On the Number of Samples Needed in Light Field Rendering With Constant-depth Assumption

Zhouchen Lin *
Peking University, China

Heung-Yeung Shum
Microsoft Research, China

Abstract

While several image-based rendering techniques have been proposed to successfully render scenes/objects from a large collection (e.g., thousands) of images without explicitly recovering 3D structures, the minimum number of images needed to achieve a satisfactory rendering result remains an open problem. This paper is the first attempt to investigate the lower bound for the number of samples needed in the Lumigraph[4]/light field rendering[5].

To simplify the analysis, we consider an ideal scene with only a point that is between a minimum and a maximum range. Furthermore, constant-depth assumption and bilinear interpolation are used for rendering. The constant-depth assumption serves to choose "nearby" rays for interpolation. Our criterion to determine the lower bound is to avoid horizontal and vertical double images, which are caused by interpolation using multiple nearby rays. This criterion is based on the causality requirement in scale-space theory, i.e., no "spurious details" should be generated while smoothing. Using this criterion, closed-form solutions of lower bounds are obtained for both 3D plenoptic function (Concentric Mosaics[8]) and 4D plenoptic function (light field). The bounds are derived completely from the aspect of geometry and are closely related to the resolution of the camera and the depth range of the scene. These lower bounds are further verified by our experimental results.

1 Introduction

A number of image-based rendering techniques have been proposed to render a novel view from a collection of densely captured image sequences, without 3D reconstruction of the scene/object, or explicit feature correspondence. These techniques collect a set of samples of the plenoptic function[1], which describes the irradiance perceived from the observer's viewpoints. The original plenoptic function is 7-dimensional, defined as the intensity of light rays passing through the optical center at every location, at every possible viewing angle, for every wavelength and at any time. The dimensionality is reduced to five by ignoring time and wavelength[6]. By restricting the viewpoints or the objects inside a box, light field[5] and Lumigraph[4] reduce the dimension of sampling to four. Concentric mosaics further reduces the sampling dimension to three by constraining the viewpoints inside a 2D planar circle[8]. The simplest plenoptic function collected at a fixed viewpoint is a two dimensional panorama (e.g., [2, 9]).

These light field rendering techniques, however, failed to analyze how many images need to be captured in order to render a new view without noticeable artifacts. In other words, what is the lower bound or the minimum number of samples needed for light field rendering? Oversampling has become a common practice in light field capturing. It is important to study the light field sampling problem. First, the minimum sampling analysis serves as a guidance on how to capture the environment, e.g., how densely cameras should be spaced. The fewer images we capture, the less storage we need. Second, until the minimum sampling problem is addressed, compression ratio of light field images/data does not make much sense. It is possible to remove part of the redundancy from uniformly oversampled dense sequence of images by compression. There is, however, a difference between traditional image/video compression and light field compression. In light field compression, every new view, not every original image, needs to be reconstructed. Sampling analysis is tightly coupled with the rendering process itself.

Light field rendering techniques avoid explicit 3D construction of the environment because it is difficult to recover 3D depth accurately from images. However, using depth knowledge greatly improves the rendering quality and also affects minimum sampling rate. In our analysis, constant-depth assumption and bilinear interpolation are used to improve the quality of rendering. We also consider an ideal scene with only a single point that is between a minimum and a maximum depth.

Given the capturing and rendering camera resolutions, our criterion to determine the lower bound is to avoid horizontal and vertical double images, which are caused by interpolation using multiple nearby rays. This criterion is based on the causality requirement in scale-space theory, i.e., no "spurious details" should be generated while smoothing. Details of minimum sampling analysis is dependent on the captur-

*National Laboratory of Machine Perception, Peking University, Beijing 100871, P.R. China. This work was done at Microsoft Research China.

ing and rendering configurations of the light fields. In this paper, we study two possible configurations: 3D Concentric Mosaics and 4D light field. But the essential idea is also applicable to other capturing configurations. The problem of pruning the samples is left for future work.

The remainder of this paper is organized as follows. In Section 2, we set up the problem and introduce necessary assumptions in order to study the sampling problem. In Section 3, we present our methodology based on the causality principle in scale-space theory. Then we apply our analysis to rendering with Concentric Mosaics in Section 4, and light field in Section 5. Finally we conclude our paper with a discussion and future work.

2 The problem statement

Our theoretical analysis of minimum sampling for light field rendering is based on a number of reasonable assumptions about scene, camera, and interpolation methods. We assume that the ideal scene consists of only a point that is between a maximum distance and a minimum distance, captured by a camera that has a finite angular resolution δ. For minimum sampling analysis, a point is representative. The angular dimension of the point is sufficiently small compared to δ, but not zero.

2.1 Camera resolution

In our analysis, a pin-hole camera model is adopted. 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 at the optical center with a low-pass filter. The same filter is applied to all pixels. The shape of the filter is compactly supported, with the width of support being the angular resolution δ. As a result, the camera is taking samples of $\tilde{\mathcal{F}}$, the convoluted plenoptic function at the optical center. The value of a pixel is exactly the value of $\tilde{\mathcal{F}}$ at the direction linking the pixel and the optical center.

Throughout this paper, we will use angular resolution, vertical and horizontal, to facilitate the deduction. And we assume uniform angular resolution if the field of view of the capturing camera is small (e.g., a typical video camera has a fairly small FOV $\approx 40^\circ$). We also assume the resolution of the rendering camera is the same as that of the capturing camera. We always decouple the sampling into horizontal and vertical directions, so that the interpolation and illustration will be simple.

2.2 Bilinear interpolation with constant-depth assumption

In light field rendering, bilinear interpolation and constant-depth assumption are used to improve the quality of rendered images. With constant-depth assumption, all the objects seen by the camera are hypothesized on a simple surface, e.g., a cylinder or a plane. Constant-depth assumption

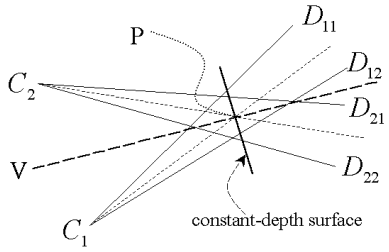

Figure 1: Bilinear interpolation with constant-depth assumption: the ray from V is interpolated using four rays from cameras at C_1 and C_2 that intersect near the point P on the constant-depth surface.

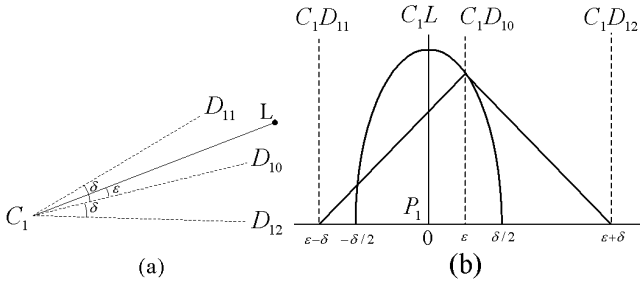

Figure 2: The intensity distribution of the ideal point changes from parabolic to wedge-like due to finite resolution of camera and linear interpolation.

is used to choose "nearby" rays in the ray space for rendering. In fact, any approach to find "nearby" rays inherently involves an assumption on depth. For example, interpolation using parallel rays assumes infinite depth. It has been shown in [10] that interpolation using constant-depth assumption yields better rendering results.

Figure 1 illustrates the concept of bilinear interpolation with constant-depth assumption. Suppose that we want to render a view at V, and the viewing ray hits the constant-depth surface at P. C_1 and C_2 are two nearby positions of the camera that are closest to VP, and C_iD_{ij} $(i,j=1,2)$ are nearby rays in camera C_i that are closest to the ray C_iP. The pixel value of VP is bilinearly interpolated, e.g., according to the angles $\angle D_{ij}C_iP$ and $\angle C_iPV$ $(i,j=1,2)$ [10].

To better understand sampling and linear interpolation, we redraw the rays in the camera coordinate, where the horizontal axis represents the angle of the ray shooting from the optical center. In Figure 2(a), camera C_1 is taking pictures 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. In Figure 2(b), the vertical line at 0 represents the ray C_1L, the parabola-like curve is the value of $\tilde{\mathcal{F}}$, or the intensity distribution of the point L in the continuous case. C_1D_{10} is by angle ε $(-\frac{\delta}{2} \le \varepsilon < \frac{\delta}{2})$ apart from C_1L, therefore the pixel value corresponding to the ray C_1D_{10} is just the value of $\tilde{\mathcal{F}}$

at ε. Since the angles of C_1D_{11} and C_1D_{12}, or $\varepsilon - \delta$ and $\varepsilon + \delta$ respectively, are outside of $[-\frac{\delta}{2}, \frac{\delta}{2})$, the intensity at the two corresponding pixels is zero. After linear interpolation, the intensity distribution of the point L will become 2δ wide and of wedge-shape.

3 Our sampling analysis methodology

3.1 When is rendering acceptable?

We now consider in what cases the rendered image is reasonable and acceptable. Note that bilinear interpolation is a good filtering choice because it preserves enough details and yet has sufficient width of support. But if the distance between successive camera locations is too wide apart, even bilinear interpolation introduces artifacts, appearing as double images. This phenomenon of double images was also observed by Levoy and Hanrahan [5] in light field rendering. Double images are the most salient and visually disturbing artifact when the sampling rate is too low. In our rendering experience, we even could hardly figure out or detect other types of artifacts with the presence of double images. Double images are caused by incorrect depth information used by corresponding interpolation rays. If the object is at a constant-depth, there should not be any artifact assuming that the object surface is Lambertian. If the object is near the constant-depth surface, the rendered image will appear blurry. When the object is farther away from the constant-depth surface, double images become more apparent when the sampling rate is inadequate.

From the point of view of image processing, sampling is a smoothing process of the plenoptic function because of the finite resolution of the capturing camera. Interpolation is another smoothing process. Under Lambertian premise, ignoring slight occlusion change, 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 procedure. From the causality requirement in scale-space theory ([7]), i.e., no "spurious detail" should be generated when smoothing, visible double images should be avoided and blurring is acceptable. Human perception is also more tolerant to blurring than to double images.

3.2 How do double images happen?

Double images degrade the visual quality of the rendered images. Now let's explain why this happens. As shown in Figure 3, C_1 and C_2 are two nearby cameras. C_iL intersects the constant-depth surface S at P_i $(i = 1, 2)$. Around P_i, there are two wedge-like intensity distributions of point L. The rays from C_1 do not contribute to the distribution around P_2 if L is not close enough to S, because the rays near C_1P_2 are zero. Similarly, the rays from C_2 do not contribute to the distribution around P_1 either. If the two patterns of the intensity distributions do not overlap, then some of the rendering rays can fall in the gap between the two patterns. As a result,

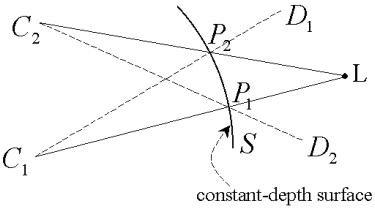

Figure 3: Double images of L appear on the rendered image: if C_1 and C_2 are too far away, or if L is not close to the constant-depth surface.

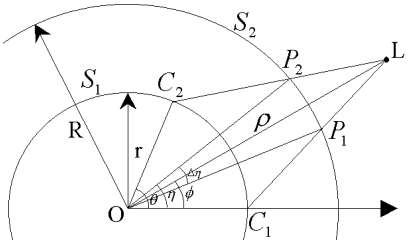

Figure 4: Geometry for concentric mosaics.

on the rendered image the double images of L will appear. Greater details will be shown in Section 4.3.

Therefore, the condition to avoid double images is that the intensity distributions of the point projected onto the constant-depth surface must at least meet each other. This resembles the Rayleigh criterion in the diffraction theory of optics ([3]), that two objects are said to be just resolved if the central spot of the diffraction pattern of one object falls on the first minimum of the diffraction pattern of the other object. Since we are treating discrete images, the criterion can be relaxed to the clearly resolved case.

Now we can apply this condition to both 3D plenoptic function (Concentric Mosaics) and 4D plenoptic function (light field) in the following sections.

4 3D plenoptic function

4.1 Review of Concentric Mosaics

Concentric Mosaics form a 3D plenoptic function by collecting all rays from a rotating camera on a plane[8]. At each rotation angle, an image with multiple verticle lines is captured. The nth concentric mosaic is created by putting together the nth verticle lines in all the images captured. Concentric Mosaics index all input image rays naturally in 3 parameters: radius, rotation angle and vertical elevation. It has been shown that any novel view inside the visible region can be rendered without any 3D reconstruction. As shown in Figure 4, a camera swings on the circle S_1. And a constant-depth circle (or cylindrical surface) S_2 is assumed for rendering. To render a novel ray (e.g., OP_2), first we find its intersection point (P_2) with the constant-depth surface. Then two nearest rays (C_1P_2 and C_2P_2) from nearby cameras (C_1 and C_2) are interpolated to generate the rendering result.

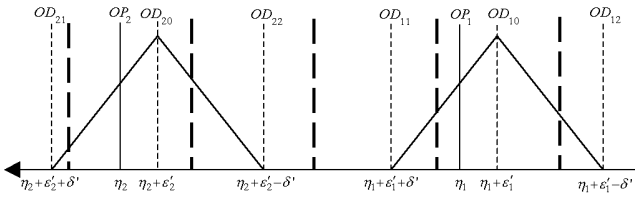

Figure 5: Double images happen when two wedges do not touch: the angle between successive positions of the camera is too wide or the point is too far from the constant-depth surface.

4.2 The minimum sampling condition

We now move from the cylindrical coordinate in Figure 4 to Figure 5 where the horizontal axis represents the angle of the ray starting from the cylinder center O, the two wedge-like intensity distributions of the point L might not overlap if C_1 and C_2 are not sufficiently close and L is not near S_2. As shown in Figure 5,

1. η_i is the position of OP_i,
2. $\eta_i + \varepsilon_i'$ is the position of OD_{i0}, where C_iD_{i0} is the nearest ray (see C_1D_{10} in Figure 2(a)) to C_iL viewed from C_i (to save notation, we assume that D_{ij} $(i = 1, 2;\ j = 0, 1, 2)$ are on S_2), and
3. $\eta_i + \varepsilon_i' \pm \delta'$ are positions of OD_{i1} and OD_{i2} (see C_1D_{11} and C_1D_{12} in Figure 2(a)) respectively,

where ε_i' is the angle between OP_i and OD_{i0}, and δ' is the angle between OD_{i0} and OD_{ij} $(j = 1, 2)$[1].

In this case, if $\eta_2 + \varepsilon_2' - \delta' > \eta_1 + \varepsilon_1' + \delta'$, when the viewer is at O and the viewing-rays are indicated by the thick dashed-lines in Figure 5, the rendered image of the point L will appear double.

Therefore, to avoid double images, that $\eta_2 + \varepsilon_2' - \delta' \leq \eta_1 + \varepsilon_1' + \delta'$ must be fulfilled, or equivalently

$$\eta_2 - \eta_1 \leq \varepsilon_1' - \varepsilon_2' + 2\delta'. \tag{1}$$

This is the condition when L is outside S_2. If L is inside S_2, it becomes

$$\eta_1 - \eta_2 \leq \varepsilon_2' - \varepsilon_1' + 2\delta'. \tag{2}$$

Because L is random, both (1) and (2) must be fulfilled. Furthermore, $\varepsilon_1' - \varepsilon_2'$ can take an arbitrary value in $[-\delta', \delta')$. Therefore, the final condition to avoid double images is

$$|\eta_2 - \eta_1| \leq \delta'. \tag{3}$$

[1]Note that since δ is extremely small and the FOV of the camera is fairly small, these four angles are nearly equal.

4.3 Lower bound analysis

Referring to Figure 4, suppose in polar coordinate $L = (\rho, \phi)$, $C_1 = (r, 0)$, $C_2 = (r, \theta)$, $P_2 = (R, \eta)$, then η satisfies:

$$\frac{R\sin\eta - r\sin\theta}{\rho\sin\phi - r\sin\theta} = \frac{R\cos\eta - r\cos\theta}{\rho\cos\phi - r\cos\theta},$$

since P_2 is on the line C_2L. The above equation can be written as:

$$\frac{1}{r}\sin(\eta - \phi) - \frac{1}{\rho}\sin(\eta - \theta) = \frac{1}{R}\sin(\theta - \phi). \tag{4}$$

Because the angle of the point is between two successive positions of the camera, the three angles $\eta - \phi$, $\eta - \theta$ and $\theta - \phi$ are all small, (4) can be linearized to:

$$\frac{1}{r}(\eta - \phi) - \frac{1}{\rho}(\eta - \theta) \approx \frac{1}{R}(\theta - \phi).$$

Hence

$$\eta \approx \frac{\rho(R - r)\phi - r(R - \rho)\theta}{R(\rho - r)}. \tag{5}$$

Therefore, for successive positions of the camera (with same R, r, ϕ and ρ),

$$|\Delta\eta| = |\eta_1 - \eta_2| \approx \left|\frac{\frac{\rho}{R} - 1}{\frac{\rho}{r} - 1}\Delta\theta\right|. \tag{6}$$

Next, we set out to find the relationship between δ' and δ. Referring to Figure 6, for the triangle $\triangle OCQ$, using the law of sines, we have

$$\frac{|CQ|}{\sin\varphi'} = \frac{R}{\sin\varphi} = \frac{r}{\sin(\varphi - \varphi')}, \tag{7}$$

Again, because φ is relatively small (typically $|\varphi| \leq \frac{\pi}{10}$, and $|\sin\frac{\pi}{10} - \frac{\pi}{10}| < 0.0052$) and $\varphi' < \varphi$, the above relation can be linearized as

$$\frac{|CQ|}{\varphi'} \approx \frac{R}{\varphi} \approx \frac{r}{\varphi - \varphi'},$$

hence $|CQ| \approx R - r$. Therefore, the relationship between δ' and δ is

$$\delta' = \frac{\widehat{PQ}}{R} \approx \frac{|CQ|\delta}{R} \approx \frac{R - r}{R}\delta = (1 - \frac{r}{R})\delta. \tag{8}$$

Combining (3), (6) and (8), we obtain:

$$|\Delta\theta| \leq \left|\frac{(\frac{\rho}{r} - 1)(\frac{r}{R} - 1)}{\frac{\rho}{R} - 1}\right|\delta.$$

If the depth of the point L is constrained between A and B ($r < A \leq \rho \leq B$ and $A \leq R \leq B$), then the maximum value of $\left|\frac{\frac{\rho}{R} - 1}{(\frac{\rho}{r} - 1)(\frac{r}{R} - 1)}\right|$ is

$$m = \frac{1}{1 - \frac{r}{R}} \cdot \max\left\{-\frac{\frac{A}{R} - 1}{\frac{A}{r} - 1}, \frac{\frac{B}{R} - 1}{\frac{B}{r} - 1}\right\} \tag{9}$$

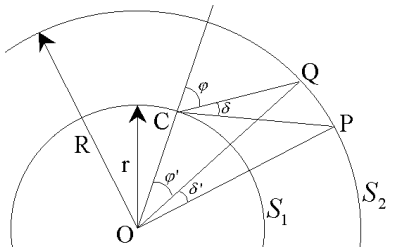

Figure 6: The relationship between δ and δ'.

Therefore a bound for the number of pictures is:

$$N_1 = \left\lceil \frac{2\pi m}{\delta} \right\rceil,$$

where $\lceil x \rceil$ denotes the smallest integer not less than x.

On the other hand, since the FOV of the camera is limited, the patches that every picture projected onto the cylinder must cover the cylinder, otherwise the bilinear interpolation could not be properly carried out. Accordingly, the number of pictures should also be larger than:

$$N_2 = \left\lceil \frac{2\pi}{\Phi} \right\rceil,$$

where $\Phi = \frac{1}{2}FOV - \arcsin\left(\frac{r}{R}\sin\left(\frac{1}{2}FOV\right)\right)$, using the second equality in (7).

Finally, the lower bound we attain is

$$N = \max\{N_1, N_2\}. \tag{10}$$

Note that we are not saying that if the number of pictures is greater than N then the visual quality will be acceptable. Rather, if the number of pictures is less than N then even the simplest scene with only a point could not be properly rendered. The theoretical lower bound should be higher than the one we have deduced above. However, when the scene becomes more complex, one might not notice too much artifact due to the characteristics of human vision. Consequently, the actual lower bound will not deviate too much from the one we estimated.

The above analysis only considers the horizontal resolution. Our study shows that adding the vertical resolution does not improve the minimum sampling rate.

4.4 Optimal constant-depth R

Instead of simply choosing constant-depth $R = \frac{A+B}{2}$, Eq.(9) can guide us to find a better constant-depth R, such that m is minimized, and the number of pictures required becomes smaller. Such an R satisfies:

$$-\frac{\frac{A}{R}-1}{\frac{A}{r}-1} = \frac{\frac{B}{R}-1}{\frac{B}{r}-1}$$

$$R = \frac{2AB-(A+B)r}{A+B-2r}. \tag{11}$$

One can easily check that $A \le R \le \frac{A+B}{2}$, and the equalities hold only for $A = B$. This choice of R is reasonable because closer objects will be distorted more and thus need 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, this isn't a difficult task to measure them. Moreover, since $B > A$, R is much less sensitive to B. Therefore, only the minimum depth needs to be accurate.

It is interesting to rewrite (11) as:

$$\frac{2}{R-r} = \frac{1}{A-r} + \frac{1}{B-r},$$

which means that the optimal constant depth is exactly the harmonic mean of the minimum and maximum depths. It is also worth noting that such choice of R can make the objects at the minimum depth and the miximum depth be rendered equally sharply.

4.5 Validity of the bound

1. If the scene is truly at a constant-depth, e.g., a painted cylinder, then $A = B = R$. In this case $N = N_2$, which is true.

2. If the scene is infinitely far away, let R chosen as (11), then

$$m = \frac{r(\frac{1}{A}-\frac{1}{B})}{2(1-\frac{r}{A})(1-\frac{r}{B})}. \tag{12}$$

When $A \to \infty$ and $B \to \infty$, $m \to 0$. Again $N = N_2$, which is also true.

3. If $r \to 0$, then the concentric mosaics will reduce to a panorama, and the number of pictures needed is N_2. In this case, $m \to 0$ and $N = N_2$. The lower bound is correct again.

4. The above examples are all extreme cases. Our analysis will proceed with real data. Figure 7 illustrates the top view of a real scene. The scene is enclosed by an ellipse and the center of the camera rig is placed near one of the focal points of the ellipse. The scales are labeled in the figure. The radius of the rig is 1.7m. The horizontal FOV of the camera is 43° and each picture taken by the camera is 360 pixels by 288 pixels. Then $A = 3.4$, $B = 16.6$, $r = 1.7$, and $\delta = \frac{43}{360} \cdot \frac{\pi}{180} \approx \frac{\pi}{1500}$. To achieve the best quality, R is chosen as (11), thus $R = 4.75$. Then $N = 1329$. Figure 10 is a panoramic view of part of the scene. Figure 8 compares the details between the rendered scenes with two different sampling rates. We can see that when the number of pictures is 1479, the scene is rather satisfactory, while clear double images appear in the scene reconstructed from 986 pictures. In this experiment, the lower bound we computed is fairly accurate.

Figure 10: Part of a panoramic view of a real scene.

5 4D plenoptic function

5.1 Lightfield rendering

In light field rendering proposed by Levoy and Hanrahan [5], a 4D light field is parameterized by light slab representation, where an oriented line is defined by connecting a point on the (u, v) plane to a point on the (s, t) plane. Both the uv and st planes are uniformly discretized. The light field is captured by placing a camera on the uv plane and making it face the st plane. Images are taken at each grid on the uv plane. To ensure that the captured light rays pass the grids on the st plane, sheared perspective projection and resampling are employed. To render a novel ray, first its intersection points with the uv and st planes are computed and then 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. The configuration of uv and st planes is flexible according to the scene to be rendered. Similar configuration was also proposed in Lumigraph [4]. Here we will only apply the criterion in Section 3.1 to the light field of "Lion" (Figure 14(d) in [5]) to analyze the maximum allowable distance between successive camera locations.

5.2 The maximum camera spacing

In the lion light field, four-slab arrangement is used for inward-looking views of a lion placed at the origin. Each slab is a copy of the one shown in Figure 9, rotated by every 90 degrees along the origin O. During rendering, the light slab is resampled. The resampling process is simply interpolating the 4D function from the nearest samples. Quadra-linear interpolation in both uv and st planes gives improved visual quality, and therefore is used in our analysis. Again, we suppose that the scene consists of only one point.

Note that quadra-linear interpolation implicitly assumes that the scene is of constant-depth: on the st plane. To derive the maximum horizontal displacement of the camera, we place the point on the base-plane, then the interpolation becomes bilinear. Without loss of generality, we set up a coordinate as in Figure 11, where $L = (x_0, y_0)$ is the point in the scene, C_1 and C_2 are two nearby places of the camera, 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. A slight difference is that we assumed a constant-depth plane that is parallel to but might not be the st plane. The distance between the two planes is R. We notice that, due to perspective projection, if the slant of C_1L is too small the original intensity distribution of the point L may stride over several intervals on st plane (Figure 12). The more the intervals, the blurrier the point becomes.

Let D_{11} be the first grid on the st plane that is outside and on the left of the "shadow" of L and D'_{11} is the point that C_1D_{11} intersects the constant-depth plane. D'_{22} is defined alike. Following the analysis in the previous section, to avoid double images, D'_{11} must be at the left of D'_{22}.

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

$$x_1 = -a + \frac{d(x_0 + a)}{y_0 + d}, \quad \text{and} \quad x_2 = a + \frac{d(x_0 - a)}{y_0 + d}$$

respectively. Suppose D_{i0} is by ε_i apart from L_i and D_{ii} is by $N_i\delta'$ away from D_{i0} $(i = 1, 2)$, where N_i is a positive integer and δ' is the sample spacing on st plane, then the x-coordinates of D'_{11} and D'_{22} are:

$$\begin{aligned} x'_{11} &= -a + \left(\tfrac{d(x_0+a)}{y_0+d} + \varepsilon_1 - N_1\delta'\right)\left(1 + \frac{R}{d}\right), \\ x'_{22} &= a + \left(\tfrac{d(x_0-a)}{y_0+d} + \varepsilon_2 + N_2\delta'\right)\left(1 + \frac{R}{d}\right). \end{aligned}$$

From $x'_{11} \le x'_{22}$, we have

$$-2a \cdot \frac{(y_0 - R)d}{(y_0 + d)(R + d)} \le (N_1 + N_2)\delta' + (\varepsilon_2 - \varepsilon_1) \quad (13)$$

Since the position of L is arbitrary, $\varepsilon_2 - \varepsilon_1$ can vary between $-\delta'$ and δ'. Therefore the following condition must be satisfied:

$$-2a \cdot \frac{(y_0 - R)d}{(y_0 + d)(R + d)} \le (N_1 + N_2 - 1)\delta'$$

The above condition is deduced when $y_0 < R$. If $y_0 > R$, the corresponding condition is

$$2a \cdot \frac{(y_0 - R)d}{(y_0 + d)(R + d)} \le (N_1 + N_2 - 1)\delta'$$

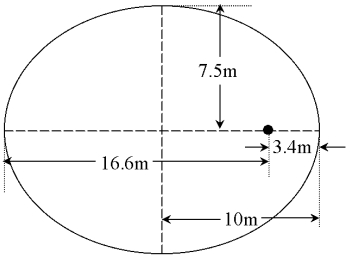

Figure 7: Top view of a real scene.

Figure 8: A close-up view of the scene reconstructed: from 986 pictures (top) and 1479 pictures (bottom).

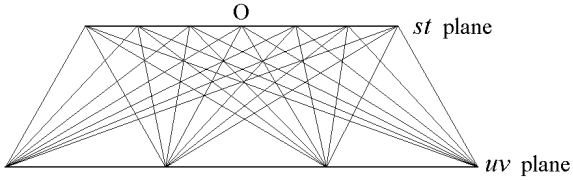

Figure 9: Top view of the light slab in the lion light field. The constant-depth is implicitly assumed to be on the st plane.

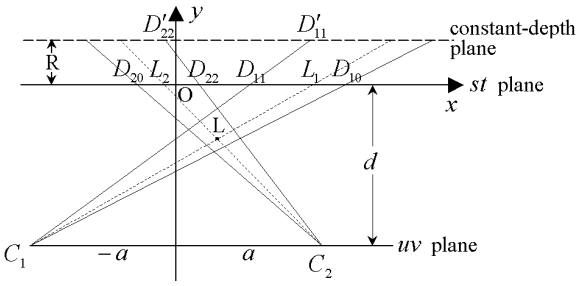

Figure 11: Rendering a single point L with light field.

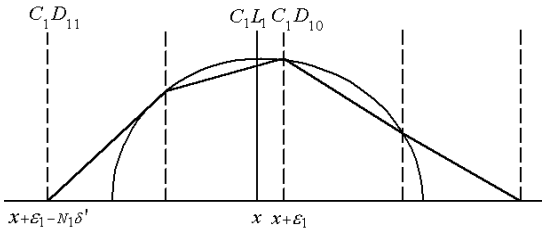

Figure 12: Due to perspective projection, the "shadow" of a point (curve-shaped) may be much wider than the sample spacing on st plane.

Summing up, the condition to avoid double images is

$$\begin{aligned} 2a &\leq (N_1 + N_2 - 1)\delta' \cdot \frac{(y_0 + d)(R + d)}{|y_0 - R|d} \\ &= (N_1 + N_2 - 1)\delta \cdot \frac{(y_0 + d)(R + d)}{|y_0 - R|} \end{aligned} \tag{14}$$

where δ is the horizontal resolution of the camera.

If y_0 is between $A(\leq R)$ and $B(\geq R)$, then taking the minimum value of the right hand side of (14) gives the maximum allowed distance between two places of the camera, namely

$$D_{max} = \delta(R + d) \cdot \min\left\{\frac{A + d}{R - A}, \frac{B + d}{B - R}\right\} \tag{15}$$

The coefficient $N_1 + N_2 - 1$ vanishes because it equals to 1 when L is near the origin.

The above formula is for horizontal spacing. Similarly, the formula for vertical spacing is

$$D^v_{max} = \delta_v(R + d) \cdot \min\left\{\frac{A + d}{R - A}, \frac{B + d}{B - R}\right\} \tag{16}$$

where δ_v is the vertical resolution of the camera.

By making $\frac{A+d}{R-A} = \frac{B+d}{B-R}$, we can find the optimal constant-depth R for light field rendering, namely

$$R = \frac{2AB + (A + B)d}{A + B + 2d}.$$

We see that since $|A| \ll d$ and $|B| \ll d$, it is important that $A + B \approx 0$ so that the st plane is a good constant-depth

plane. Therefore, the object must be placed carefully so that its center is at the origin, as done in [5].

Again, one can check that the optimal R given above is also the harmonic mean of the minimum and maximum depths, provided that the zero-depth plane is moved to the uv plane.

5.3 An Example

In the "Lion" light field rendering shown in [5], it was reported that $d = 0.5$m and picture size is 256 by 256. We guessed[2] $\widehat{FOV} = 26°$.

- For side view of the lion, $\hat{A} = -0.02$m, $\hat{B} = 0.02$m, then $\delta = \frac{26}{256} \cdot \frac{\pi}{180} \approx \frac{\pi}{1800}$ and $D_{max} \approx 2$cm.
- For front view of the lion, $\hat{A} = -0.05$m, $\hat{B} = 0.05$m, then $D_{max} \approx 8$mm.

Therefore, the sampling rates of the lion light field should be very different for front and side views. In [5], it is reported that the adopted spacing is 3.125cm, regardless of front and side views. The authors also revealed that double images appeared in their experiments, and the situation became the worst near the head and tail of the lion. It appears that the computed spacing is comparable to the spacing used in the capturing rig for the Lion data set.

6 Conclusion and future work

In this paper, we have presented an investigation of the lower bound for the number of samples required in light field/Lumigraph rendering. Our analysis serves as a guidance for how to capture light field and for how to effectively compress a large amount of light field data. Our analysis is based on the casuality requirement in scale-space theory, i.e., blurred images are acceptable but double images resulting from interpolation must be eliminated. By simplifying the scene to a point, we infer that the two patterns of intensity distribution of the point, generated by projecting the point onto the constant-depth surface from two nearby positions of the camera, must at least meet each other. We have concluded that a minimum number of images have to be captured to avoid artifacts from light field rendering, and the lower bound is closely related to the camera resolution and the scene depth complexity. In addition, an optimal constant-depth can be found for the best rendering quality.

We have applied the lower bound analysis to 3D Concentric Mosaics and 4D light field. For Concentric Mosaics, only the horizontal resolution is analyzed to determine the sampling rate. The lower bound from our analysis agrees fairly well with the real data in our experiments. For light field, maximum camera spacing in both horizontal and vertical directions is studied. The lower bound for light field is also comparable with that listed in [5] considering the visual quality reported in the paper.

While the experimental results are encouraging, the lower bound would be more accurate if we also incorporate frequency distribution of the scene (e.g., textured regions) and the characteristics of human vision. Our analysis is carried out completely from the geometric aspect. Variation in the scene texture will significantly affect the minimum sampling rate. For example, few data points should be required if the scene is uniformly textured. Characteristics of human vision are also important for the sampling analysis.

We are currently working on how to reduce the large amount of light field data based on the sampling analysis. From this analysis, we see that without correct depth information during interpolation, incorrect rays will be selected, unless very dense samples are captured. A simple way to reduce data size is to discard some of the samples in slowly-varying areas in the scene or ray space. Another way is to estimate the approximate scene depth. Estimating depth from large amount of light field data could be easier than that in traditional computer vision. We are also working on how accurately the depth should be estimated and which samples need to be kept with the estimated approximate depth information.

[2] Unfortunately, we do not know the exact values of FOV, A and B in the Lion data set.

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] S. E. Chen. QuickTime VR – an image-based approach to virtual environment navigation. *Computer Graphics (SIGGRAPH'95)*, pages 29–38, August 1995.

[3] W. Driscoll, editor. *Handbook of Optics*. McGraw-Hill, New York, 1978.

[4] 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.

[5] 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.

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

[7] B. M. t. H. Romeny, editor. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, Netherlands, 1994.

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

[9] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and texture-mapped models. *Computer Graphics (SIGGRAPH'97)*, pages 251–258, August 1997.

[10] M. Wu. *Real-Time Stereo Rendering of Concentric Mosaics with Linear Interpolation*. To appear in VCIP'2000.