International Journal of Computer Vision 58(2), 121–138, 2004

A Geometric Analysis of Light Field Rendering

ZHOUCHEN LIN AND HEUNG-YEUNG SHUM
Microsoft Research Asia, Beijing 100080, People's Republic of China

Received May 17, 2002; Revised November 19, 2002; Accepted December 9, 2002

Abstract. Recently, many image-based modeling and rendering techniques have been successfully designed to render photo-realistic images without the need for explicit 3D geometry. However, these techniques (e.g., light field rendering (Levoy, M. and Hanrahan, P., 1996. In *SIGGRAPH 1996 Conference Proceedings*, Annual Conference Series, Aug. 1996, pp. 31–42) and Lumigraph (Gortler, S.J., Grzeszczuk, R., Szeliski, R., and Cohen, M.F., 1996. In *SIGGRAPH 1996 Conference Proceedings*, Annual Conference Series, Aug. 1996, pp. 43–54)) may require a substantial number of images. In this paper, we adopt a geometric approach to investigate the minimum sampling problem for light field rendering, with and without geometry information of the scene. Our key observation is that anti-aliased light field rendering is equivalent to eliminating the "double image" artifacts caused by view interpolation.

Specifically, we present a closed-form solution of the minimum sampling rate for light field rendering. The minimum sampling rate is determined by the resolution of the camera and the depth variation of the scene. This rate is ensured if the optimal constant depth for rendering is chosen as the harmonic mean of the maximum and minimum depths of the scene. Moreover, we construct the minimum sampling curve in the joint geometry and image space, with the consideration of depth discontinuity. The minimum sampling curve quantitatively indicates how reduced geometry information can be compensated by increasing the number of images, and vice versa. Experimental results demonstrate the effectiveness of our theoretical analysis.

Keywords: light field, lumigraph, image-based modeling and rendering, view interpolation, sampling, double image

1. Introduction

Image-based modeling and rendering (IBMR) has become an active research area in computer vision and computer graphics. Many IBMR techniques have been proposed to produce realistic-looking images from prerecorded images, often at low rendering costs. Unlike traditional geometry-based rendering, they do not necessarily rely on exact geometry. For instance, light field rendering (Levoy and Hanrahan, 1996) requires no geometry information, yet generates novel view images by simply interpolating the captured samples of the light field or the plenoptic function (Adelson and Bergen, 1991).

However, light field and Lumigraph rendering techniques require a substantial number of sample images (typically thousands or more). As a result, data reduction is important. Many techniques, such as vector quantization (Levoy and Hanrahan, 1996), multiple reference frame (Zhang and Li, 2000), eigen-texture (Nishino et al., 2001), and light field mapping (Chen et al., 2002), have been proposed to achieve high compression ratio while enabling random access of samples for real-time rendering. Despite so much work on light field data compression, a more important issue that needs to be addressed is the minimum sampling problem for anti-aliased light field rendering. It is obvious that the fewer images we need to capture, the less storage and acquisition time required. Ideally, we should only acquire the minimum number of images needed.

In this paper, we adopt a geometric approach to analyze the light field rendering, and in particular its sampling issues. We first derive the minimum sampling rate of light field when there is no geometry information

available. We show that the minimum sampling rate for anti-aliased light field rendering is equivalent to eliminating the "double image" artifact, which is caused by interpolation with inaccurate geometry. The minimum sampling analysis can provide guidance as to how the acquisition and rendering systems should be set up, e.g., how closely cameras should be spaced and how much pre-filtering (Levoy and Hanrahan, 1996; Halle, 1994) need to be applied to the images.

Furthermore, we derive the minimum sampling curve of the light field in the joint image and geometry space when geometry information becomes available. This curve quantitatively shows how the same rendering quality can be achieved by trading off the amount of geometry information and the number of images.

The remainder of this paper is organized as follows. In Section 2, we review related work. In Section 3, we present our criterion of acceptable rendering quality. The minimum sampling rate and minimum sampling curve of the light field are then studied in Sections 4 and 5, respectively. The issue of sampling with occlusion is discussed in Section 6. Finally, we present concluding remarks in Section 7.

2. Background and Previous Work

Both light field (Levoy and Hanrahan, 1996) and Lumigraph (Gortler et al., 1996) use a two-plane parameterization of the 4D plenoptic function (Adelson and Bergen, 1991). As shown in Fig. 1, an oriented line is defined by connecting a point on the uv plane (or the camera plane) to a point on the st plane (or the focal plane). Therefore, each ray in the free space is uniquely determined by a quadruple (u, v, s, t). Both the camera and focal planes are uniformly discretized (Fig. 1(a)) for simplicity.

The light field could be captured by placing a camera on the uv plane facing the st plane. Images are taken at each grid on the uv plane. To ensure that the captured light rays pass through the grids on the st plane, sheared perspective projection and resampling are performed. To render a novel ray, its intersection points with the uv and st planes are first computed. Subsequently, the nearest 16 (or part of the 16) sampling rays in the light slab around the novel ray are selected to interpolate the novel ray. Figure 1(b) shows four of these sample rays used in interpolation.

Other sampling techniques have also been proposed to capture the light field. Two-sphere and sphere-plane parameterizations (Camahort et al., 1998) can be used for more uniform sampling. To minimize redundancy, Schirmacher et al. (1999) proposed to adaptively add a new view by determining if its addition improves the rendering quality. Alternatively, Hlaváč et al. (1996) and Sloan et al. (1997) discarded an input view if it can be predicted by nearby frames. Concentric Mosaics (Shum and He, 1999) sampled the light field along a circle rather than on a plane. The dynamically reparameterized light field (Isaksen et al., 2000) used a variable focal plane/surface, and the surface light field (Wood et al., 2000; Nishino et al., 2001; Chen et al., 2002) indexed rays by points on the object surface for high-quality rendering. While all the above techniques relied on regular parameterization, the unstructured Lumigraph (Buehler et al., 2001) handled the problem of rendering with irregularly sampled images. However, none of these methods provide a

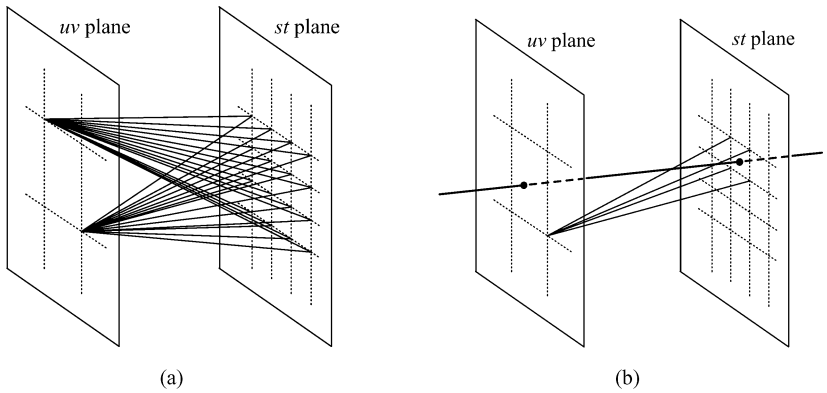

Figure 1. The representation, sampling and reconstruction of a light field. (a) The light slab after sampling and resampling. The uv and st planes are the camera plane and the focal plane, respectively. (b) A novel ray can be interpolated from nearby sample rays (for clarity, only partial nearby sample rays are shown here) by computing its intersection points with the uv and st planes.

quantitative analysis on the number of samples needed and how the sampling rate is related to geometry.

On the theoretical front, Halle (1994) pointed out the relationship between pixel disparity and aliasing in the context of holographic stereograms. He suggested, yet without justification, that a sampling rate be chosen such that the images of a point from all the contributing views should abut each other without gaps. No explicit formulae were given for the sampling rate and the size of prefiltering in Halle (1994).

The first-ever analysis on minimum sampling rates for light fields (and Concentric Mosaics) was proposed in Lin and Shum (2000). Through a geometric analysis, it was shown in Lin and Shum (2000) that the sampling rate is dependent on scene disparity variation and camera resolution. Soon after, the same results were obtained in Chai et al. (2000) where a spectral analysis was adopted to take scene texture into account as well. The trade-off between the number of images and the amount of geometry information was also studied in Chai et al. (2000) and the minimum sampling curve in the joint space of image and geometry was derived without the consideration of occlusion.

This paper is an extension of our previous work (Lin and Shum, 2000). From the geometric perspective, we study the artifact of "double image" (a geometric counterpart of spectral aliasing), optimal constant depth, and depth layers. Our geometric analysis is an alternative to the spectral analysis in Chai et al. (2000). Yet it is more intuitive and flexible, and is applicable to irregular capturing and rendering configurations. For example, the results on Concentric Mosaics (Lin and Shum, 2000), which are difficult to obtain using spectral analysis, can be easily obtained. Moreover, by taking the discontinuity factor into account, our sampling analysis is more accurate than that in Chai et al. (2000).

3. Problem Formulation

Before introducing our geometric approach, we need to make some assumptions about the camera, scene, and interpolation methods used in light field rendering.

3.1. Assumptions

The list of assumptions we make in our analysis is as follows:

- *Camera*: pin-hole with a finite resolution.
- *Scene*: occlusion-free and Lambertian.
- *Interpolation method*: bilinear.

3.1.1. Camera. In our analysis we adopt a pin-hole camera model with a finite resolution. Thus, the camera records a blurred version of the plenoptic function or the light field. A pixel value is a weighted integral of the illumination of the light arriving at the camera plane. Alternatively, a pixel value is the convolution of the plenoptic function at the optical center with a low-pass filter. The shape of the filter is compactly supported, with the width of support being the angular resolution of camera. Equivalently, the camera simply samples the convoluted plenoptic function at the camera center. The value of a pixel is exactly the value of the blurred plenoptic function at the direction linking the pixel and the optical center.

Throughout this paper, we use uniform angular resolution, in both vertical and horizontal directions. Both capturing and rendering cameras have the same resolution.

3.1.2. Scene. To simplify our analysis, we study the characteristics of the scene element and bound its depth. The angular extent of a point is sufficiently small compared to the camera resolution, but not zero. Since a scene is composed of points, if every point can be correctly rendered, so can the scene. The scene points are first dealt with independently (i.e., ignoring occlusion) and the discussion on sampling problem with occlusion is postponed until Section 6. Our microscopic analysis methodology is inspired by common practices in physics where theories are often built on the analysis of independent particles and the interaction between particles. Moreover, by assuming that the scene is Lambertian, we focus our analysis on the scene geometry and ignore the illumination effects.

3.1.3. Interpolation. The rendering process involves choosing "nearby" rays for each viewing ray and interpolating using them. Therefore, an estimate on the depth along the viewing ray is required for ray query. In fact, any approach to find "nearby" rays inherently involves an assumption of some scene depth. It is around where the rays intersect. For example, infinite depth is assumed in rendering with Concentric Mosaics, because interpolation using parallel rays is equivalent to infinite depth. In rendering with the light field, the depth is always implicitly assumed to be at the focal plane.

Usually, the nearest samples are the most important to reconstruct a signal. Bilinear interpolation has been commonly used in existing light field rendering

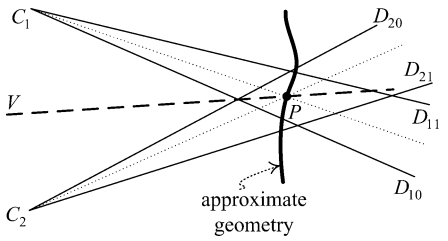

Figure 2. Rendering with bilinear interpolation: the scene point along a ray *VP* is estimated at P and the ray is interpolated using the four nearest rays from cameras at C_1 and C_2.

systems (e.g., Chen and Williams, 1993; Levoy and Hanrahan, 1996; Shum and He, 1999) because it is simple and can produce good rendering quality. In the presence of more accurate depth information, better rendering results can be obtained by interpolating more light rays.

Figure 2 illustrates the rendering process. Suppose that we want to render a view at V, and the viewing ray intersects the approximate geometry at P. C_1 and C_2 are two nearby positions of the camera that are closest to *VP*, and C_iD_{ij} ($i = 1, 2$; $j = 0, 1$) are nearby rays in camera C_i that are closest to the ray C_iP. Then the pixel value of *VP* can be bilinearly interpolated from rays C_iD_{ij} (e.g., Wu and Shum, 2000), by assigning weights w_{ij} to rays C_iD_{ij} in the following manner:

$$w_{10} = \frac{\angle VPC_2 \cdot \angle PC_1D_{11}}{(\angle VPC_1 + \angle VPC_2) \cdot \angle D_{10}C_1D_{11}},$$
$$w_{11} = \frac{\angle VPC_2 \cdot \angle PC_1D_{10}}{(\angle VPC_1 + \angle VPC_2) \cdot \angle D_{10}C_1D_{11}},$$
$$w_{20} = \frac{\angle VPC_1 \cdot \angle PC_2D_{21}}{(\angle VPC_1 + \angle VPC_2) \cdot \angle D_{20}C_2D_{21}},$$
$$w_{21} = \frac{\angle VPC_1 \cdot \angle PC_2D_{20}}{(\angle VPC_1 + \angle VPC_2) \cdot \angle D_{20}C_2D_{21}}.$$

3.2. Anti-Aliasing Condition

In this section, we investigate the visual artifacts caused by rendering with interpolation and inaccurate depth. We show that the anti-aliased light field rendering is equivalent to eliminating "double images" for each scene point.

3.2.1. Widening of Intensity Contribution after Interpolation. We first consider within-view (intra-view) interpolation. As shown in Fig. 3(a), camera C_1 is taking a snapshot of a point L. C_1D_{10} is the nearest sampling ray to C_1L while C_1D_{11} and C_1D_{12} are two nearby rays. Figure 3(b) maps the intensity contribution of L as a function of camera angular position, where

- the vertical line at 0 represents the ray C_1L,
- δ is the angular resolution of the camera,
- the parabola-like curve is the intensity contribution of L in the continuous case (or the shape of low-pass filter for the blurred plenoptic function),
- and C_1D_{10} is displaced by angle ε ($-\frac{\delta}{2} \le \varepsilon < \frac{\delta}{2}$) from C_1L.

As a result, the contribution of L to the pixel that corresponds to the ray C_1D_{10} is just the value of the continuous contribution at ε. Since the angles of C_1D_{11} and C_1D_{12}, corresponding to $\varepsilon-\delta$ and $\varepsilon+\delta$ in Fig. 3(b), respectively, are outside the interval of $[-\frac{\delta}{2}, \frac{\delta}{2})$, neither intensities of the two corresponding pixels are affected by the point L. Subsequent to linear interpolation, the intensity contribution of the point L becomes a wedge of width 2δ. As will be seen, it is the *width*, not the shape of the intensity contribution, that matters.

Suppose C_1 is one of the nearby cameras to the novel view V (Fig. 4(a)), C_1L intersects the approximate

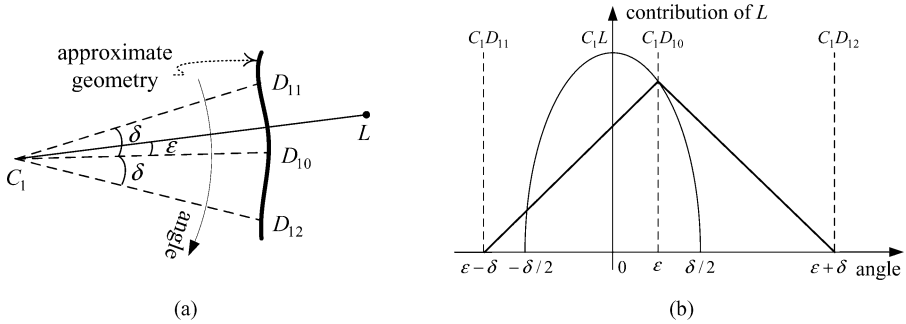

Figure 3. The change of intensity contribution. (a) The camera C_1 is imaging a scene point L. (b) The intensity contribution of L changes from parabola-like to wedge-shape due to finite camera resolution and linear interpolation.

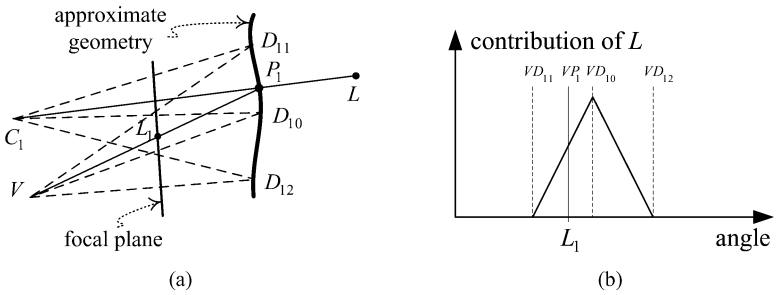

Figure 4. Using the rays in camera C_1 generates a wedge of intensity contribution around L_1 in the novel view.

geometry at P_1, and VP_1 intersects the focal plane at L_1. Then the part of the contribution of L to the novel view, transferred by camera C_1, centers around L_1 (Fig. 4(b)).

3.2.2. Rendering Quality vs. Geometry Information. Now we consider the between-view (inter-view) interpolation. As shown in Fig. 5(a), C_1 and C_2 are two nearby cameras, V is the novel view, and L is the scene point of interest. Suppose $C_i L$ intersects the approximate geometry at P_i and VP_i intersects the focal plane at L_i ($i = 1, 2$). Then around L_i lie two intensity contributions of the point L (Fig. 5(b)) on the plane.

The contribution of L to the viewing rays is interpolated between these two intensity contributions. Obviously, the rendering quality depends on the distance between them. In Fig. 6, the relative positions between the two intensity contributions are shown, where the horizontal axis represents the angular position of the rays in view V and the vertical axis represents the amount of contribution. The thick vertical dash-lines represent the viewing rays. When they nearly overlap (Fig. 6(a)), there is only one pixel strongly influenced by L. So L will appear sharp on the rendered image. When they partially overlap (Fig. 6(b)), then L contributes to two consecutive pixels and L will become blurred. When they no longer touch (Fig. 6(d)), some viewing rays can fall in the gap between them. If the contrast between L and its neighborhood is large, the intensities of in-between viewing rays are different from the intensity of L. As a result, there will be two separate images of L on the rendered image. This phenomenon is the "double image" artifact.

The distance between the two intensity contributions is dependent on the sample spacing and the geometry, so is the rendering quality. When the sample spacing becomes larger and larger, or when the geometry information becomes less and less accurate, the rendered point gradually changes from being sharp to being blurry and further to becoming a double image artifact. The double image artifact is a result of an inadequate sampling rate.

Strictly speaking, the change from being sharp to having double images is continuous. Nevertheless, the condition of two intensity contributions touching each other (Fig. 6(c)) is a critical one.

3.2.3. When is Rendering Quality Acceptable? We now consider the circumstances under which a rendered image is considered acceptable. A rendered scene point may either be sharp, be blurry, or have double images. Sharpness is certainly what we desire. Thus it is important that we make a distinction between blurring and double images.

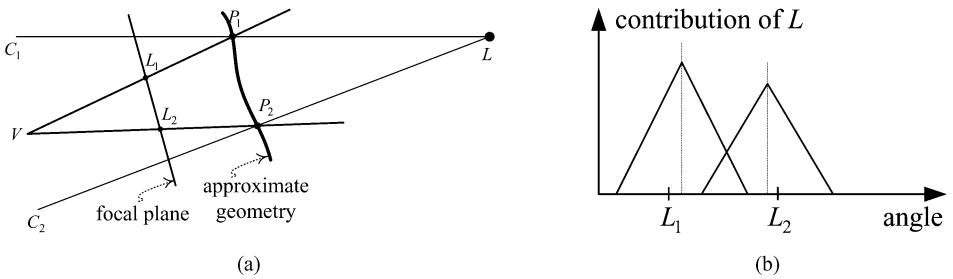

Figure 5. Bilinear interpolation generates two wedges of intensity contribution around L_1 and L_2. Double images of L may appear on the rendered image if the two wedges do not overlap.

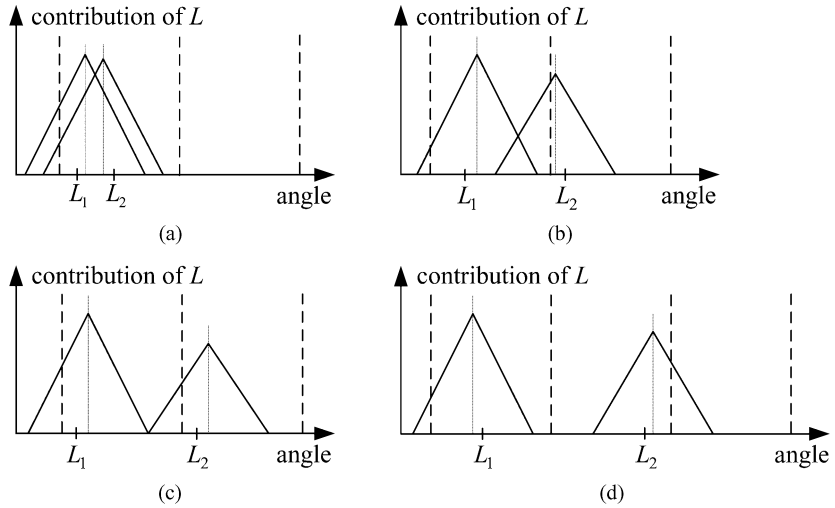

Figure 6. The rendering quality depends on the distance between the two intensity contributions. The viewing rays are indicated by the thick vertical dash-lines. (a) When the intensity contributions are very close, the rendered point appears sharp. (b) When the intensity contributions partially overlap, it looks blurred. (c) That the two intensity contributions meet their ends is the critical condition that double images may occur. (d) Double images may occur when the two intensity contributions do not overlap.

The phenomenon of double images has also been observed by Levoy and Hanrahan (1996) and Halle (1994). Double images are the most salient and visually disturbing artifact, particularly when the sampling rate is low. Human perception is more tolerant to blurring than to double images. People are accustomed to seeing blurred objects in photos, which correspond to off-focus locations. It stands to reason that an object should only appear blurred with no double images should the geometry be inaccurate.

The above human perception has been summarized by the causality principle in scale-space theory (Romeny, 1994), i.e., no "spurious detail" should be generated when smoothing an image. The light field rendering should obey this principle because both the intra- and inter-view interpolations are smoothing processes on a single image. The intra-view case is obvious. For the inter-view case, under the Lambertian premise and ignoring the visibility problem, the value of every pixel on the interpolated image can be computed by weighting several pixels on only one of the images chosen for interpolation. The weighting template is pixel-dependent, but it is still a smoothing operation because all the weights are non-negative. As a result, blurring is harmless but double images should be eliminated.

From the viewpoint of signal processing, light field rendering is a reconstruction process of a 4D signal. When the sampling rate is inadequate, aliasing shows up as double images on the rendered image. Indeed, we can prove that the overlap between two successive intensity contributions on the focal plane is equivalent to the anti-aliasing condition. The details of the proof are presented in Appendix A.

In conclusion, rendering quality is acceptable when all objects in the rendered image appear either sharp or blurred. No double images are allowed.

4. Minimum Sampling Rate

We now study the minimum sampling rate for light field rendering without any geometric information. Specifically, we apply the criterion in Section 3.2.3 to analyze the maximum allowable distance between successive camera locations. However, interpolation is impossible if no depth along a viewing ray is assumed. Therefore, the minimum geometric information, described by a globally constant depth, is still necessary. In fact, the interpolation method of the original light field rendering (Levoy and Hanrahan, 1996) implicitly assumes a globally constant depth at the focal plane.

4.1. Maximum Camera Spacing

Without loss of generality, we set up a coordinate system as in Fig. 7, where

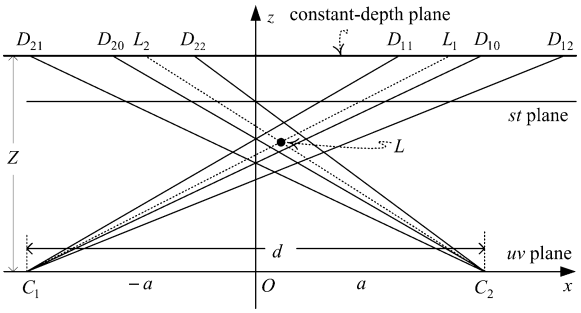

Figure 7. Rendering a point L with a light field.

1. $L = (x_0, z_0)$ is a point in the scene,
2. C_1 and C_2 are two adjacent cameras, and
3. C_1D_{10} and C_2D_{20} are two rays in the light slab that are nearest to C_1L_1 and C_2L_2, respectively.

The global constant-depth plane, at a depth Z, is parallel to but might not be identical to the focal plane. This resembles a dynamically reparameterized light field (Isaksen et al., 2000). We omit the projection from the constant-depth plane to the focal plane of the novel views as they are parallel.

We now examine the intensity contribution of L on the constant-depth plane in Fig. 7. The details are shown in Fig. 8, where

1. x_i is the coordinate of L_i $(i = 1, 2)$ on the constant-depth plane,
2. $x_i + e_i$ is the coordinate of D_{i0}, where C_iD_{i0} is the nearest sampling ray to C_iL viewed from C_i, and
3. $x_i + e_i \pm \Delta$ are the coordinates of D_{i1} and D_{i2}, respectively,

in which e_i is the offset between L_i and D_{i0}, and $\Delta = \delta Z$ is the sample spacing on the constant-depth plane.

To avoid double images, the two wedge-shape intensity contributions of the point L must overlap. Hence, D_{11} must be on the left of D_{22}, or

$$x_1 + e_1 - \Delta \le x_2 + e_2 + \Delta, \tag{1}$$

It is easy to compute that the coordinates of L_1 and L_2 are:

$$x_1 = -a + Z(x_0 + a)/z_0, \quad \text{and}$$
$$x_2 = a + Z(x_0 - a)/z_0,$$

respectively. Therefore, (1) becomes

$$\frac{2a(Z - z_0)}{z_0} \le 2\Delta + (e_2 - e_1). \tag{2}$$

Since the position of L is arbitrary, $e_2 - e_1$ can vary between $-\Delta$ and Δ. Therefore the following condition must be satisfied:

$$\frac{2a(Z - z_0)}{z_0} \le \Delta. \tag{3}$$

The above condition is deduced when $z_0 \le Z$. If $z_0 > Z$, the corresponding condition is

$$\frac{2a(z_0 - Z)}{z_0} \le \Delta.$$

Summing up, the sample spacing d must satisfy:

$$d = 2a \le \Delta \cdot \frac{z_0}{|z_0 - Z|} = \delta \cdot \frac{z_0 Z}{|z_0 - Z|}. \tag{4}$$

If z_0 is bounded between $z_{\min}(\le Z)$ and $z_{\max}(\ge Z)$, then the minimum value of the right hand side of (4) is the maximum allowable distance between two

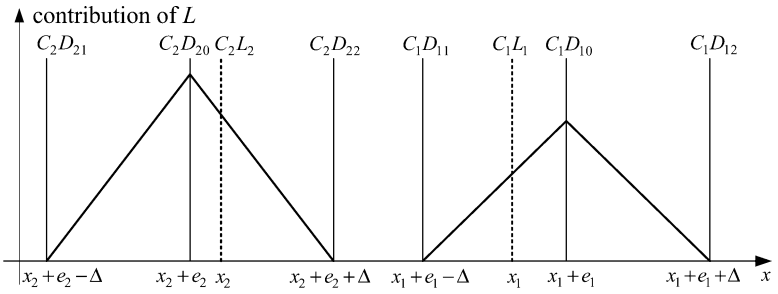

Figure 8. The details of the intensity contributions of L on the constant-depth plane. The horizontal axis indicates the ray position and the vertical axis represents the intensity value of the rays.

locations of the camera, namely

$$d_{\max} = \delta \min_{z_{\min} \le z_0 \le z_{\max}} \left\{ \frac{z_0 Z}{|z_0 - Z|} \right\}$$
$$= \delta Z \cdot \min\left\{ \frac{z_{\min}}{Z - z_{\min}}, \frac{z_{\max}}{z_{\max} - Z} \right\}. \quad (5)$$

We may rewrite (4) as:

$$\left| \frac{1}{z_0} - \frac{1}{Z} \right| d \le \delta.$$

This means that when Z is the estimate of exact depth, the disparity (Xu and Zhang, 1996)[1] error viewed from nearby cameras must not exceed *one* pixel. Since (5) is the minimization of (4) over all scene points, it means that if the globally constant depth Z is chosen for the scene, then the sample spacing should ensure that the disparity errors of all scene points between successive views must not exceed one pixel.

4.2. The Optimal Constant Depth

Equation (5) indicates that $d_{\max}$ is maximized when $Z = Z_{\text{opt}}$ satisfies:

$$\frac{z_{\min}}{Z_{\text{opt}} - z_{\min}} = \frac{z_{\max}}{z_{\max} - Z_{\text{opt}}},$$

or

$$Z_{\text{opt}} = \frac{2 z_{\min} z_{\max}}{z_{\min} + z_{\max}}. \quad (6)$$

One can easily check that $z_{\min} \le Z_{\text{opt}} \le \frac{1}{2}(z_{\min} + z_{\max})$, and the equalities hold only for $z_{\min} = z_{\max}$. This choice of Z_{opt} is reasonable because closer objects require more accurate depth information.

One should be cautious to provide relatively accurate minimum and maximum depths so that the computed optimal constant depth can really take effect. Fortunately, it is possible to estimate or measure them in practice. Moreover, Z_{opt} is less sensitive to $z_{\max}$ than to $z_{\min}$.

4.3. Interpretation of the Optimal Constant Depth

It is interesting to note that (6) can be rewritten as:

$$\frac{1}{Z_{\text{opt}}} = \frac{1}{2}\left(\frac{1}{z_{\min}} + \frac{1}{z_{\max}} \right).$$

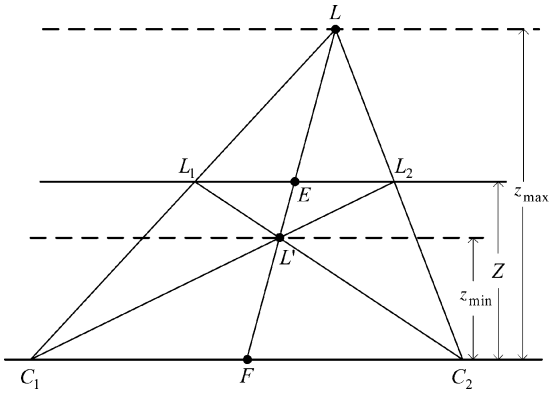

Figure 9. Graphical determination of the optimal constant depth.

In this formulation, the optimal constant depth is exactly the harmonic mean of the minimum and maximum depths. Equation (6) can also be written as:

$$\frac{1}{z_{\min}} - \frac{1}{Z_{\text{opt}}} = \frac{1}{Z_{\text{opt}}} - \frac{1}{z_{\max}}.$$

This implies that the nearest and farthest objects can be rendered with equal sharpness.

The optimal constant depth can be determined graphically. Referring to Fig. 9, where

1. C_i $(i = 1, 2)$ are on the camera plane,
2. F is the mid-point of C_1C_2,
3. L and L' are the farthest and the nearest points in the scene, and
4. C_2L' intersects C_1L at L_1 and C_1L' intersects C_2L at L_2.

Then it is guaranteed that L_1L_2 is parallel to C_1C_2. From projective geometry (Ayres, 1967), one can prove that $\{L,\ E,\ L',\ F\}$ is a harmonic set of points, where E is the intersection point of LL' and L_1L_2. Therefore

$$\frac{|EL| \cdot |FL'|}{|FL| \cdot |EL'|} = 1,$$

which gives

$$\frac{(z_{\max} - Z) z_{\min}}{z_{\max}(Z - z_{\min})} = 1.$$

Thus this is a convenient way of finding the optimal constant depth Z_{opt}.

With the optimal constant depth, the sample spacing in (5) becomes:

$$d_{\max} = \frac{2\delta}{z_{\min}^{-1} - z_{\max}^{-1}}. \quad (7)$$

It is totally determined by the disparity variation of the scene and the camera resolution. The above equality can also be written as:

$$\left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}}\right) d_{\max} = 2\delta.$$

This means the maximum allowed sample spacing should make the disparity variation of the scene between successive cameras be 2 pixels.

4.4. *Prefiltering the Light Field*

Let the sample spacing be d. If this sampling rate is inadequate, i.e. $d > d_{\max}$, where $d_{\max}$ is given by (5), then it is necessary to prefilter the light field in order to eliminate the artifact of double images. As mentioned by Levoy and Hanrahan (1996), the prefiltering can be done on the camera or focal plane, or both. Filtering on the focal plane reduces the image resolution, whereas filtering on the camera plane reduces the depth of focus (Isaksen et al., 2000). It is easy to see that the size for focal-plane prefiltering should be $d/d_{\max}$ pixels, and the focal-plane prefiltering can be done more conveniently by postfiltering on the rendered images. However, in theory the camera-plane prefiltering cannot be effective because the samples are not taken after low-pass filtering the camera plane.

4.5. *Disparity-Based Analysis*

In fact, the minimum sampling rate can be found by simply observing the maximum and minimum disparities in pixels in two captured images, without any measurement over the scene or any camera calibration. So do rendering and prefiltering.

Suppose the maximum and minimum disparities found in two images are $N_{\max}^D$ and $N_{\min}^D$ pixels, respectively. From Section 4.3, we know that the sampling rate must make the maximum disparity difference in the scene be 2 pixels. Because the disparity variation is proportional to the sample spacing, the sampling interval should be shrunk by $(N_{\max}^D - N_{\min}^D)/2$ times so that the disparity variation is 2 pixels. Therefore $N = \text{ceil}((N_{\max}^D - N_{\min}^D)/2) + 1$ sample images are required for the interval between the two cameras.

The rendering can also be done conveniently with epi-polar images (Chai et al., 2000), where a global motion compensation vector is needed for picking out appropriate pixels in different sample images and blending among them. The optimal constant depth now corresponds to the optimal motion compensation vector, which is $(N_{\min}^D + N_{\max}^D)/(2(M-1))$ pixels between successive sample images, where M is the number of sample images uniformly taken in the interval.

Finally, the size of prefiltering is $(N_{\max}^D - N_{\min}^D)/(2(M-1))$ pixels.

4.6. *Experiments*

We now verify our analysis in previous sections. In our "Toy" light field, we capture two images with a large distance between them, and find that the minimum and maximum disparities are 27 and 59 pixels, respectively. Then we know that $(59-27)/2+1 = 17$ images are required for the interval. The relevant data are listed in Table 1. Figure 10(a) is one of the sample images. In order to detect the existence of double images, sharp features, which appear as thin vertical lines, are added to the scene. The upper and lower boxes are at the maximum and minimum depths, respectively.

First, we test the effectiveness of the optimal constant depth. With the optimal constant depth chosen at $1.3721 z_{\min}$,[2] the light field can be correctly rendered with equal sharpness at $z_{\min}$ and $z_{\max}$ (Fig. 11(b)). If the optimal constant depth is not chosen, e.g., if the mean depth at $1.5926 z_{\min}$ is used instead, the scene points at $z_{\min}$ appear to be very blurry (Fig. 11(c)), though those at $z_{\max}$ look sharper.

Second, we test the sufficiency of the minimum sampling rate. When the sampling rate doubles, the visual improvement is nearly unnoticeable (Fig. 11(b) and (d)). However, when the sampling rate is reduced by half, double images are clearly visible at both $z_{\min}$ and $z_{\max}$ (Fig. 11(e)).

Third, we test the effectiveness of the prefiltering size. The double images in Fig. 10(e) become invisible when a prefiltering size of 2 pixels on the focal plane

Table 1. The data for the light field "Toy".

st sampling rate	$N_{\min}^D$	$N_{\max}^D$	Minimum sampling rate	Optimal const. depth
610×455	27	59	17×17	$1.3721 z_{\min}$

Figure 10. The sample and rendered images of the light field "Toy". (a) A sample image of the light field "Toy". The vertical lines in the white boxes are for the detection of double images. (b) The rendered image from 17 × 17 images. (c) The same as (b) with the mean constant depth chosen. (d) Rendered from 30 × 30 images. (e) Rendered from 9 × 9 images. (f) Image (e) prefiltered with size of 2 pixels. (b), (d)–(f) all use the optimal constant depth.

is chosen (Fig. 11(f)). However, the rendered image becomes blurred.

5. Minimum Sampling Curve of the Light Field

Now we study how geometry information can help reduce the sampling rate. The theory is based on the analysis on the minimum sampling rate. We find that the sampling rate is inversely proportional to both geometry information and camera resolution. We differentiate whether the approximate geometry is smooth or not. When the approximate geometry contains steep jumps, our results are different from those in Chai et al. (2000) because we consider the artifact caused by the depth discontinuity.

5.1. *Sample Spacing and Depth Uncertainty*

From Section 4.1, we know that in order to properly render a scene point with depth estimated between z_{near} and z_{far}, the sample spacing must satisfy

$$\frac{1}{z_{\text{near}}} - \frac{1}{z_0} \le \frac{\delta}{d}, \quad \text{and} \quad \frac{1}{z_0} - \frac{1}{z_{\text{far}}} \le \frac{\delta}{d},$$

where z_0 is the exact depth of the point. Therefore,

$$\frac{2\delta}{d} \ge \frac{1}{z_{\text{near}}} - \frac{1}{z_{\text{far}}}. \tag{8}$$

On the other hand, if (8) is satisfied, then the scene point can be properly rendered by choosing the optimal depth

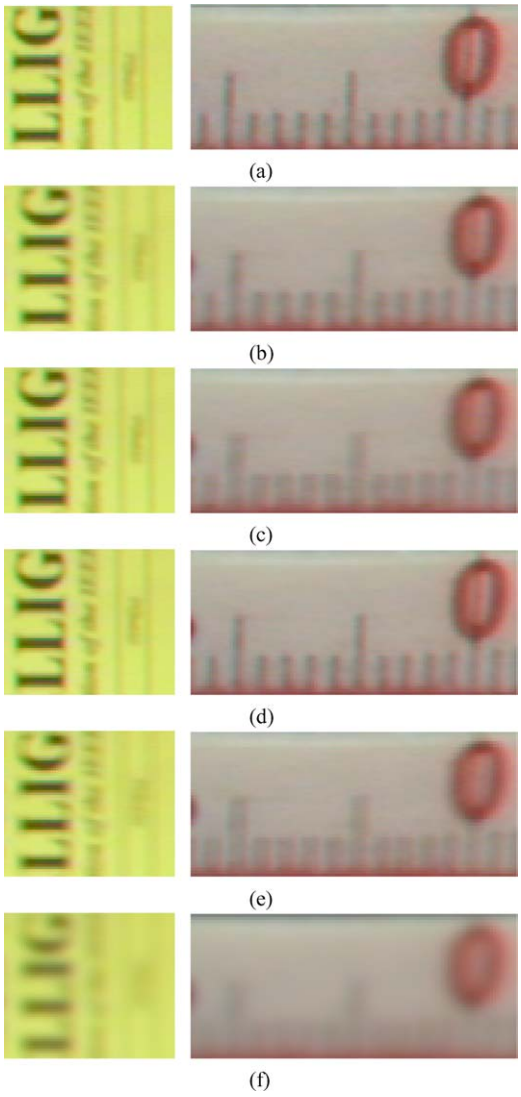

Figure 11. Blow-up of the sampling and rendered images of the light field "Toy".

z_{opt} at

$$\frac{2}{z_{\text{opt}}} = \frac{1}{z_{\text{near}}} + \frac{1}{z_{\text{far}}}.$$

Summing up, the maximum allowable sample spacing should satisfy:

$$\frac{2\delta}{d} = \frac{1}{z_{\text{near}}} - \frac{1}{z_{\text{far}}},$$

or the disparity difference between z_{near} and z_{far} is 2 pixels. In this case, the disparity differences between z_{near} and z_{opt} as well as z_{opt} and z_{far} are both 1 pixel.

We may define $z_{\text{near}}^{-1} - z_{\text{far}}^{-1}$ as the geometric uncertainty of a scene point and the geometric uncertainty of a scene is the maximum uncertainty of all scene points. Then it is clear that the sample spacing is inversely proportional to both the uncertainty of geometry information and the camera resolution, as shown in Fig. 12.

This analysis is valid only under the premise that the approximate geometry is smooth, or having no steep depth jumps, so that the depth along the novel viewing ray is well defined. If the jumps exist, another aliasing, which we call the texture discontinuity, may occur. Here we explain how this happens. In Fig. 13, the object surface is approximated by a local geometry with a steep jump, indicated by a dashed line segment, and C_i $(i = 1, 2)$ are two nearby cameras. P_0 is a point on

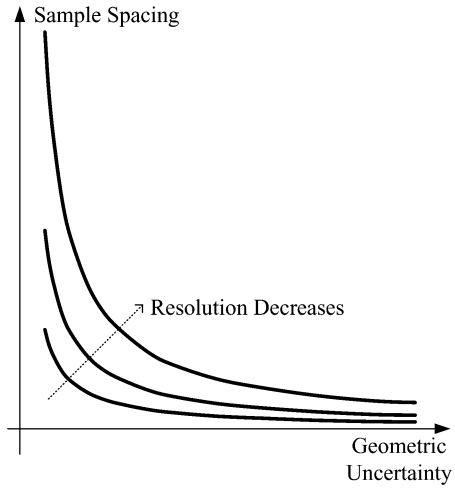

Figure 12. The curve of sample spacing as a function of geometric uncertainty when the approximate geometry is smooth.

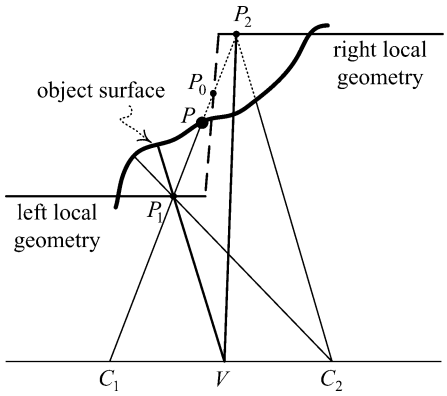

Figure 13. The discontinuity in approximate geometry may cause the discontinuity in texture.

the jump and P is the surface point on the ray C_1P_0. P_1 and P_2 are the intersection points of C_1P_0 with the approximate geometry. Then P might also appear twice in a novel view V, at the rays VP_1 and VP_2, respectively. If double images of P appear, the texture of object surface will become discontinuous. To eliminate such discontinuity, the distance between the two rays on the focal plane should be less than *one* pixel. As V is free between C_1 and C_2, this implies that the disparity difference between P_1 and P_2, viewed between C_1 and C_2, must not exceed one pixel.

Taking the texture discontinuity into account, when the approximate geometry is not smooth, the sampling rate also has to make all the disparity differences of depth jumps within one pixel.

5.2. Minimum Sampling Curve

To date, the exact depth is still hard to compute with computer vision techniques, even from a sequence of images. Fortunately, for rendering purpose, only approximate depth is necessary. As shown in the last subsection, every depth has a certain degree of tolerance on the error. If the sampling rate is already fixed, a globally constant depth may not be sufficient for all objects in the environment. We have to choose several constant depths so that all objects can be properly rendered using an appropriate depth. The concept of depth layers has appeared in other papers, such as Shade et al. (1998) and Isaksen et al. (2000). Depth layers serve for depth quantization. They are very useful in depth recovery. Often, the number of depth layers is small, which makes sweeping on the depth efficient. In this subsection we derive the minimum sampling curve that describes the relationship between the number of images and the number of depth layers.

Suppose the minimum and maximum depths of the scene are $z_{\min}$ and $z_{\max}$, respectively. Then between two nearby cameras the total disparity variation is $(d/\delta)(z_{\min}^{-1} - z_{\max}^{-1})$ pixels, where d is the sample spacing. As the representative depths are discontinuous, from the conclusion in last subsection, the disparity difference between successive representative depths should be 1 pixel. Each representative depth should be best chosen as the harmonic mean of the nearest and the farthest depths of each depth layer. Therefore, the disparity difference between each representative depth and the nearest or farthest depth of the depth layer should be a half pixel. However, the nearest and the farthest representative depths are the exception because there are no depth layers before or after them. Hence the disparity difference between the nearest depth and the first representative depth can be 1 pixel. Same for the farthest layer. Therefore, if we want to divide the scene into N layers, then it allows $N + 1$ pixels of disparity variation in total. Hence the optimal sample spacing should satisfy:

$$\frac{\delta}{d_{\text{opt}}} = \frac{1}{N+1}\left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}}\right). \tag{9}$$

This gives the minimum sampling curve, which depicts the relation among the number of sample images, the number of depth layers and the camera resolution. This curve resembles that in Fig. 12.

Equation (9) is different from the result of Chai et al. (2000), which gives:

$$\frac{\delta}{d_{\text{opt}}} = \frac{1}{2N}\left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}}\right). \tag{10}$$

According to our analysis, the sampling rate given by the latter can only eliminate double images *inside* the constant-depth planes. This sampling rate is inadequate for eliminating the texture discontinuity *across* the depth planes. Note that Eqs. (9) and (10) are the same when $N = 1$.

5.3. Decomposition of Depth Layers

In this subsection, we study how the depth layers should be decomposed. Suppose the representative depths are Z_i^r $(i = 0, 1, \ldots, N-1)$. Each Z_i^r is the representative depth of objects whose depth is between z_i and z_{i+1}. Obviously, $z_0 = z_{\min}$.

We now choose Z_i^r and z_{i+1} recursively. The disparity difference between z_0 and Z_0^r is one pixel. So

$$\frac{1}{Z_0^r} = \frac{1}{z_0} - \frac{\delta}{d_{\text{opt}}}.$$

For intermediate depth layers, the disparity difference between z_i and Z_i^r and z_{i+1} and Z_i^r are both a half pixel. So

$$\frac{1}{Z_i^r} = \frac{1}{z_i} - \frac{\delta}{2d_{\text{opt}}}, \quad (i = 1, 2, \ldots, N-1)$$

$$\frac{1}{z_{k+1}} = \frac{1}{Z_k^r} - \frac{\delta}{2d_{\text{opt}}}, \quad (k = 0, 1, \ldots, N-2)$$

Therefore,

$$\frac{1}{z_{i+1}} = \frac{1}{z_i} - \frac{\delta}{d_{\text{opt}}} = \frac{1}{z_1} - \frac{i\delta}{d_{\text{opt}}}, \quad (i = 0, 1, \cdots, N-2)$$

$$\frac{1}{Z_k^r} = \frac{1}{z_{k+1}} + \frac{\delta}{2d_{\text{opt}}} = \frac{1}{z_1} - \frac{(k-0.5)\delta}{d_{\text{opt}}}, \quad (k = 0, 1, \ldots, N-1)$$

$$\frac{1}{z_1} = \frac{1}{Z_0^r} - \frac{\delta}{2d_{\text{opt}}}.$$

Finally,

$$\begin{aligned} \frac{1}{z_i} &= \frac{1}{z_{\min}} - \frac{(i+0.5)\delta}{d_{\text{opt}}}, \quad (i = 1, \ldots, N-1) \\ \frac{1}{Z_k^r} &= \frac{1}{z_{\min}} - \frac{(k+1)\delta}{d_{\text{opt}}}, \quad (k = 0, 1, \ldots, N-1). \end{aligned} \tag{11}$$

5.4. *Experiments*

Since ground truth of the depth information is needed to verify our analysis, we perform some synthetic experiments. The relevant data of the light field "Outdoor" are listed in Table 2. As the depth variation is large, the minimum sampling rate is quite high (364×364). However, adding geometry information can help reduce the sampling rate. According to our analysis, using 45, 90, and 181 depth layers requires only 17×17, 9×9 and 5×5 images, respectively. Note that the analysis in Chai et al. (2000) claims that only 23, 46 and 91 depth layers are sufficient for 17×17, 9×9 and 5×5 samples, respectively. The solid line in Fig. 14 shows our minimum sampling curve for this light field, whereas the dashed line is the one predicted by Chai et al. (2000). Our experiments can show that our analysis is more accurate. In Fig. 15, the rendering results are shown, where from left to right the sampling rate increases from 5×5 to 17×17, and from top to bottom the depth layers decrease from infinity to 24. Due to the occlusion in the scene, bilinear interpolation is not used if the depth difference of two interpolation rays is too large, and the holes are preserved if the corresponding areas are not visible to both nearby views.

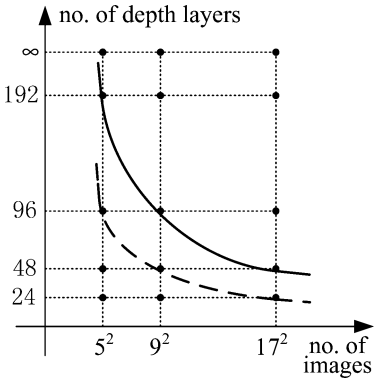

Figure 14. The minimum sampling curves of the light field "Outdoor". The upper solid curve is our minimum sampling curve, while the lower dashed one is that from Chai et al. (2000). The rendering results with different numbers of depth layers and sample images, indicated by the dots, are shown in Fig. 15 at the same relative positions.

It can be seen that:

1. Below our minimum sampling curve (Fig. 15(g), (j), (k), and (m)–(o)), texture discontinuity is discernible in these images even though some (Fig. 15(g), (k) and (o)) meet the minimum sampling curve in Chai et al. (2000), e.g., the horizontal streaks on the water or the soil (Fig. 16(g), (j), (k), and (m)–(o)).
2. On or above our minimum sampling curve (Fig. 15 (d)–(f), (h), (i), and (l) and Fig. 16(d)–(f), (h), (i), and (l)), the artifacts that are not caused by occlusion are undetectable and the rendering quality is nearly identical to that using exact depth (Figs. 15(a)–(c) and 16(a)–(c)).
3. Along the curve, the rendering quality is nearly unchanged (Figs. 15(d), (h), (l) and 16(d), (h), (l)), if the artifacts caused by occlusion are ignored. This fully demonstrates the tradeoff between sampling rate and geometric accuracy.

Another important observation is that for eliminating occlusion artifacts the sampling rate is more important

Table 2. The data for the light field "Outdoor".

st sampling rate	*st* size	*uv* size	uv-st dist.	$z_{\min}$	$z_{\max}$	Minimum sampling rate	Depth layers needed given sampling rate		
							5×5	9×9	17×17
480×480	100×100	50×50	100	31.27	573.77	364×364	181	90	45

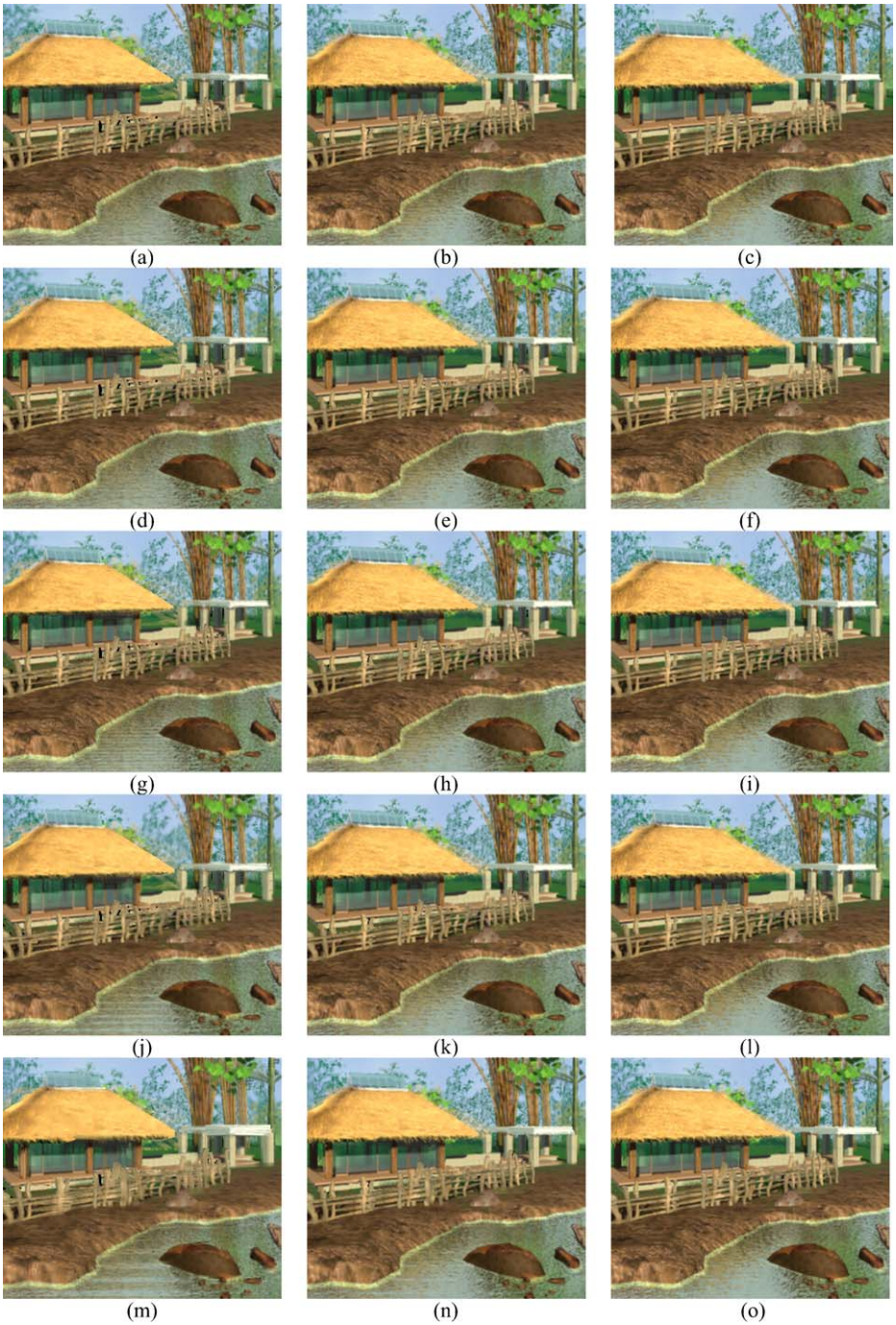

Figure 15. The rendered images of the light field "Outdoor". The first, second and third columns are rendered from 5×5, 9×9 and 17×17 images, respectively. The first row uses exact depth. The second to fifth rows use 192, 96, 48, and 24 depth layers, respectively.

than the geometric information. If the sampling rate is inadequate, the occlusion artifacts cannot be eliminated by adding the geometric information, not even with exact geometry (Fig. 15(a) and (b)).

6. Dealing with Occlusion

In previous sections, the scene is always assumed to be occlusion-free. In a real scene, occlusion

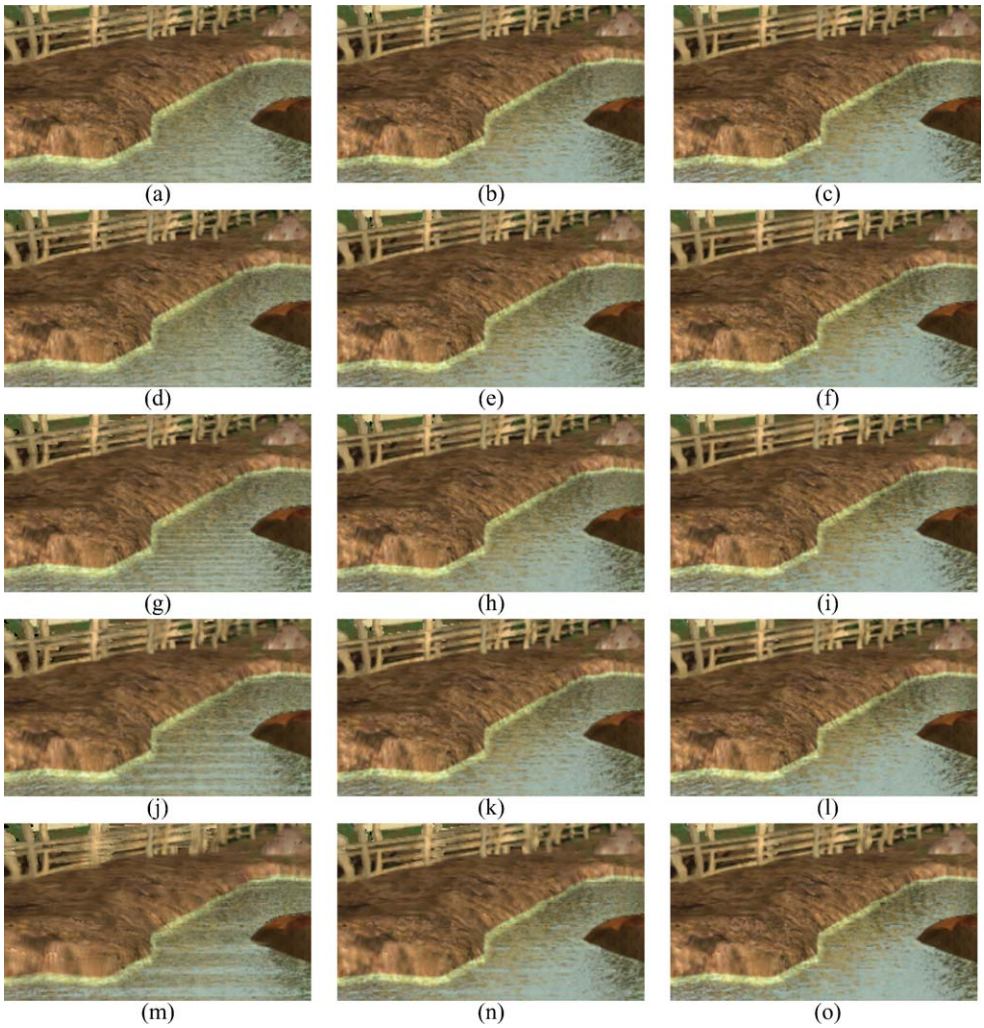

Figure 16. Blown-up views of corresponding rendered images in Fig. 15.

always exists and sampling with occlusion must be analyzed.

Occlusion destroys the desirable scene property that all scene points are visible from all views. It is easy to see that in general the spectrum of a light field with occlusion is *not* bandlimited, thus demanding the reconstruction of completely artifact-free (in the sense of signal reconstruction) light field from discrete samples is impossible for a complex scene. Therefore, the minimum sampling rate of a scene with occlusion can only be found according to how much aliasing is tolerable. In our experiments, we find that there is no visually unacceptable artifacts when scenes with occlusion are rendered at the minimum sampling rate for occlusion-free scenes. Thus we recommend that all scenes with the same disparity variation share the same minimum sampling rate.

Unfortunately, the minimum sampling curve no longer exists for scenes with occlusion. As shown in Fig. 17, under the Lambertian premise, if there is no occlusion in the scene, interpolating C_1P_1 with C_2P_2 is equivalent to interpolating C_1P_1 and C_3P_2, or C_4P_2, etc., because the pixel values of rays C_2P_2, C_3P_2, etc., are identical. This is the basis for different combinations of sampling rate and geometry information producing the same rendering result. On the other hand, if occlusion exists, the pixel values of rays C_2P_2, C_3P_2, etc., may not be identical. Therefore interpolating C_1P_1 with C_2P_2 is not equivalent to interpolating C_1P_1 and C_3P_2, or C_4P_2, etc. As a

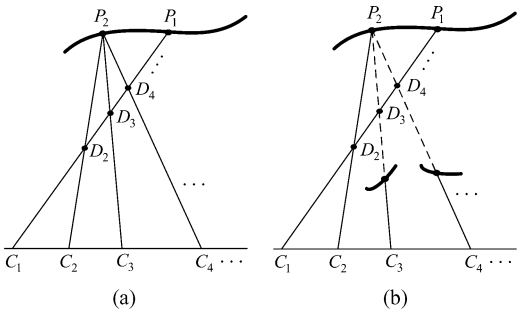

Figure 17. Only in an occlusion-free scene can the sampling rate be traded for geometry information. (a) When the scene is occlusion-free and Lambertian, the interpolation between C_1P_1 and C_2P_2 is equivalent to those between C_1P_1 and C_3P_2, and between C_1P_1 and C_4P_2, etc. Therefore different combinations of sampling rate and geometry information can give the same rendering result. (b) When occlusion exists, interpolating C_1P_1 with C_2P_2 is not equivalent to interpolating C_1P_1 and C_3P_2, or C_4P_2, etc. Therefore different combinations of sampling rate and geometry information cannot produce the same rendering result.

result, the sampling rate cannot be traded for geometry information.

7. Conclusion

In this paper, we have presented a geometry-based analysis of the minimum sampling rate and minimum sampling curve for light field/Lumigraph rendering. Our analysis provides guidance on how to capture and prefilter the light field effectively. We equate acceptable rendering quality with the absence of any double image artifacts. This criterion is equivalent to the anti-aliasing condition in spectral analysis and can also be based on human perception and the causality principle in scale-space theory. By examining each individual point in a scene, we infer that the two intensity contributions of a point on the focal plane of the novel view must at least touch each other. We conclude that the minimum sampling rate is determined by the camera resolution and the scene disparity variation. In addition, an optimal globally constant depth for the best rendering quality is found to be the harmonic mean of the maximum and minimum depths.

The minimum sampling curve quantitatively shows how reducing the geometry information can be compensated for by increasing the number of images. We define the geometric uncertainty of a scene as the maximum disparity uncertainty of each scene point. The sample spacing is found to be inversely proportional to both the camera resolution and the scene geometric uncertainty. Our results are more accurate than those in plenoptic sampling (Chai et al., 2000) because the depth discontinuity in the approximate geometry is considered.

While the experimental results are encouraging, the sampling rate would be further reduced if the characteristics of human vision are considered. For example, manifold hopping (Shum et al., 2001) can greatly reduce the size of the input database. Finally, studying more thoroughly the sampling problem of a scene with occlusion is another challenging task.

Appendix A: Anti-Aliasing Condition in Terms of Geometry

In this appendix, we prove that eliminating double images in the geometric viewpoint is the anti-aliasing condition in the viewpoint of signal processing.

First, we show the equivalence between the constant overlap of the wedge-shape intensity contributions in discrete images and the constant overlap of those in continuous ones (the parabola-like curve in Fig. 3(b)). On one hand, it is apparent that the overlap of continuous intensity contributions guarantees the overlap of discrete intensity contributions (Fig. 18(a)). On the

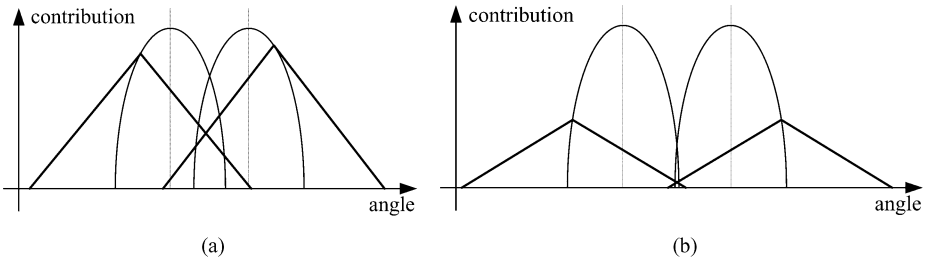

Figure 18. Equivalence between types of overlap. (a) The overlap of parabola-like intensity contributions ensures the overlap of wedge-shape ones. (b) The overlap of wedge-shape intensity contributions at an extreme case implies the overlap of parabola-like ones.

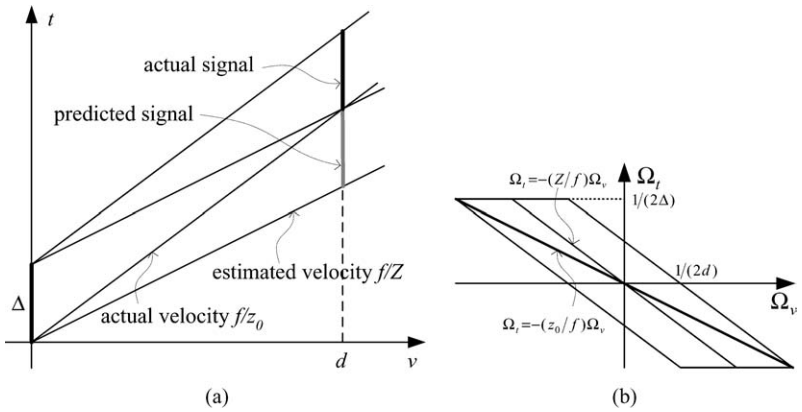

Figure 19. The anti-aliasing conditions in the spatial-temporal domain and spectral domain. (a) In the spatial-temporal domain, the condition is the overlap of parabola-like intensity contributions. (b) In the frequency domain, the condition is that the motion-compensated reconstruction filter contain the entire spectrum of the video.

other hand, as shown in Fig. 3(b), if the offset ε of one of the intensity contributions is close to $-\delta/2$ and the other offset is close to $\delta/2$, then the overlap of these two wedge-shape intensity contributions requires that the two parabola-like ones at least touch each other (Fig. 18(b)). In fact, the transfer from (2) to (3) in Section 4.1 has demonstrated such equivalence.

Next, we prove that the constant overlap of parabola-like intensity contributions is equivalent to the anti-aliasing condition in the frequency domain. To show this, we borrow the framework of video processing because the relevant results are already well known in video processing and we will not repeat their deductions.

For a point at depth z_0, its perceived velocity in the video is f/z_0 (Xu and Zhang, 1996), where f is the focal length. Using a constant depth Z is equivalent to estimating its velocity at f/Z. The parabola-like intensity contributions from other frames backprojected from the constant-depth plane to the current frame are the predicted positions of the point, using the estimated velocity f/Z. So the actual intensity contribution and the predicted one must overlap. From Fig. 19(a), in spatial-temporal domain, the "time" duration d must satisfy

$$d\,|\,f/z_0 - f/Z\,| \le \Delta, \tag{12}$$

in order to ensure the overlap, where $\Delta = \delta f$ is the sample spacing on the focal plane.

Now let's consider the anti-aliasing condition in the frequency domain. The 2D light field is parameterized by v and t, where v is the "virtual time". Since the actual velocity is constant, the spectrum of the point is simply a slant line segment, with the highest frequency in t being $1/(2\Delta)$ (Tekalp, 1995). Using the estimated velocity is equivalent to using a motion-compensated reconstruction filter to reconstruct the video. The filter is a parallelogram shown in Fig. 19(b). Then the anti-aliasing condition is that the spectrum of the video must completely lay inside the parallelogram, or

$$1/(2d) \ge 1/(2\Delta)\,|f/z_0 - f/Z|.$$

We see that the above inequality is identical to (12).

Acknowledgment

The authors would like to thank Dr. Xin Tong for his generous help in preparing for the experiments. Dr. Sing Bing Kang and Dr. Steve Lin's careful revision and proof-reading is also greatly appreciated.

Notes

1. We also use "disparity" for the abbreviation of "the reciprocal of depth" if no unit of "pixel" follows.
2. The optimal motion compensation vector is $(59+27)/(2\times(17-1)) = 2.6875$ pixels between successive sample images. The real value of $z_{\min}$ is insignificant.

References

Adelson, E.H. and Bergen, J.R. 1991. The plenoptic function and the elements of early vision. In *Computational Models of Visual*

Processing, M.S. Landy and J.A. Movshon (Eds.), MIT Press, chap. 1, pp. 3–20.

Ayres, F. Jr. 1967. *Theory and Problems of Projective Geometry*. McGraw-Hill Inc.

Buehler, C., Bosse, M., McMillan, L., Gortler, S., and Cohen, M. 2001. Unstructured lumigraph rendering. In *SIGGRAPH 2001 Conference Proceedings*, Annual Conference Series, Aug. 2001, pp. 425–432.

Camahort, E., Lerios, A., and Fussell, D. 1998. Uniformly sampled light fields. In *Proceedings of 9th Eurographics Workshop on Rendering*, pp. 117–130.

Chai, J.-X., Tong, X., Chan, C.-C., and Shum, H.-Y. 2000. Plenoptic sampling. In *SIGGRAPH 2000 Conference Proceedings*, Annual Conference Series, July 2000, pp. 307–318.

Chen, W.-C., Bouguet, J.-Y., Chu, M.H., and Grzeszczuk, R. 2002. Light field mapping: Efficient representation and hardware rendering of surface light fields. In *SIGGRAPH 2002 Conference Proceedings*, Annual Conference Series, Aug. 2002, pp. 447–456.

Chen, S.E. and Williams, L. 1993. View interpolation for image synthesis. In *SIGGRAPH 1993 Conference Proceedings*, J.T. Kajiya, (Ed.), vol. 27 of *Annual Conference Series*, Aug. 1993, pp. 279–288.

Gortler, S.J., Grzeszczuk, R., Szeliski, R., and Cohen, M.F. 1996. The lumigraph. In *SIGGRAPH 1996 Conference Proceedings*, Annual Conference Series, Aug. 1996, pp. 43–54.

Halle, M. 1994. Holographic stereograms as discrete imaging systems. In *Practical Holography VIII*, volume Proc. SPIE Vol. 2176, pp. 73–84.

Hlaváč, V., Leonardis, A., and Werner, T. 1996. Automatic selection of reference views for image-based scene representations. In *Proc. of ECCV'96*, pp. 526–535.

Isaksen, A., McMillan, L., and Gortler, S.J. 2000. Dynamically reparameterized light fields. In *SIGGRAPH 2000 Conference Proceedings*, Annual Conference Series, July 2000, pp. 297–306.

Levoy, M. and Hanrahan, P. 1996. Light field rendering. In *SIGGRAPH 1996 Conference Proceedings*, Annual Conference Series, Aug. 1996, pp. 31–42.

Lin, Z. and Shum, H.-Y. 2000. On the number of samples needed in light field rendering with constant-depth assumption. In *Proceedings of IEEE Computer Vision and Pattern Recognition 2000 (CVPR 2000) Conference*, Hilton Head Island, South Carolina, USA, June 2000, pp. 588–579.

Nishino, K., Sato, Y., and Ikeuchi, K. 2001. Eigen-texture method: Appearance compression and synthesis based on a 3d model. *IEEE Transactions on Pattern Pattern Analysis and Machine Intelligence*, 23(11):1257–1265.

Romeny, B. (Ed.) 1994. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers: Netherlands.

Schirmacher, H., Heidrich, W., and Seidel, H. 1999. Adaptive acquisition of lumigraphs from synthetic scenes. In *Proceedings of Eurographics'99*, pp. 151–159.

Shade, J., Gortler, S., He, L.-W., and Szeliski, R. 1998. Layered depth images. In *SIGGRAPH 1998 Conference Proceedings*, Annual Conference Series, July 1998, pp. 231–242.

Shum, H.-Y. and He, L.-W. 1999. Rendering with concentric mosaics. In *SIGGRAPH 1999 Conference Proceedings*, Annual Conference Series, Aug., pp. 299–306.

Shum, H.-Y., Wang, L., and Chai, J. 2001. Rendering by manifold hopping. In *SIGGRAPH 2001 Technical Sketch*, Aug. 2001, p. 253.

Sloan, P.P., Cohen, M.F., and Gortler, S.J. 1997. Time critical lumigraph rendering. In *Symposium on Interactive 3D Graphics*, pp. 17–23.

Tekalp, A.M. 1995. *Digital Video Processing*. Prentice Hall.

Wood, D.N. et al. 2000. Surface light fields for 3d photography. In *SIGGRAPH 2000 Conference Proceedings*, Annual Conference Series, pp. 287–296.

Wu, M. and Shum, H.-Y. 2000. Real-time stereo rendering of concentric mosaics with linear interpolation. In *IEEE/SPIE Visual Communications and Image Processing (VCIP) 2000*, June 2000, pp. 23–30.

Xu, G. and Zhang, Z. 1996. *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*. Kluwer Academic Publishers.

Zhang, C. and Li, J. 2000. Compression of lumigraph with multiple reference frame (mrf) prediction and just-in-time rendering. In *IEEE Data Compression Conference*, pp. 253–262.