ON GENERALIZED SAMPLING FOR IMAGE-BASED RENDERING DATA

Cha Zhang and Tsuhan Chen

Dept. of Electrical and Computer Engineering, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
{czhang, tsuhan}@andrew.cmu.edu

ABSTRACT

In this paper, we apply generalized sampling to image-based rendering (IBR) data, more specifically, the light-field. We show that in theory the lowest sampling rate of lightfield when we use generalized sampling can be as low as half of that when we use rectangular sampling. However, in practice rectangular sampling has several advantages over generalized sampling. We analyze the pros and cons for each sampling approach, and explain why in practice rectangular sampling is still more preferable.

1. INTRODUCTION

Image-based rendering (IBR) has attracted a lot of attention recently. Proposed by Adelson and Bergen [1] as a 7D plenoptic function, IBR models a 3D dynamic scene by recording the light rays at every space location, towards every possible direction, over any range of wavelengths and at any time. By removing the time and the wavelength dimensions, McMillan and Bishop [2] defined plenoptic modeling as generating a continuous 5D plenoptic function from a set of discrete samples. The concepts of Lightfield by Levoy and Hanrahan [3] and Lumigraph by Gortler et al. [4] effectively parameterized the plenoptic function into a 4D function under the constraint that the object lies in free space.

The principle of the lightfield [3] can be briefly explained as follows. Light rays from an object can be parameterized by their intersections with two planes, one plane indexed with coordinate (u,v) and the other with coordinate (s,t). In Figure 1, we show an example where the two planes, the camera plane and the focal plane, are parallel. An example light ray is shown and indexed as (u_0,v_0,s_0,t_0). The coordinates are then discretized so that a finite number of light rays are recorded. In practice, we put on the camera plane a 2D array of cameras, which share the same focal plane. The images captured form a 2D array of images. To create a new view of the object, we just split the view into its light rays, which are then calculated by interpolating existing nearby light rays in the image array. The new view is then generated by reassembling the split rays together.

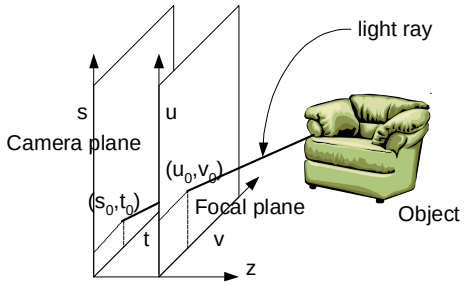

Figure 1 Lightfield parameterization.

Since the number of images taken in lightfield is often huge, it is very important to know the minimum sampling requirement for a specific scene. In this paper, we apply generalized sampling (GS) [7] to lightfield. We adopt the spectrum analysis results by Chai *et al.* [5] and show that with generalized sampling the sampling efficiency can be improved by a factor of 2 compared with rectangular sampling (RS). However, when we use GS in practice, we find many limitations. In this paper, we will analyze the two approaches and show that RS is preferable for practical IBR applications.

The paper is organized as follows. Section 2 gives a brief review on some previous work. We apply GS to lightfield and discuss the reconstruction problem in Section 3. Some experimental results are given in Section 4. Discussions and conclusions are presented in Section 5.

2. REVIEW OF THE PREVIOUS WORK

Recently Chai *et al.* [5] proposed a method to study the frequency spectrum of lightfield. Let the lightfield be $l(u,v,s,t)$ and its Fourier transform be $L(\Omega_u,\Omega_v,\Omega_s,\Omega_t)$. Assuming Lambertian surfaces and no occlusion, they showed that the spectral support of a lightfield signal is bounded by the minimum and maximum depths of objects in the scene, no matter how complicated the scene is. Figure 2 (a) shows the spectrum of a 2D lightfield predicted by their theory, where f is the focal length of the cameras, $z_{\min}$ and $z_{\max}$ are the minimum and maximum

depths of the scene. Without loss of generality we assume that along Ω_v the spectrum is bounded due to finite resolution of the camera (capturing camera or rendering camera) or the texture on the objects, whichever is lower, and we let the resolution be unity.

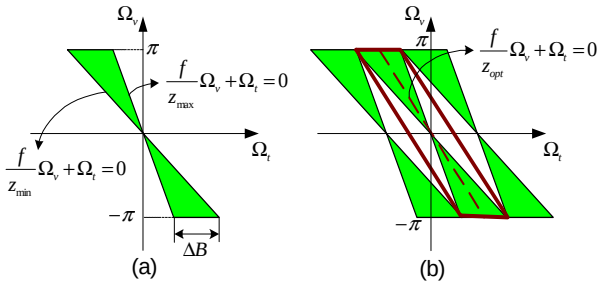

Figure 2 Sampling analysis of lightfield in [5]. (a) Frequency support of a scene with no occlusion and with Lambertian surfaces. (b) The "Optimal" compacting and the reconstruction filter in [5].

The analysis in [5] provides a good first-order approximation of the spectrum analysis of IBR. It further claimed that by knowing the bound of the support, the maximum sampling density could be achieved by compacting copies of the spectral support rectangularly in the frequency domain as Figure 2 (b). The optimal reconstruction filter is also shown in Figure 2 (b) with bold contours. It happens to be well approximated by the depth-driven bilinear interpolation filter already used in the Lumigraph paper [4]. The optimal rendering depth z_{opt} can also be obtained from the above figure, which is

$$z_{opt} = 2/[1/z_{\min} + 1/z_{\max}] \qquad (1)$$

In [6], Chan and Shum proposed to view the plenoptic sampling problem as a multidimensional sampling problem. Their analysis is conceptually important because it implies that we can apply generalized sampling to lightfield. They mentioned that "it is also possible to determine the minimum sampling densities for the quincunx and hexagonal sampling lattices" without further detail. We will explore this in more details in the following sections.

3. GENERALIZED SAMPLING FOR LIGHTFIELD

Figure 3 (a) shows how we can compact the frequency support and the replicas better with generalized sampling theory. With this sampling strategy, the sampling efficiency can be improved by a factor of 2 compared with Figure 2 (b), which means we only need 50% of the samples. The reconstruction filter is marked in bold contour in Figure 3 (a), which is a tilted fan-like filter. For detailed information about the generalized sampling theory, we refer the reader to [7].

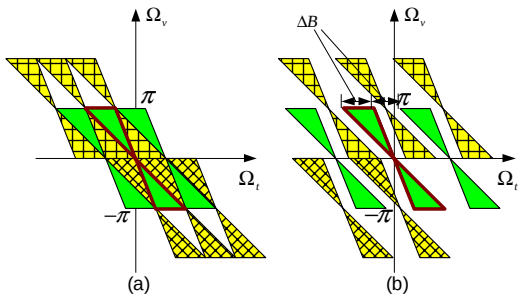

Figure 3 Generalized sampling of lightfield. (a) The most compact way to pack the lightfield spectrum. (b) Reduce the sampling rate such that GS has the same efficiency as RS in Figure 2 (b).

There are several problems to be concerned for the GS approach. First, the corresponding sampling lattice in the spatial domain for Figure 3 (a) may not be consistent with how we take images for the scene. We propose to sample the scene with RS at a higher sampling rate and then down-sample it to the required lattice. Interpolation may be required during the down-sampling process. Of course, interpolation will introduce extra errors in practice. This is what GS has to pay to achieve a lower sampling rate.

Second, during the rendering, there can be two approaches to reconstruct the continuous lightfield signal from the GS sampled data. One is to introduce a preprocessing stage before rendering. At this stage we up-sample the data to a rectangular grid by a discrete fan-like reconstruction filter. During the rendering we can simply apply depth-driven bilinear interpolation as before. This approach requires huge amount of memory to store the up-sampled data, but has a fast rendering speed. Another approach is to use a continuous fan-like reconstruction filter directly for rendering. Since it is difficult to find the best continuous filter given its finite support (which is required in rendering due to speed consideration), we design a discrete optimal reconstruction filter first and interpolate it to get the continuous filter. To speed up the rendering, look-up tables can be used to store the filter values. We adopt the second approach in the follow sections because it does not need much memory to store the up-sampled data.

Third, in theory we may be able to design an ideal fan-like reconstruction filter to get back the original lightfield signal without losing any information; in practice, however, we cannot design a filter without any transition band. We choose to reduce the sampling density for the ease of filter design. To give a fair comparison between GS and RS, we let them have the same sampling density, as is shown in Figure 2 (b) and Figure 3 (b). We will focus on the optimal discrete reconstruction filter design for these two cases.

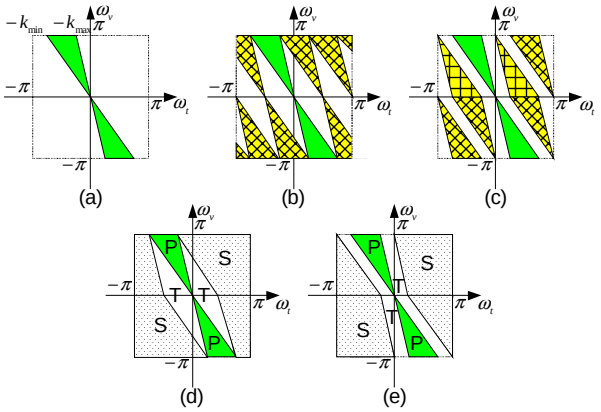

Figure 4 (a) The Fourier transform of the original discrete signal. (b) The Fourier transform of the signal with minimum sampling rate for RS (c) The Fourier transform of the signal with the same sampling rate for GS. (d) Up-sampling filter specification for RS. (e) Up-sampling filter specification for GS.

Assume we start with a RS sampled data at a sampling rate that is much higher than the minimum requirement. Its discrete Fourier transform is shown in Figure 4 (a), where the two slopes $-k_{\min}$ and $-k_{\max}$ are determined by the minimum and maximum depths of the scene. If we down-sample the data with RS to its minimum sampling rate, we get a new spectrum as in Figure 4 (b). If we down-sample it to the same rate with GS, we have Figure 4 (c)[1]. Figure 4 (d) and (e) are up-sampling filter specifications for (b) and (c), respectively. S, T and P stand for stop-band, transition-band and pass-band, respectively. We can observe that the transition-band of the up-sampling filter for RS is very narrow for high frequency components, but very wide for low frequency components. On the other hand, the transition-band of the filter for GS has constant width along all the frequencies. This observation implies that RS is better for scenes that has less high-frequency components, while GS is better otherwise.

We use the eigenfilter approach [8][9] to design the filters in Figure 4 (d) and (e). Let $D(\omega_v,\omega_t)$ be the desired frequency response in the pass-band, $H(\omega_v,\omega_t)$ be the filter we try to design. The Eigenfilter approach finds:

$$\arg\min_{H(\omega_v,\omega_t)} E = \arg\min_{H(\omega_v,\omega_t)} \{\alpha E_P + \beta E_S\}$$

$$= \arg\min_{H(\omega_v,\omega_t)} \left\{ \alpha \iint_P |D(\omega_v,\omega_t) - H(\omega_v,\omega_t)|^2 \, d\omega_v d\omega_t + \beta \iint_S |H(\omega_v,\omega_t)|^2 \, d\omega_v d\omega_t \right\} \quad (2)$$

where E is the overall square error measured by the weighted sum of the pass-band error E_P and the stop-band error E_S; α and β are weighting constants which control the accuracies of the approximation. In a normal setup, we often have $D(\omega_v,\omega_t)=1$.

Eigenfilter is optimal for filter design itself, but it is not necessarily optimal in terms of the reconstruction error for a certain signal. Let the down-sampled signal spectrum (such as Figure 4 (b) and (c)) be $X(\omega_v,\omega_t)$. We try to find the optimal reconstruction filter through:

$$\arg\min_{H(\omega_v,\omega_t)} E^r = \arg\min_{H(\omega_v,\omega_t)} \{E_P^r + E_S^r\}$$

$$= \arg\min_{H(\omega_v,\omega_t)} \left\{ \iint_P |X(\omega_v,\omega_t)|^2 |D(\omega_v,\omega_t) - H(\omega_v,\omega_t)|^2 \, d\omega_v d\omega_t + \iint_S |X(\omega_v,\omega_t)|^2 |H(\omega_v,\omega_t)|^2 \, d\omega_v d\omega_t \right\} \quad (3)$$

where E^r is the overall reconstruction error, E_P^r and E_S^r are the reconstruction errors in pass-band and stop-band, respectively. Equation (3) can still be solved through the eigenfilter approach, as it differs from Equation (2) only by a weighting function. Notice that the optimal filter is related with the signal $X(\omega_v,\omega_t)$. In our experiments in Section 4, we show that by using the Fourier transform of a first order auto-regressive (AR-1) signal to model $X(\omega_v,\omega_t)$, we get better reconstruction than the regular eigenfilter approach.

4. EXPERIMENTAL RESULTS

We show some experimental results on the two different sampling approaches. In order to have full control on the scenes and the cameras, we choose two scenes rendered from 3D models with texture. These scenes are shown in Figure 5, where scene (a) and (b) are named *Duck* and *Containers*, respectively. We take the center horizontal lines to construct the epipolar images (EPIs). Figure 5 (a1) and (b1) are snapshots of the scenes; (a2) and (b2) are their EPIs; (a3) and (b3) are the Fourier transform of the EPIs. Although occlusions can be observed in the scenes, we ignore them in our analysis since they are not significant in these two examples.

From the Fourier transform of the scenes in Figure 5, we can find the corresponding sampling rate and compact the spectrum to Figure 2 (b) and Figure 3 (b). One thousand random images are then rendered for each scene with different reconstruction filters. These images are also synthesized through a 3D model rendering engine. The difference between the synthesized images and the

[1] Strictly speaking, Figure 4 (c) is a transformed version of the original discrete Fourier transform. We perform this transform so that we can compare the filter specifications of the two sampling strategies. Please refer to [7] for details about such transforms.

rendered images is used to measure the quality of the sampling process. In our experiments, PSNR is used to measure such differences.

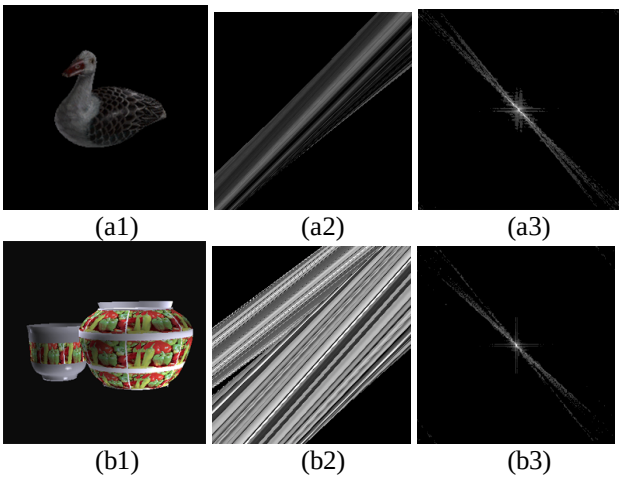

Figure 5 The testing scene EPIs and their Fourier transforms. (a) The scene Duck. (b) The scene Containers.

Table I Rendered images qualities for different sampling methods and reconstruction filters (PSNR: dB)

	Duck	Containers
RS, regular eigenfilter	33.99	20.44
RS, optimal reconstruction filter	35.89	21.81
GS, regular eigenfilter	35.68	22.04
GS, optimal reconstruction filter	36.06	22.21
RS, bilinear interpolation	36.10	21.67

The experimental results are shown in Table I. We test the regular eigenfilter (REF) and the optimal reconstruction filter (ORF) presented in Section 3 for both RS and GS. To design the ORF, we model each replica of the Fourier spectrum as the Fourier transform of an AR-1 signal with $\rho = 0.9$ along ω_v, and constant along ω_t. For example, the replica centered at $\omega_v = \omega_t = 0$ is represented as:

$$\left| X(\omega_v, \omega_t) \right|_P = \left| \frac{1}{1 - \rho e^{j\omega_v}} \right| \qquad (4)$$

where the subscript P simply means that it is valid only in the pass-band. We also list depth-driven bilinear interpolation for RS in Table I for comparison. Several conclusions can be drawn from the above table. RS REF performs much worse than GS REF. This is because the specification for RS REF design has a zero transition-band at high frequency components, which is hard to design. In all cases, ORF is significantly better than designing a general-purpose filter. This is because extra knowledge was employed during the ORF design process. With ORF, the difference between RS and GS becomes very small, which again shows the power of ORF. Interestingly, the simple approach that uses RS with bilinear interpolation gives comparable performance as GS with optimal filter for reconstruction. This is unexpected but was verified in some other scenes we tested. Comparing the two test scenes, there is more improvement by using GS for the scene *Containers*, because *Containers* has more high frequency components in its spectrum. This is consistent with our analysis in Section 3.

5. DISCUSSION AND CONCLUSIONS

Real world scenes often has weak high frequency components and strong low frequency components due to reasons such as occlusions, thus RS is usually more suitable than GS. Even for the *Containers* scene, the improvement by using GS is minor, which cannot justify the increased complexity in GS. Since GS is inconsistent with how the images are taken, the required re-sampling may introduce extra error. The rendering speed is another concern. Unless we build a huge look-up table for the reconstruction filter and always do simple rounding when searching for a filter value, bilinear interpolation is required to get the filter values at arbitrary viewpoints. Even if the designed reconstruction filter has the same size of support as bilinear interpolation, the filter interpolation will slow down the rendering by a factor of 4. Therefore, we conclude that in practice rectangular sampling is preferable to non-rectangular sampling for lightfield.

REFERENCES

[1] E. H. Adelson, and J. R. Bergen, "The plenoptic function and the elements of early vision", *Computational Models of Visual Processing*, Chapter 1, Edited by Michael Landy and J. Anthony Movshon. The MIT Press, Cambridge, Mass. 1991.

[2] L. McMillan and G. Bishop, "Plenoptic modeling: an image-based rendering system", *Computer Graphics (SIGGRAPH'95)*, pp. 39-46, Aug. 1995.

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

[4] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The Lumigraph", *Computer Graphics (SIGGRAPH'96)*, pp. 43-54, Aug. 1996.

[5] J.X. Chai, X. Tong, S.C. Chan and H. Y. Shum, "Plenoptic sampling", *Computer Graphics (SIGGRAPH'00)*, pp.307-318, July 2000.

[6] S. C. Chan and H. Y. Shum, "A Spectral Analysis for Light Field Rendering", *ICIP 2000*.

[7] D.E. Dudgeon and R.M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-hall signal processing series, 1984.

[8] P. P. Vaidyanathan and T. Q. Nguyen, "Eigenfilters: A New Approach to Least-Squares FIR Filter Design and Applications Including Nyquist Filters", *IEEE Trans. on CAS*, pp. 11-23, Jan. 1987.

[9] T. Chen, "Unified Eigenfilter Approach: with Applications to Spectral/Spatial Filtering", *IEEE ISCAS*, Chicago, May 1993.