic sampling is also applicable to other IBR systems, e.g. concentric mosaics, layered-depth image, view interpolation, and image warping, to name a few. With plenoptic sampling, there are a number of exciting areas for future work.

For example, we have used depth value in this paper to encode the geometry information. Depth is also used in image-assisted geometry simplification. However, no surface normal has been considered. In the future, we would like to experiment with different techniques to generate image-assisted geometry simplification using geometrical representations other than depth. We plan to incorporate the surface normal into image-based polygon simplification. The efficiency of geometry simplification can be further enhanced by considering the standard techniques in geometrical simplification, e.g. visibility culling.

Another interesting line of future work is on how to design a new rendering algorithm for the joint image and geometry representation. The complexity of the rendering algorithm should be proportional to the number of depth in use. In addition, error-bounded depth reconstruction should be considered as an alternative to traditional vision reconstruction, if the reconstruction result is to be used for rendering. Given the error bounds that are tolerable by the rendering algorithms, the difficulty of vision reconstruction can be much alleviated.

Lastly, we plan to study view-dependent plenoptic sampling. Current analysis of plenoptic sampling is based on the assumption that the surface is diffuse and little view-dependent variance can occur. It is conceivable that view dependent surface property will increase the minimum sampling rate for light field.

6 Acknowledgements

The authors benefited from discussions with Zhouchen Lin on truncated models and Tao Feng on optical analysis of light field rendering. Mr. Yin Li's incredible help on preparing Siggraph video is greatly appreciated. The last author also wishes to thank Pat Hanrahan for his helpful discussion on the minimum sampling of concentric mosaics while visiting Stanford in April 1999. Finally, the authors thank Siggraph reviewers' comments which have tremendously helped to improve the final manuscript.

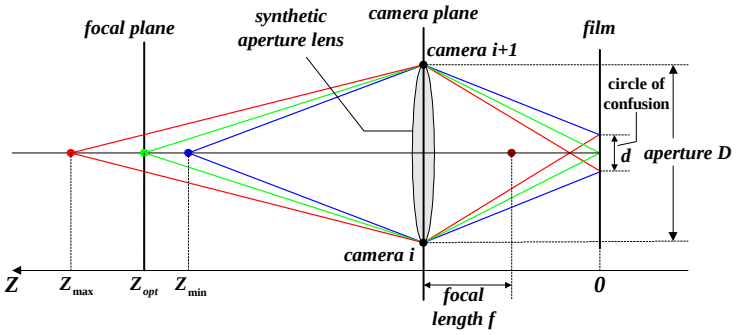


Figure 10: A discrete synthetic aperture optical system for light field rendering.

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] M. Bass, editor. *Handbook of Optics*. McGraw-Hill, New York, 1995.

[3] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1:7–55, 1987.

[4] E. Camahort, A. Lerios, and D. Fussell. Uniformly sampled light fields. In *Proc. 9th Eurographics Workshop on Rendering*, pages 117–130, 1998.

[5] J.-X. Chai and H.-Y. Shum. Parallel projections for stereo reconstruction. In *Proc. CVPR 2000*, 2000.

[6] C. Chang, G. Bishop, and A. Lastra. Ldi tree: A hierarchical representation for image-based rendering. *SIGGRAPH'99*, pages 291–298, August 1999.

[7] S. Chen and L. Williams. View interpolation for image synthesis. *Computer Graphics (SIGGRAPH'93)*, pages 279–288, August 1993.

[8] T. Feng and H.-Y. Shum. An optical analysis of light field rendering. Technical report, Microsoft Research, MSR-TR-2000-38, May 2000.

[9] B. Girod. Motion compensation: visual aspects, accuracy, and fundamental limits. In *Motion Analysis and Image Sequence Processing*. Kluwer, 1995.

[10] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Computer Graphics Proceedings, Annual Conference Series*, pages 43–54, Proc. SIGGRAPH'96 (New Orleans), August 1996. ACM SIGGRAPH.

[11] M. Halle. Holographic stereograms as discrete imaging systems. In *Proc. SPIE Vol.2176, Practical Holography VIII*, pages 73–84, May 1994.

[12] V. Hlavac, A. Leonardis, and T. Werner. Automatic selection of reference views for image-based scene representations. In *Proc. ECCV*, pages 526–535, 1996.

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

[14] S. Kang. A survey of image-based rendering techniques. In *VideoMetrics, SPIE Vol. 3641*, pages 2–16, 1999.

[15] J. Lengyel. The convergence of graphics and vision. Technical report, IEEE Computer, July 1998.

[16] M. Levoy and P. Hanrahan. Light field rendering. In *Computer Graphics Proceedings, Annual Conference Series*, pages 31–42, Proc. SIGGRAPH'96 (New Orleans), August 1996. ACM SIGGRAPH.

[17] Z.-C. Lin and H.-Y. Shum. On the numbers of samples needed in light field rendering with constant-depth assumption. In *Proc. CVPR 2000*, 2000.

[18] W. Mark, L. McMillan, and G. Bishop. Post-rendering 3d warping. In *Proc. Symposium on I3D Graphics*, pages 7–16, 1997.

[19] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *Computer Graphics (SIGGRAPH'95)*, pages 39–46, August 1995.

[20] H. Schirmacher, W. Heidrich, and H. Seidel. Adaptive acquisition of lumigraphs from synthetic scenes. In *Eurographics'99*, pages 151–159, Sept 1999.

[21] S. M. Seitz and C. M. Dyer. View morphing. In *Computer Graphics Proceedings, Annual Conference Series*, pages 21–30, Proc. SIGGRAPH'96 (New Orleans), August 1996. ACM SIGGRAPH.

[22] 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.

[23] H.-Y. Shum and L.-W. He. Rendering with concentric mosaics. In *Proc. SIGGRAPH 99*, pages 299–306, 1999.

[24] P. P. Sloan, M. F. Cohen, and S. J. Gortler. Time critical lumigraph rendering. In *Symposium on Interactive 3D Graphics*, pages 17–23, Providence, RI, USA, 1997.

[25] A. Tekal. *Digital Video Processing*. Prentice Hall, 1996.

A An optical analysis of light field rendering

Similar to [16, 13], we consider the light field rendering system as a discrete synthetic aperture optical system, as shown in Figure 10. Analogous to the Gaussian optical system, we can define the following optical parameters:

- Focal length f;
- Smallest resolvable feature (on the image plane) d;
- Aperture D. Distance between two adjacent cameras;
- Circle of confusion $c = d/f$;
- Hyperfocal distance $D_H = D/c$.

Let the plane of perfect focus be at the distance z_{opt}, the minimum and maximum distances at which the rendering is acceptable be z_{min} and z_{max}, respectively. The following relations exist ([2], vol. 1, p.1.92)

$$z_{min} = \frac{D_H z_{opt}}{D_H + z_{opt}}, \quad \text{and} \quad z_{max} = \frac{D_H z_{opt}}{D_H - z_{opt}},$$

which lead to,

$$\begin{aligned} \frac{1}{z_{opt}} &= (\frac{1}{z_{min}} + \frac{1}{z_{max}})/2 \\ \frac{1}{D_H} &= (\frac{1}{z_{min}} - \frac{1}{z_{max}})/2 \end{aligned}$$

Therefore, to have the best rendering quality, no matter which optical system is used, the focus should be always at z_{opt}. Moreover, to guarantee the rendering quality, D_H has to be satisfied, i.e.,

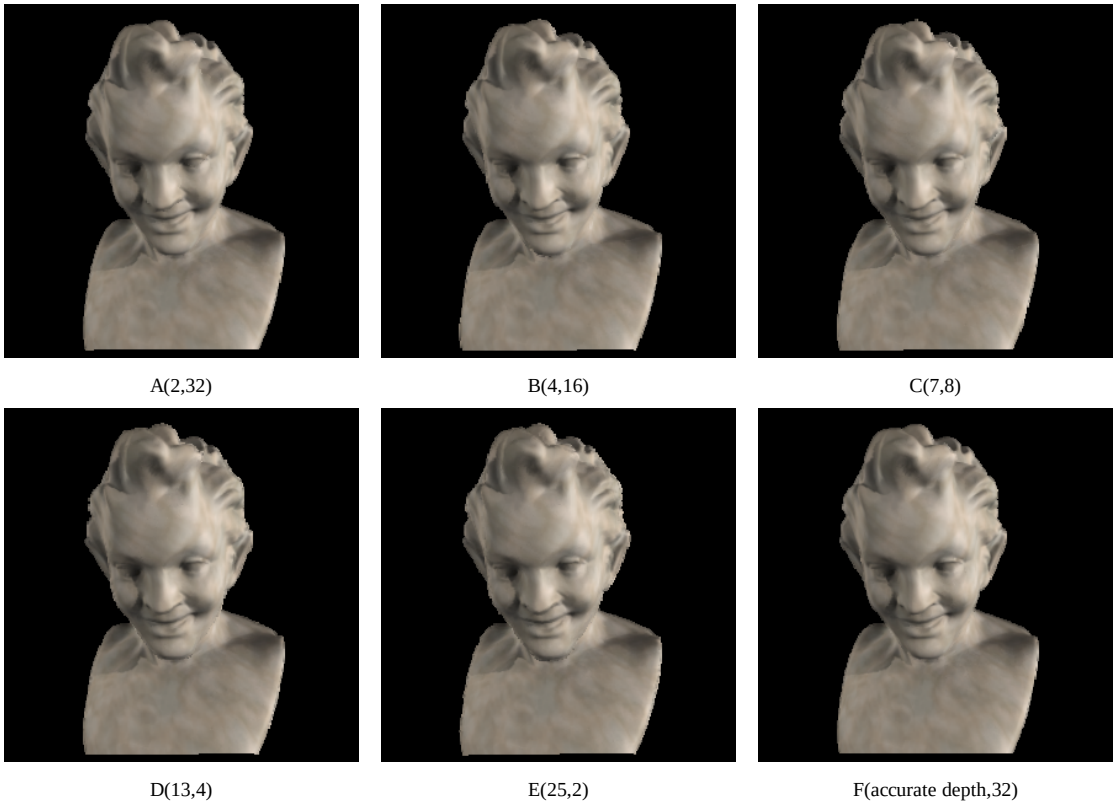
$$\frac{D}{d/f} = (\frac{1}{z_{min}} - \frac{1}{z_{max}})/2 \tag{15}$$

In other words, given the minimum and maximum distances, the maximum camera spacing can be determined in order to meet the specified rendering quality. The hyperfocal distance describes the relationship among the rendering resolution (circle of confusion), the scene geometry (depth of field) and the number of images needed (synthetic aperture). Intuitively, the minimum sampling rate is equivalent to having the maximum disparity less than the smallest resolvable feature on the image plane, e.g, camera resolution or one pixel, i.e., $d = \delta_v = 1$. The same result was also obtained by Lin and Shum [17] using a geometrical approach.

Equation 15, not surprisingly, is almost exactly the same as Equation 8 because $D_H = 2/h_d$. However, our approach using spectral analysis of light field signals incorporates the textural information in the sampling analysis. More detailed optical analysis of light field rendering can be found in [8].

	Focal length	Maximum depth	Minimum depth	(u, v) interval	(s, t) interval	Pixels per image	Image per slab	Spacing Δt_{max}
Head	160.0	308.79	183.40	0.78125	1.5625	256×256	64×64	4.41
Statue	3000.0	5817.86	2326.39	15.625	31.25	256×256	64×64	40.38
Table	350.0	3235.47	362.67	2.4306	7.29	288×288	96×96	5.67

Table 1: A summary of parameters used in three data sets in our experiments.



A(2,32) B(4,16) C(7,8)

D(13,4) E(25,2) F(accurate depth,32)

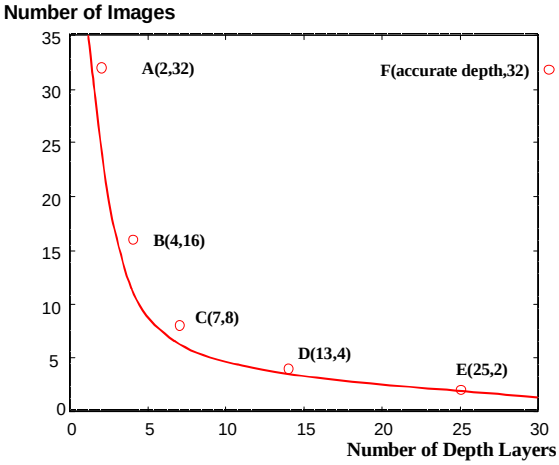
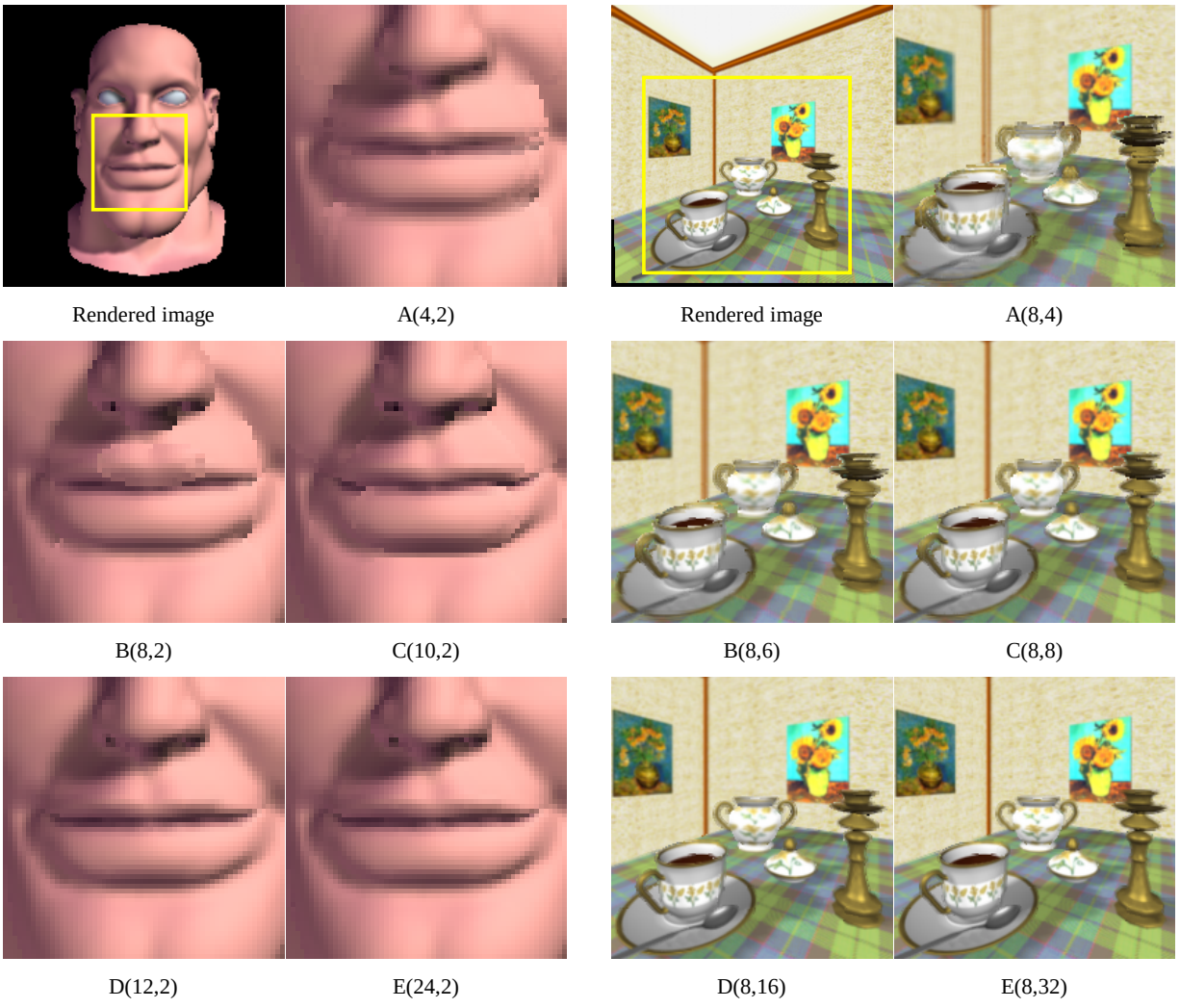
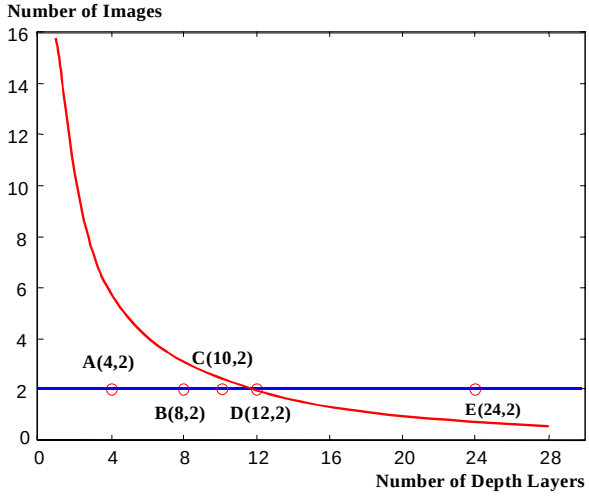
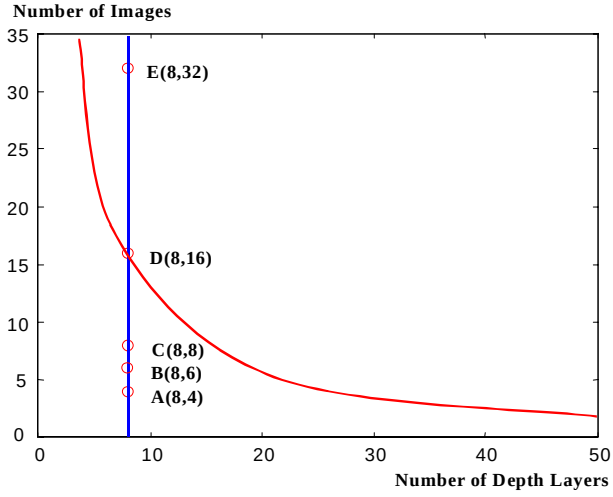


Figure 11: Minimum sampling curve for the object ”Statue” in the joint image and geometry space with accurate geometry. Sampling points in the figure have been chosen to be slightly above the minimum sampling curve due to quantization.





(a)



(b)

Figure 12: Minimum sampling points in the joint image and geometry space: (a) for the object ”Head”, when the number of images is 2×2; (b) for the ”Table” scene, when the number of depth layers is 8.