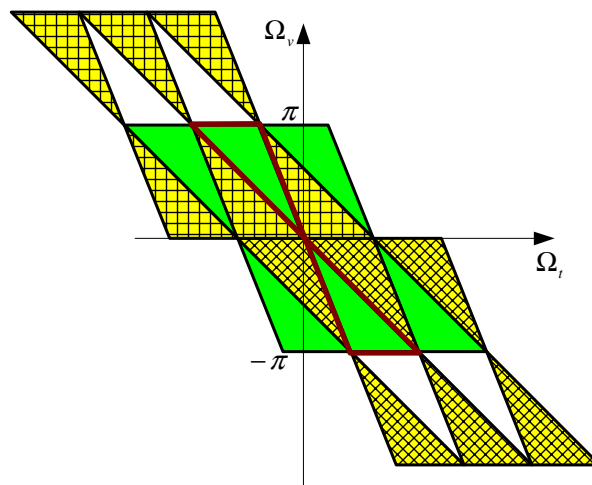Electrical *&* Computer
ENGINEERING

Generalized Plenoptic Sampling

Cha Zhang and Tsuhan Chen
Advanced Multimedia Processing Lab

Technical Report AMP 01-06
September 2001

Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Image-based rendering (IBR) has become a very active research area in recent years. The optimal sampling problem for IBR has not been completely solved. In this paper, we show that theoretically, the optimal sampling efficiency can be achieved by employing the generalized periodic sampling theory with arbitrary geometry. When there is no occlusion in the scene and the scene is Lambertian, we show that the sampling density can be twice of that when we use rectangular sampling. We then propose a general framework for IBR sampling. Begin with an over-sampled dataset, we down-sample the data first in the discrete domain. To render an image from the down-sampled data, two approaches are proposed, i.e., to reconstruct the over-sampled dataset first through up-sampling and then rendering, or to use a continuous interpolation filter to calculate the desired light rays directly. Eigenfilter method is employed to design filters during down-sampling and up-sampling. We show that if the proposed approach adopts the same down-sampling density as the previous work, the reconstruction filter of the proposed approach is easier to design, and the reconstructed scene could have a higher quality.

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.

Keywords: plenoptic sampling, image-based rendering, periodic sampling with arbitrary geometry, eigenfilter.

I. 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 wave-lengths 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 parameterize the plenoptic function into a 4D function under the constraint that the object or the viewer could be constrained within a 3D bounding box. Shum and He [5] proposed concentric mosaics, where the number of dimensions of the plenoptic function is decreased further to three by restricting the viewer's movement inside a circle.

The principle of the Lightfield can be briefly addressed as follows. By placing the object in its bounding box which is surrounded by another larger box, the Lightfield indexes all possible light rays entering and exiting one of the six parallel planes of the double bounding boxes. The Lightfield data is thus composed of six 4D functions, where the plane of the inner box is indexed with coordinate (u, v) and that of the outer box with coordinate (s, t). In Figure 1, we show one pair of the parallel planes and one example light ray which can be 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. Usually, it is discretized more precisely for the inner bounding box closer to the object, and more coarsely for the outer bounding box. Alternatively, the Lightfield can be considered as six two-dimensional image arrays, with all the light rays coming from a fixed (s, t) coordinate forming one image. This is equivalent to setting a camera at each coordinate (s, t) and taking a picture of the object with the imaging plane being the (u, v) plane. An example of the Lightfield image array is shown in Figure 2. 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 arrays. The new view is then generated by reassembling the split rays together.

The advantage of IBR is that it requires little or no geometrical information about the scene to realistically render it, if the number of sample images are dense enough. This is very attractive due to the difficulty to construct geometrical models for objects in a real-world scene. Also, the rendering speed of IBR is independent to the scene complexity. However, the amount of data to be stored for IBR is huge. Recently Chai et al. [6] proposed to study the plenoptic sampling problem in order to reduce the amount of data for IBR. A key observation of that paper is that the spectral support of a light field signal is bounded by the minimum and maximum depths of objects in the scene only, no matter how complicated the spectral support may be. They claim that by knowing the bound of the support, the maximum sampling density can be achieved by compacting copies of the spectral support in the frequency domain. The above conclusions are drawn under the assumption that the Lightfield has no occlusion and the scene is Lambertian.

Chan and Shum went one step further and proposed to view IBR sampling problem as a multi-dimensional signal processing problem [7]. In this paper, we examine the approach to replacing rectangular sampling with generalized sampling strategy in the high dimensional space [8]. The analysis is performed in both the continuous spatial domain and the discrete spatial domain. We propose a general framework for IBR sampling. Starting from an over-sampled dataset, we down-sample the data first in the discrete domain. To render an image from the down-sampled data, two approaches are proposed, i.e., to reconstruct the over-sampled dataset first through up-sampling followed by rendering, or to use a continuous interpolation filter to calculate the desired light rays directly. Throughout the analysis, the eigenfilter method is used to design optimal multi-dimensional filters we need during down-sampling and up-sampling process. To compare our proposed sampling strategy with the conventional one, we show that if the proposed approach has the same down-sampling density as the previous work, the reconstruction filter of the proposed approach is easier to design, and the reconstructed scene has a higher quality. Currently, we focus our analysis on 2D Lightfield.

In this case, the pair of planes in Figure 1 degenerates into two lines, e.g., the *v* axis and the *t* axis, which are the two dimensions of the Lightfield. The full analysis of 4D Lightfield and 3D concentric mosaics is under study.

This paper is organized as follows. In Section II, we introduce the notations used in this report and present some background knowledge for multi-dimensional signal processing. In Section III, we give a complete analysis of IBR sampling in the continuous spatial domain. Section IV gives a detailed analysis in the discrete spatial domain, including a general framework on a practical IBR sampling process, discrete domain down-sampling and discrete domain to continuous domain interpolation. Experimental results are shown in Section Conclusions and future work are given in Section VI.

II. Notations and Background Knowledge

A. Notations

The sampling problem of 2D Lightfield involves many notations for both discrete-time and continuous-time signals. In this section we give a clarification of all the notations we use.

We use lowercase boldfaced letters to denote vectors (mostly column vectors, unless specified otherwise), and capital boldfaced letters to denote matrices. The symbol $\mathbf{I}_k$ denotes the $k \times k$ identity matrix. The notations $\mathbf{A}^T$, $\mathbf{A}^{-1}$, $\mathbf{A}^{-T}$ and $\mathbf{A}*$ denote the transpose, the inverse, the inverse transpose and the conjugate of the matrix $\mathbf{A}$. The continuous signal is typically written as $s(\mathbf{x})$, where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_D \end{bmatrix}^T$ is the spatial domain index (continuous) of the signal, with D denoting the number of dimensions. Since we are dealing with 2D Lightfield, $\mathbf{x}$ is often a 2D vector. For historical reason, when we use $s(\mathbf{x})$ to represent the Lightfield signal, we denote $\mathbf{x} = \begin{bmatrix} x_t & x_v \end{bmatrix}^T$, where *t* and *v* stand for the camera line and the focal line, respectively. The continuous Fourier transform of signal $s(\mathbf{x})$ is denoted as $S(\mathbf{\Omega})$, which is defined as:

$$S(\mathbf{\Omega}) = \int_{-\infty}^{\infty} s(\mathbf{x}) e^{-j\mathbf{\Omega}^T \mathbf{x}} d\mathbf{x} \tag{1}$$

For the same reason, when we are describing 2D Lightfield signals, we let $\mathbf{\Omega} = \begin{bmatrix} \Omega_t & \Omega_v \end{bmatrix}^T$. The inverse Fourier transform of a multidimensional continuous signal is given by:

$$s(\mathbf{x}) = \frac{1}{(2\pi)^D} \int_{-\infty}^{\infty} S(\mathbf{\Omega}) e^{j\mathbf{\Omega}^T \mathbf{x}} d\mathbf{\Omega} \tag{2}$$

When the spatial domain index is discrete, which we call a discrete signal, we represent the signal as $s_d(\mathbf{n})$ (with subscript *d* often omitted if the context is clear). $\mathbf{n} = \begin{bmatrix} n_1 & n_2 & \cdots & n_D \end{bmatrix}^T$ is the spatial domain index (discrete) of the signal. Note that all n_i's are integers. Let symbol $\mathbb{N}$ denote the set of all $D \times 1$ integer

vectors. Therefore, we have $\mathbf{n} \in \mathsf{N}$. The real vector $\boldsymbol{\omega} = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_D \end{bmatrix}^T$ is the frequency domain variable of the multidimensional discrete signal. The Fourier transform of signal $s_d(\mathbf{n})$ is defined as:

$$S_d(\boldsymbol{\omega}) = \sum_{\mathbf{n} \in \mathsf{N}} s_d(\mathbf{n}) e^{-j\boldsymbol{\omega}^T \mathbf{n}} \tag{3}$$

if the summation converges. Subscript d is often omitted if no confusion exists. The inverse Fourier transform is given by:

$$s_d(\mathbf{n}) = \frac{1}{(2\pi)^D} \int_{\boldsymbol{\omega} \in [-\pi, \pi)^D} S_d(\boldsymbol{\omega}) e^{j\boldsymbol{\omega}^T \mathbf{n}} d\boldsymbol{\omega} \tag{4}$$

where $[a,b)^D$ denotes the set of $D \times 1$ real vectors $\mathbf{x}$ with components x_i in the range $a \le x_i < b$. When presenting the 2D Lightfield signal, we denote $\mathbf{n} = \begin{bmatrix} n_t & n_v \end{bmatrix}^T$ and $\boldsymbol{\omega} = \begin{bmatrix} \omega_t & \omega_v \end{bmatrix}^T$.

B. Fundamentals of Multidimensional Signal Processing

B.1. The sampling theory in the multidimensional case

In the multidimensional case, it is well known that the rectangular sampling strategy is not necessary the optimal [8]. For example, given a circularly band-limited signal in 2D case, the optimal sampling strategy is to sample it hexagonally, which will increase the sampling efficiency by around 13.4%. In the multidimensional space, periodic samples can be taken with arbitrary sampling geometry, as shown in Figure 3. Take the 2D case as an example. Let $s(\mathbf{x})$ be the spatial domain continuous signal, and $S(\boldsymbol{\Omega})$ be its Fourier transform. Define the sampling matrix $\mathbf{V}$ as:

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \tag{5}$$

where $\mathbf{v}_1$ and $\mathbf{v}_2$ are two linearly independent vectors. Sampling the continuous signal $s(\mathbf{x})$ produces the discrete signal:

$$s_d(\mathbf{n}) = s(\mathbf{V}\mathbf{n}) \tag{6}$$

where $s_d(\mathbf{n})$ stands for the sampled signal, $\mathbf{n} = \begin{bmatrix} n_1 & n_2 \end{bmatrix}^T \in \mathsf{N}$ is an integer vector representing the index of the samples along $\mathbf{v}_1$ and $\mathbf{v}_2$. Notice that $1/|\det \mathbf{V}|$ represents the density of samples per unit area, which should be as small as possible to save the number of samples as long as there is no aliasing. Denote $S_d(\boldsymbol{\omega})$ as the Fourier transform of $s_d(\mathbf{n})$. There is a close relationship between the Fourier transform of $S_d(\boldsymbol{\omega})$ and $S(\boldsymbol{\Omega})$ [8]:

$$S_d(\boldsymbol{\omega}) = \frac{1}{|\det \mathbf{V}|} \sum_{\mathbf{k}} S\left(\mathbf{V}^{-T}(\boldsymbol{\omega} - 2\pi\mathbf{k})\right)$$

$$\text{or, } S_d(\mathbf{V}^T \boldsymbol{\Omega}) = \frac{1}{|\det \mathbf{V}|} \sum_{\mathbf{k}} S(\boldsymbol{\Omega} - \mathbf{U}\mathbf{k}) \tag{7}$$

where $\mathbf{k} = \begin{bmatrix} k_1 & k_2 \end{bmatrix}^T \in \mathsf{N}$; $\mathbf{U}$ is a matrix that satisfies:

$$\mathbf{U}^T \mathbf{V} = 2\pi\mathbf{I} \tag{8}$$

and $\mathbf{I}$ is the 2x2 identity matrix. From Equation (7), we can see that periodic sampling on a parallelepiped lattice in the spatial domain leads to the replication of the Fourier transform on another parallelepiped lattice in the frequency domain. This is shown in Figure 3, where the two lattices satisfy Equation (8).

B.2. Down-sampling for multidimensional discrete signal

The $\mathbf{M}$-fold down-sampled version of $s(\mathbf{n})$ is defined as $r(\mathbf{n}) = s(\mathbf{Mn})$, where $\mathbf{M}$ is a nonsingular integer matrix called the down-sampling matrix. In the frequency domain, the relation is [8]:

$$R(\mathbf{\omega}) = \frac{1}{|\det \mathbf{M}|} \sum_{\mathbf{k} \in \mathsf{N}(\mathbf{M}^T)} S(\mathbf{M}^T(\mathbf{\omega} - 2\pi\mathbf{k})) \tag{9}$$

where $\mathsf{N}(\mathbf{M}^T)$ is the set of all integer vectors of the form $\mathbf{M}^T\mathbf{x}$, $\mathbf{x} \in [0,1)^D$. Similar to the continuous case, in the frequency domain, down-sampling the signal with $\mathbf{M}$ is equivalent to duplicating the original spectrum on the grid $\mathbf{Nk}$ with $\mathbf{N} = 2\pi\mathbf{M}^{-T}$ and $\mathbf{k} \in \mathsf{N}(\mathbf{M}^{-T})$, followed by transforming the whole spectrum with matrix $\mathbf{M}^T$.

B.3. Up-sampling for multidimensional discrete signal

For a nonsingular integer matrix $\mathbf{L}$, the $\mathbf{L}$-fold up-sampling of $s(\mathbf{n})$ is defined as [8]:

$$r(\mathbf{n}) = \begin{cases} s(\mathbf{L}^{-1}\mathbf{n}) & \mathbf{n} \in LAT(\mathbf{L}) \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $LAT(\mathbf{L})$ (the lattice generated by $\mathbf{L}$) denotes the set of all vectors of the form $\mathbf{Lm}$, $\mathbf{m} \in \mathsf{N}$. Clearly, the condition $\mathbf{n} \in LAT(\mathbf{L})$ above is equivalent to $\mathbf{L}^{-1}\mathbf{n} \in \mathsf{N}$. The corresponding frequency-domain relation of up-sampling is:

$$R(\mathbf{\omega}) = S(\mathbf{L}^T\mathbf{\omega}) \tag{11}$$

B.4. Down-sampling filters and up-sampling filters

Similar to the 1D case, a multidimensional signal has to be pre-filtered before entering the down-sampler, and post-filtered after the up-sampler. Such systems are shown in Figure 4 (a) and (b). Given the down-sampling/up-sampling matrix $\mathbf{M}$ (or $\mathbf{L}$), these filters typically have parallelepiped-shaped passband support in the region:

$$\mathbf{\omega} = \pi\mathbf{M}^{-T}\mathbf{x} + 2\pi\mathbf{k}, \qquad \mathbf{x} \in [-1,1)^D, \mathbf{k} \in \mathsf{N} \tag{12}$$

where $2\pi\mathbf{k}$ is only to indicate the periodicity of the spectrum. In this paper, the down-sampling filters and up-sampling filters play very important roles in the optimal sampling of 2D Lightfiled.

B.5. Down-sampling and up-sampling with a rational matrix

In general, we can process any down-sampling or up-sampling matrix with rational elements in a discrete system. Take down-sampling as an example. Let $\mathbf{M}$ be the rational sampling matrix. It is easy to rewrite $\mathbf{M}$ as the multiplication of two matrices:

$$\mathbf{M} = \mathbf{M}_{up}^{-1} \cdot \mathbf{M}_{down} \tag{13}$$

where both $\mathbf{M}_{up}$ and $\mathbf{M}_{down}$ are integer matrices. Such a down-sampling process can be realized through Figure 4 (c), where we first up-sample the signal with matrix $\mathbf{M}_{up}$, then pass it through an up-sampling

filter, and finally down-sample the signal with matrix $\mathbf{M}_{down}$. The overall effect is that the signal is down-sampled with rational matrix $\mathbf{M}$. Rational matrix up-sampling can be implemented in the same way, as is shown in Figure 4 (d). For more detailed discussion, please refer to [9].

B.6.   Multidimensional filter design with eigenfilter

Multidimensional filter design has been studied for many years. We are especially interested in the FIR multidimensional filter design, because FIR filter can give the result of the filtering in a fixed amount of time, which is essential in Lightfield rendering. Most of the 1D FIR filter design technologies, such as the windowing method [8], the Parks-McClellan algorithm [10] and the eigenfilter approach [11][12], have been extended to multidimensional case. We take the eigenfilter approach throughout this paper.

The optimization criterion of eigenfilter approach is to minimize a quadratic measure of the passband and/or stopband error (sometimes also the transitionband error). Assume that we are designing 2D linear phase FIR digital filters. Let $D(\omega_1, \omega_2)$ be the desired frequency response in the passband, $G(\omega_1, \omega_2)$ be the one in the transition-band, and $H(\omega_1, \omega_2)$ be the filter we try to optimize. The Eigenfilter approach finds:

$$
\begin{aligned}
\arg\min_{H(\omega_1,\omega_2)} E &= \arg\min_{H(\omega_1,\omega_2)} \left\{ \alpha E_P + \beta E_S + \gamma E_T \right\} \\
&= \arg\min_{H(\omega_1,\omega_2)} \left\{ \alpha \iint_P [D(\omega_1,\omega_2) - H(\omega_1,\omega_2)]^2 \, d\omega_1 d\omega_2 \right. \\
&\qquad\qquad + \beta \iint_S [H(\omega_1,\omega_2)]^2 \, d\omega_1 d\omega_2 \\
&\qquad\qquad \left. + \gamma \iint_T [G(\omega_1,\omega_2) - H(\omega_1,\omega_2)]^2 \, d\omega_1 d\omega_2 \right\}
\end{aligned}
\tag{14}
$$

where E is the overall square error measured by the weighted sum of the passband error E_P, the stopband error E_S and the transitionband error E_T; α, β and γ are weighting constants which control the accuracies of the approximation. Usually, γ is set to be zero because we are not care about the transitionband response of the filter too much[1]. Equation (14) means that the eigenfilter approach minimizes the errors in the passband and the stopband in the square error sense. Through some manipulations, the overall error in equation (14) can be reformulated as [14][15]:

$$
E = \mathbf{a}^T \mathbf{Q} \mathbf{a}
\tag{15}
$$

where $\mathbf{a}$ is a real vector related with the 2D filter impulse response $h(n_1, n_2)$; $\mathbf{Q}$ is a real, symmetric and positive-definite matrix; T stands for transpose. It is well known that the eigenvector associated with the minimum eigenvalue of matrix $\mathbf{Q}$ minimizes the overall error.

---

[1] However, in some cases, the transition band constrain has to be included because otherwise the transition band response goes very bad.

III. Continuous Spatial-Domain Analysis

A. Overview of the previous work

When there is no occlusion and the scene is Lambertian, we can show that the intensity of a light ray entering a specific camera is equal to that of a corresponding light ray entering the camera at the origin. As shown in Figure 5, suppose we have cameras lying along a horizontal line x_t. Horizontal line x_v is the focal line of the cameras. To be consistent with [6], we set up local coordinate on the focal line base on where the camera is. Denote $\mathbf{x} = \begin{bmatrix} x_t & x_v \end{bmatrix}^T$ as a column vector. In the two-dimensional space where vector $\mathbf{x}$ is defined, each point maps one-to-one to a light ray unparallel to the two axes. Denote $s(\mathbf{x})$ as the intensity of the light ray corresponding to $\mathbf{x}$. If there is no occlusion and the scene is Lambertian, it is obvious from Figure 5 that

$$s(\mathbf{x}) = s\left(\begin{bmatrix} x_t \\ x_v \end{bmatrix}\right) = s\left(\begin{bmatrix} 0 \\ x'_v \end{bmatrix}\right) = s\left(\begin{bmatrix} 0 \\ x_v - fx_t/z(\mathbf{x}) \end{bmatrix}\right) \tag{16}$$

where $z(\mathbf{x})$ is the depth of the object surface hit by light ray $\mathbf{x}$, f is the focal length of the cameras.

An epipolar image (EPI) is constructed by plotting the intensity function $s(\mathbf{x})$ in the $\mathbf{x}$ space. An example EPI is shown in Figure 6, where we have a two-layer scene, with the contents of each layer be sinusoids for a finite length. Figure 6 (b) shows the EPI of the scene (left), and its Fourier transform (right). The sampling analysis is basically the Fourier analysis of the EPI. Starting from the simplest case, if a scene lies at a constant distance to the cameras, e.g., $z(\mathbf{x}) = z_0$, the Fourier transform of the EPI can be obtained by:

$$S(\mathbf{\Omega}) = S\left(\begin{bmatrix} \Omega_t \\ \Omega_v \end{bmatrix}\right) = \int_{-\infty}^{\infty} s(\mathbf{x})e^{-j\mathbf{\Omega}^T\mathbf{x}}d\mathbf{x} = \int_{-\infty}^{\infty} s\left(\begin{bmatrix} 0 \\ x_v - fx_t/z_0 \end{bmatrix}\right)e^{-j\mathbf{\Omega}^T\mathbf{x}}d\mathbf{x}$$
$$= 2\pi S'(\Omega_v) \cdot \delta(f\Omega_v/z_0 + \Omega_t) \tag{17}$$

where $S'(\Omega_v)$ is the 1D Fourier transform of the continuous signal observed at camera position $x_t = 0$, $\delta(\cdot)$ is the 1D Dirac delta function. For optimal sampling, we are only interested in the support of the frequency response of the EPI. Therefore, from Equation (17), we know that the frequency support of a scene at a constant depth lies along a line with slope $\Omega_v/\Omega_t = -z_0/f$.

Since any complex scenes can be decomposed into separate layers, the above deduction can be generalized to complex scenes *when there is no occlusion between layers*, in which case the Fourier transform of the whole scene is simply the sum of the Fourier transforms of all the layers. The overall frequency support of a complex scene is bounded by the two slopes determined by the maximum scene depth z_{max} and the minimum scene depth z_{min}. In Figure 6, we have a simple two-layer scene. At each layer, the original 1D signals are sinusoids with only two basic frequency components. Consequently, in the frequency domain of the EPI, we can see several peaks, which correspond to certain frequencies. These peaks lie on the two lines pretty well, and the slope of each line is determined by the layer's depth. One thing to notice is that there are some additional artifacts in the frequency response. The horizontal and vertical additional artifacts are

due to the windowing of the camera. The additional artifacts on the line corresponding to the farther layer, which is actually parallel to the line of the closer layer, is due to occlusion. For the analysis of non-Lambertian and occluded scenes, please refer to [17].

For now if we ignore the occlusion between layers, the frequency support of a general continuous EPI is shown in Figure 7 (a). We assume here without loss of generality that the resolution of the camera (capturing camera or rendering camera) or the contents of the EPI image, whichever is lower, is unity. Therefore, the sampling period along the focal line is unity, and the continuous frequency response of the EPI is bounded within $[-\pi, \pi]$ along Ω_v.

When a continuous signal is sampled in the spatial domain, its frequency response is replicated at a certain distance. The optimal sampling algorithm is to find the minimum replicas' distance in the frequency domain so that there is no aliasing effect. Figure 7 (b) shows the "optimal" sampling strategy proposed in [6]. Along the Ω_t axis, the frequency support is compacted as much as possible. The maximum sampling distance along the x_t line is:

$$\Delta x_{t\,max} = 2\pi/\Delta B \tag{18}$$

where

$$\Delta B = \pi f\left(1/z_{min} - 1/z_{max}\right) \tag{19}$$

where z_{max} and z_{min} are the maximum and minimum scene depth, respectively; f is the focal length of the cameras.

The reconstruction filter for the above sampling strategy is also given in Figure 7 (b). It is a tilted parallelogram, as marked with bold contours. It happens to be well approximated by the bilinear interpolation filter already used in the Lumigraph [4], where depth information of the scene is used to help improve the rendering quality. Moreover, through Figure 7 (b), we get the optimal rendering depth z_{opt} if only one depth can be used during the rendering, i.e., z_{opt} satisfies:

$$2/z_{opt} = \left(1/z_{max} + 1/z_{min}\right) \tag{20}$$

In the spatial domain, the explicit form of the ideal reconstruction filter is:

$$f_r(\mathbf{x}) = \frac{\sin\left(\Delta B x_t/2\right)}{\Delta B x_t/2} \cdot \frac{\sin \pi\left(x_v - fx_t/z_{opt}\right)}{\pi\left(x_v - fx_t/z_{opt}\right)} \tag{21}$$

It is the product of two sinc functions along two non-orthogonal directions, which we call the *principle directions* of the interpolation filter.

B. The optimal sampling strategy
B.1. The spectrum analysis
We notice that the sampling problem for image-based rendering is a multidimensional sampling problem. The question is, can we do better than the rectangular sampling approach in [6]? The answer is affirmative based on the sampling theory for multidimensional signal outlined in Section II-B.1.

Figure 8 shows how we can compact the frequency support and the replicas better. With this sampling strategy, the sampling efficiency can be improved by a factor of 2, which means we only need 50% of the samples. The reconstruction filter is a tilted fan-like filter, which will be discussed later.

Given the frequency support of an image-based rendering scene as in Figure 7 (a), the optimal replication lattice $\mathbf{U}$ can be easily found. We define:

$$\mathbf{U} = \begin{bmatrix} \pi f / z_{min} & -\pi f / z_{max} \\ -\pi & \pi \end{bmatrix} \tag{22}$$

where z_{max} and z_{min} are the maximum and minimum scene depth, respectively; f is the focal length of the cameras. This replication lattice compacts the frequency support of Figure 7 (a) into Figure 8, which is the optimal. In the spatial domain, the above replication in the frequency domain corresponds to a parallelepiped lattice sampling structure with sampling matrix $\mathbf{V}$:

$$\mathbf{V} = 2\pi \mathbf{U}^{-T} = \begin{bmatrix} 2z_e / f & 2z_e / f \\ 2z_e / z_{max} & 2z_e / z_{min} \end{bmatrix} \tag{23}$$

where z_e is an equivalent depth defined by $1/z_e = 1/z_{min} - 1/z_{max}$.

To show what happens in the spatial domain when we sample with the sampling matrix given in Equation (23), we show an example in Figure 9. The circles represent the data samples we keep after the sampling. The lattice is derived by assuming $z_{min} = \frac{4}{3} f$, and $z_{max} = 4f$. Therefore, our sampling matrix is $\mathbf{V} = \begin{bmatrix} 4 & 4 \\ 1 & 3 \end{bmatrix}$. As a comparison, the rectangular sampling approach has a sampling matrix $\mathbf{V} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$, which is diagonal. Comparing the number of samples per unit area, $1/|\det \mathbf{V}|$, of the two sampling matrices, it is obvious that the new approach saves half of the samples.

B.2. The continuous spatial-domain reconstruction filter

Image-based rendering is essentially the interpolation of sampled light rays in order to reconstruct the continuous IBR scene. Although many mature interpolation technologies exist, such as the bilinear interpolation, the cubic interpolation and the spline interpolation, they are mostly approximations of the ideal low-pass filter. As will be shown later, under the new sampling approach, the reconstruction filter has to be high-pass along some direction. Principally, the reconstruction filter is exactly the inverse of its frequency response. The tilted fan-like filter marked as bold contour in Figure 8 can be used directly to reconstruct the continuous signal. Although the explicit form of the tilted fan-like filter is available, the filter in Figure 8 is not fixed when the scene changes. A better way to reconstruct the continuous scene is to first take a transform from the $\mathbf{x}$ space to another space $\mathbf{y} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^T$ through:

$$\mathbf{y} = \mathbf{V}^{-1}\mathbf{x} \tag{24}$$

After the transform, sampling in the $\mathbf{x}$ space by sampling matrix $\mathbf{V}$ is equivalent to sampling in the $\mathbf{y}$ space by a 2x2 identity matrix, and the sampling index is given by the integer vector $\mathbf{n}$ in Equation (6).

The Fourier transform of the sampled signal in the $\mathbf{y}$ space is shown in Figure 10 (a). The reconstruction filter is drawn in bold in Figure 10 (a) and its spatial domain response is shown in Figure 10 (b). The explicit form of the spatial domain reconstruction filter is:

$$f_r(y_1, y_2) = \frac{y_2 \sin^2 \pi y_1 + y_1 \sin^2 \pi y_2}{\pi^2 y_1 y_2 (y_1 + y_2)} \tag{25}$$

Unfortunately, this reconstruction filter is not a low pass filter. Along the direction where $y_1 + y_2 = 0$, this filter can be represented as:

$$\lim_{y_1 + y_2 \to 0} f_r(y_1, y_2) = \frac{\sin \pi y_1 (2\pi y_1 \cos \pi y_1 - \sin \pi y_1)}{\pi^2 y_1^2} \tag{26}$$

which is a high pass filter and has a minimum peak of value –0.587 around $y_1 = -y_2 = \pm 0.65$. In practice, both the reconstruction filter for the previous approach in [6] and the one for the proposed approach have infinite support. Truncation or other techniques have to be employed to roughly reconstruct the original continuous signal.

IV. Discrete Spatial-Domain Analysis

In the last section, we discussed about the sampling of the 2D Lightfield in the continuous spatial-domain. However, given a real scene, typically the minimum and maximum distances of the scene are unknown. At the first try of capturing the scene, a more natural approach is to first take images for the scene as many as possible, and then analyze the captured data and try to remove the redundant images that have been taken. This gives rise to the analysis of IBR sampling problem in the discrete spatial-domain, which we believe is more practical than the continuous analysis given in the last section.

A. The discrete spatial-domain sampling framework

We assume that we are given a set of images captured at equal distances on the camera line in Figure 5. The distance between neighboring cameras is small enough so that there is no aliasing effect between the replicated spectrums. The Fourier transform of the discrete signal is shown in Figure 11. The frequency response is still bounded by two lines passing the origin, with slopes $-k_{min}$ and $-k_{max}$, as has been stated in Section III-A. However, since the distance between neighboring cameras may not be the same as the distance between neighboring pixels in the captured images (as was assumed in Section III), $k_{min} = z_{min}/f$ and $k_{max} = z_{max}/f$ are not necessary held. Nevertheless, k_{min} and k_{max} can be easily obtained through the Fourier analysis of the original discrete signal, and we do not care about what z_{min} and z_{max} really are in the following analysis[2].

The sampling framework is illustrated in Figure 12. Starting with the original discrete signal, we first analyze the spectrum so that we know what are k_{min} and k_{max}. We then down-sample the data by matrix

---

[2] The relation between z_{min}, z_{max} and k_{min}, k_{max} can be easily obtained if we know the sampling distance of the camera with respect to the camera resolution.

$\mathbf{M}$ after the down-sampling filter[3]. The down-sampled data are then stored as a compact representation of the IBR scene. To reconstruct the continuous IBR scene for rendering, two approaches can be employed. In approach A, we first try to reconstruct the original discrete signal by up-sampling the compact data with matrix $\mathbf{M}$ followed by an up-sampling filter. After that, a continuous interpolation filter H_A is designed to get the continuous signal for rendering. Since the original discrete signal is supposed to be over-sampled, the interpolation filter H_A is easy to design. Approach B combines the three steps in approach A and designs only one continuous interpolation filter H_B to reconstruct the continuous signal from the compact representation directly. The advantage of approach A is that all the major filter designs are done in the discrete domain, which has been studied extensively in the signal processing literature. If the reconstructed original discrete signal can be held in the main memory completely, rendering the scene can be very fast because the continuous interpolation filter H_A can be very simple due to the over-sampling. For example, H_A can be simply bilinear interpolation. On the other hand, approach A has to reconstruct the original discrete signal before any rendering can be done, and the storage space for the reconstructed data is extremely huge. As a comparison, approach B requires much less memory during the rendering, because only the compact representation of the scene is required. The price of saving the memory space is to have a complex interpolation filter B, which slows down the rendering speed.

We want to compare the sampling approach in [6] and the proposed approach using periodic sampling with arbitrary geometry. As has been shown in Section III-B, the most aggressive compacting of the spectrum with the proposed approach can save half of the samples with respect to the previous approach. Theoretically, such aggressive compacting causes no aliasing and the signal can be perfectly reconstructed. However, infinite-length reconstruction filter is required and thus impractical. In this section, we compare the two approaches by making them have the same sampling density or the same number of samples per unit area. We show that under the same setting, the filter of the proposed approach is easier to design, thus having smaller design and reconstruction errors compared with the previous approach.

B. The down-sampling of the given data

Given that the original discrete signal has a frequency response bounded by two lines passing the origin and with slopes $-k_{\min}$ and $-k_{\max}$, as is shown in Figure 14, we can easily obtain the down-sampling matrix for the previous approach in [6] and the proposed approach. For the previous approach, the duplication matrix in the frequency domain is defined as:

$$\mathbf{N}_1 = \begin{bmatrix} \pi/k_{e1} & -2\pi/k_{e2} \\ 0 & 2\pi \end{bmatrix} \tag{27}$$

where k_{e1} is defined by $1/k_{e1} = 1/k_{\min} - 1/k_{\max}$ and k_{e2} is defined by $2/k_{e2} = 1/k_{\min} + 1/k_{\max}$. The sampling matrix in the spatial domain can be achieved as:

---

[3] The discussion of the down-sampling filter is ignored in this paper. We assume that it has been done before we get the original discrete signal, such as what the camera's limited resolution does.

$$\mathbf{M}_1 = 2\pi\mathbf{N}_1^{-T} = \begin{bmatrix} 2k_{e1} & 0 \\ 2k_{e1}/k_{e2} & 1 \end{bmatrix} \tag{28}$$

Notice that matrix $\mathbf{M}_1$ is not diagonal. We deliberately make it so that the Fourier transform of the down-sampled data has a standard fan shape, as is shown in Figure 13 (a). This eases the design of the interpolation filter.

For comparison, we let the proposed sampling approach have the same sampling density as the previous one. If the new sampling matrix is $\mathbf{M}_2$, we want to have $\left| \det \mathbf{M}_1 \right| = \left| \det \mathbf{M}_2 \right|$. In the frequency domain, if the new duplication matrix is $\mathbf{N}_2$, then $\left| \det \mathbf{N}_1 \right| = \left| \det \mathbf{N}_2 \right|$. From Figure 11, we can obtain that the new duplication matrix should be:

$$\mathbf{N}_2 = \begin{bmatrix} \pi/k_{\min} + \pi/2k_{e1} & -\pi/k_{\max} + \pi/2k_{e1} \\ -\pi & \pi \end{bmatrix} \tag{29}$$

And the sampling matrix in the spatial domain is:

$$\mathbf{M}_2 = 2\pi\mathbf{N}_2^{-T} = \begin{bmatrix} k_{e1} & k_{e1} \\ k_{e1}/k_{\max} - 1/2 & k_{e1}/k_{\min} + 1/2 \end{bmatrix} \tag{30}$$

It is easy to verify that $\left| \det \mathbf{M}_1 \right| = \left| \det \mathbf{M}_2 \right|$. The down-sampled signal has a standardized spectrum as in Figure 13 (b).

In the discrete down-sampling process, the down-sampling matrix $\mathbf{M}$ typically has to be an integer matrix. In this case, all we need to do is to remove those samples that are not on the down-sampling grid, or, those samples whose index cannot be written as $\mathbf{Mn}$, where $\mathbf{n} \in \mathbb{N}$. Unfortunately, neither $\mathbf{M}_1$ nor $\mathbf{M}_2$ is guaranteed to be an integer matrix. There are two ways to solve this problem. One of them is to sacrifice the sampling efficiency and force $\mathbf{M}_1$ and $\mathbf{M}_2$ to be integer matrices with a higher sampling density than they need to be. For example, if only down-sampling along the camera line is considered, as in [6], one of the best integer sampling matrices (in the sense of lowest sampling density and integer elements) is:

$$\mathbf{M}_1 = \begin{bmatrix} \lfloor 2k_{e1} \rfloor & 0 \\ 0 & 1 \end{bmatrix} \tag{31}$$

where $\lfloor \alpha \rfloor$ stands for the largest integer that is smaller than α.

When arbitrary geometry sampling is employed, we may be able to find integer sampling matrix that gives a lower sampling density than that of the sampling matrix in Equation (31). Since the entries of the sampling matrix are all integers, exhaustive search can be used to find the optimal matrix that has the smallest sampling density without aliasing effect.

A more elegant way to deal with non-integer down-sampling matrix is through the rational matrix sampling theory introduced in Section II-B.5. With this approach, we first approximate the sampling matrix with a rational matrix, and then employ the theory in the discrete domain. Since this is a mature research area in the signal processing literature, we refer the reader to [9]. In the following sections, we assume that we are able to take this approach to down-sample and up-sample the data. Therefore, the down-sampling matrices

of the previous approach in [6] and the new approach are taken exactly as Equation (28) and Equation (30). $\left|\det \mathbf{M}_1\right| = \left|\det \mathbf{M}_2\right|$ is satisfied and we can have a fair comparison between these two down-sampling approaches.

C. The Approach A

The flow of Approach A is shown in Figure 12. We up-sample the stored signal with the same matrix $\mathbf{M}$ as before. After the up-sampling, the signal has the same density as the original discrete signal before down-sampling. In the frequency domain, the Fourier transform of the up-sampled signal is a duplication of that of the original discrete signal, as is shown in Figure 14. Figure 14 (a) shows the Fourier transform of the original signal; Figure 14 (b) shows the Fourier transform of the up-sampled signal with the previous approach; Figure 14 (c) shows the Fourier transform of the up-sampled signal with the new approach. Notice the difference between (b) and (c) for the arrangement of duplications.

When we design the up-sampling filter for the up-sampled data in order to reconstruct the original discrete signal, the two different approaches have different design specifications. Figure 14 (d) and Figure 14 (e) shows the difference, where Figure 14 (d) corresponds to the previous approach and Figure 14 (e) corresponds to the proposed approach. P, S and T stand for passband, stopband and transitionband, respectively. We claim that the up-sampling filter for Figure 14 (d) is easier to design than the filter for Figure 14 (e). Therefore, with spectrum distribution of Figure 14 (c), we can reconstruct the signal better.

We give an example to verify our claim. Assume we have a scene with $k_{\min} = 4/3$ and $k_{\max} = 4$. Therefore, with the previous approach, the sampling matrix is $\mathbf{M}_1 = \begin{bmatrix} 4 & 0 \\ 2 & 1 \end{bmatrix}$; with the proposed approach, the sampling matrix is $\mathbf{M}_2 = \begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix}$.

In the literature on 2D FIR Eigenfilter design [14][15], quadrantal symmetric FIR filters design was discussed predominantly. However, the Fourier transform of the impulse response of our reconstruction filter has symmetric axis $H(\omega_t, \omega_v) = H(-\omega_t, -\omega_v)$ only, as is observed in Figure 14 (d) and (e). The design process has to be modified to produce the correct result.

We employ the eigenfilter algorithm on the design of the filters in Figure 14 (d) and (e), where the two bounding slopes are $k_{\min} = 4/3$ and $k_{\max} = 4$. The passband response is set to be a constant, and the reference point is chosen to be at $(-\pi/4, \pi/2)$. We find that with the specification given in Figure 14 (d), sometimes the result is really bad if there is no transitionband constrain. For example, if we design an 11x11 eigenfilter for Figure 14 (d), we get the Fourier transform of the resultant filter as in Figure 15 (a). The two peaks in Figure 15 (a) correspond to the two transitionbands, while the passband and stopband are both close to zero. Because the eigenfilter method does not control the amplitude of the filter in the passband, which happens to result in the minimum overall error of Equation (14) when the passband response is close to zero. In contrast, Figure 15 (c) shows the same filter design result for Figure 14 (e), which is more reasonable.

To overcome the above problem, we add transitionband constrain $G(\omega_t, \omega_v)$ as the linear horizontal slope connecting the passband and the stopband. The transitionband error weight is set to be 0.1. The resultant design outputs are shown in Figure 15 (b) and (d). Both of them are very reasonable. Eigenfilter does not output a filter with unit passband amplitude. To compare different filter design results, we scale the reference point response to unit amplitude, and the corresponding overall square error given in Equation (14) is used as the criterion of the performance. As is shown in Figure 16, the horizontal axis is the order of the designed filter, e.g., an order of 5 means we design a 5x5 filter that tries to give the frequency response in Figure 14 (d) and (e). The vertical axis is the overall square errors. A curve closer to the origin performs better because it brings smaller square error with respect to the specification. It is obvious that the new filter design specification of Figure 14 (e) is easier to fulfill and thus gives a smaller design error.

D. The Approach B

In Approach A, we first reconstruct the original over-sampled signal in the discrete domain, and then get the continuous signal though simple interpolations such as bilinear interpolation. One of the disadvantages of Approach A is that it requires huge memory to store the reconstructed original signal. Considering that IBR typically has a very huge data amount, Approach A is not practical in real applications. A much more attractive approach would be to generate the continuous signal from the down-sampled data directly. Approach B is such an approach. Basically Approach B combines the steps in Approach A and simplifies the problem into the design of a continuous interpolation filter.

Notice that after down-sampling with matrix $\mathbf{M}$, the spatial domain has been distorted. Given a light ray coordinate $\mathbf{x} = \begin{bmatrix} x_t & x_v \end{bmatrix}^T$ in the original spatial domain to be rendered, we first employ a transform on $\mathbf{x}$, i.e., let $\mathbf{y} = \mathbf{M}^{-1}\mathbf{x}$. After the transform, the $\mathbf{y}$ space is coincident with the down-sampled signal. All we need to do is to design discrete-to-continuous interpolation filters for the spectrums given in Figure 13.

Since the explicit form of the optimal continuous interpolation filter is not available, we first design a discrete filter and then get the continuous filter through interpolation. To achieve a good approximation of the optimal continuous filter, we up-sample the discrete filter by a factor of M in each dimension through zero padding, which is equivalent to squeeze the pass band and transition band of the filter specification by M, as is shown in Figure 17 (b). After the design, we scale the filter back by M in the spatial domain; hence we have the filter response on some non-integer points. Bilinear interpolation is followed to obtain the spatial response at arbitrary points. We call this process *interpolated eigenfilter design*, which is essentially the same as the interpolated FIR filter design approach in [13]. Again we use eigenfilter to design the optimal discrete filter we want. Figure 18 shows one of the design results in the spatial domain. The filter has window size 4x4, and we up-sample the filter by 8x8. The discrete filter we designed has order 33x33, where $33 = 8 \times 4 + 1$.

E. Enhanced Eigenfilter Design for Optimal Reconstruction

We can improve the reconstructed image quality through a modified eigenfilter approach. This approach applies for both Approach A and B above. We know that 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 14 (b) and (c)) be $X(\omega_v, \omega_t)$. We try to find the optimal reconstruction filter (ORF) through:

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

$$= \arg\min_{H(\omega_v,\omega_t)}\left\{\iint_P \left|X(\omega_v,\omega_t)\right|^2 \left|D(\omega_v,\omega_t) - H(\omega_v,\omega_t)\right|^2 d\omega_v d\omega_t \right. \tag{32}$$

$$\left. + \iint_S \left|X(\omega_v,\omega_t)\right|^2 \left|H(\omega_v,\omega_t)\right|^2 d\omega_v d\omega_t\right\}$$

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 (32) can still be solved through the eigenfilter approach, as it differs from Equation (14) only by a weighting function. Notice that the optimal filter is related with the signal $X(\omega_v, \omega_t)$. 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:

$$\left|X(\omega_v, \omega_t)\right|_P = \left|\frac{1}{1 - \rho e^{j\omega_v}}\right| \tag{33}$$

where the subscript P simply means that it is valid only in the pass-band.

V. Experimental Results

In this subsection, we show some experimental results on the optimal IBR sampling approach. In order to have full control on the scenes and the cameras, we choose a synthetic scene and two scenes rendered from 3D models with texture. These scenes are shown in Figure 23, where scene (a) is the synthetic scene and scene (b) and (c) are 3D model scenes named *Duck* and *Containers*, respectively. Since we focus on 2D lightfield in this paper, we take the center horizontal line to construct the EPIs. Figure 23 (a1), (b1) and (c1) are snapshots of the scenes; (a2), (b2) and (c2) are their EPIs; (a3), (b3) and (c3) are the Fourier transform of the EPIs. Although occlusions can be observed in the scenes, we analyze them as if occlusions do not happen since the ringing artifacts caused by occlusions are not serious.

The synthetic scene has two layers, with 4 sinusoids on each layer. The closer layer has finite width occluding the other layer with infinite width. The depth of the layers are $z_0 = 1.1f$ and $z_1 = 2.8f$, where f stands for the focal length. Considering the ringing artifacts, we leave a margin for the frequency support and assume that the support is bounded by two slopes $-z_0/f = -1.0$ and $-z_1/f = -3.0$. Follow the discussion given in Section IV-B, for fair comparison, the two down-sampling matrix are $\mathbf{M}_1 = \begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}$ for the previous approach, and $\mathbf{M}_2 = \begin{bmatrix} 3/2 & 3/2 \\ 0 & 2 \end{bmatrix}$ for the proposed approach. They have the same number

of samples per unit area. As $\mathbf{M}_2$ is not an integer matrix, we interpolate from the original EPI to get the samples we want to preserve, as was discussed in Section II-B.5 as rational down-sampling.

The two scenes rendered with 3D textured models are aimed to simulate real-life lightfields. The originally sampled EPIs have roughly the same support of frequency responses. These responses are bounded by $-k_{\min} = -0.9$ and $-k_{\max} = -1.3$. The down-sampling matrices can be easily obtained as $\mathbf{M}_1 = \begin{bmatrix} 5.8500 & 0 \\ 5.500 & 1 \end{bmatrix}$

for the previous approach and $\mathbf{M}_2 = \begin{bmatrix} 2.9250 & 2.9250 \\ 1.7500 & 3.7500 \end{bmatrix}$ for the proposed approach. We control the camera

and recapture the scene with distance 5.8500 times and 2.9250 times of the original distances when we construct the original EPIs. After that, rational down-sampling is employed to get the desired overall down-sampling results.

In this experiment, we mainly test the ability to render the scene properly with the proposed Approach B, since this is the more practical approach. Notice that if we want to reconstruct the original EPI as in Approach A, Approach B is also applicable since we only need to render the images at those integer points along axis x_t where the original EPI is captured[4]. Starting from different down-sampled data, we test different interpolation filters in this experiment. For data down-sampled with matrix $\mathbf{M}_1$, we test interpolated eigenfilters designed for Figure 13 (a), while for data down-sampled with matrix $\mathbf{M}_2$, we test interpolated eigenfilters designed for Figure 13 (b).

As we use synthetic and virtual scenes to perform our experiment, we are able to evaluate the IBR quality by comparing the rendered images and what they should be if we really capture them. We randomly choose 1000 camera positions along the camera axis and render their images with our optimally sampled data. These images are then synthesized or virtually recaptured through the 3D model rendering engine. The difference between the synthesized images and the rendered images is a good measurement for the quality of the sampled data under a certain rendering strategy. PSNR is used to measure such differences.

Table I shows the results of different filters for different down-sampled data. For simplicity, we use D1 and D2 to represent downsamplings with matrix $\mathbf{M}_1$ and $\mathbf{M}_2$. W(M,l) stands for interpolation eigenfilter designed with upsampling factor M and window size l by l (See Section IV-D). For example, W(8,4) means we upsample the desired filter by a factor of 8, and a discrete filter of support 33x33 ($33 = 8 \times 4 + 1$) is designed, which is then bilinearly interpolated to get the continuous interpolation filter with window size 4x4. The character T means transitionband constrain is added, which will be addressed later.

To design the interpolation eigenfilter, we need to upsample the filter first, thus squeeze the filter design specifications. Notice that after squeezing, Figure 13 (a) becomes a problematic specification. Although for the W(10,2) setting the resultant intermediate discrete filter looks good and its frequency response seems

---

[4] The Approach A and Approach B can be considered as a dual problem. While Approach A up-sample the signal and interpolate the signal, Approach B up-sample the filter and interpolate the filter. Essentially they are the same approach.

reasonable, the same problem as Figure 15 (a) happens when we change to W(8,4). Moreover, the rendering quality with the filter of W(10,2) setting is very bad. Therefore, we add transitionband constrains for the filter design of Figure 13 (a) and list the experimental results in Table I, too. For the proposed approach's specification in Figure 13 (b), no transitionband constrain is needed. From Table I we can see that the rendering quality of the proposed approach is always better than the previous approach, with an improvement of 0.5-1.7 dB.

We then apply the optimal reconstruction filter (ORF) design for the scenes. The results are shown in Table II. To give a complete comparison, the results for rectangular sampling with bilinear interpolation are also reported. For different eigenfilter designs, the best results are reported here for different downsample/upsample factor combinations. Several conclusions can be drawn from the table. First, rectangular sampling with D1 and regular eigenfilter performs much worse than generalized sampling with regular eigenfilter, which has been shown before. 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 D1 and D2 becomes very small, which again shows the power of ORF. Interestingly, the simple approach that uses rectangular sampling with bilinear interpolation gives comparable performance as generalized sampling 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 generalized sampling for the scene *Containers*, because *Containers* has more high frequency components in its spectrum.

VI. Conclusions and Future Work

In this paper, we gave a complete analysis of the optimal sampling problem. First, we improved the sampling efficiency by employing the generalized periodic sampling theory with arbitrary geometry. When there is no occlusion and the scene is Lambertian, we showed that the sampling density in the spatial domain could be twice of that when we use rectangular sampling. We then proposed a framework for IBR sampling. Two approaches to render the continuous IBR scene are discussed in detail. When the proposed approach has the same down-sampling density with the traditional rectangular sampling, we showed that the reconstruction filter for the proposed approach is easier to design, and thus gives higher rendering quality.

However, with optimal reconstruction filter, the difference between rectangular sampling and the proposed approach is minor. Rectangular sampling with bilinear interpolation even sometimes gives better rendering quality than the complex proposed method.

If we re-examine Figure 14 (b) and (c), we can observe that the transition-band of the up-sampling filter for rectangular 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 generalized sampling has constant width along all the frequencies. This observation implies that rectangular sampling is better for scenes that have less high-frequency components, while generalized sampling is better otherwise. Real world scenes often has weak

high frequency components and strong low frequency components due to reasons such as occlusions, thus rectangular sampling is usually more suitable than generalized sampling. Even for the *Containers* scene, the improvement by using the latter is minor, which cannot justify the increased complexity in the latter. Since generalized sampling 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.

Acknowledgement

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]   H.Y. Shum and L.-W. He. "Rendering with concentric mosaics", *Computer Graphics (SIGGRAPH'99)*, pp.299-306, Aug. 1999.

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

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

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

[9]   P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-hall signal processing series, 1993.

[10] Thomas W. Parks and James H. McClellan, "Chebyshev Approximation for Nonrecursive Digital Filters with Linear Phase", *IEEE Trans. Circuit Theory*, pp. 189-194, CT-19, No. 2, Mar. 1972.

[11] 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.

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

[13] Y. Neuvo, C. Y. Dong and S. K. Mitra, "Interpolated Finite Impulse Response Filters", *IEEE Trans. on ASSP*, pp. 563-570, June 1984.

[14] S.-C. Pei and J.-J. Shyu, "2-D FIR Eigenfilters: A Least-Squares Approach", *IEEE Trans. on Circuits and Systems*, pp. 24-34, Vol. 37, No. 1, Jan. 1990.

[15] A. Nashashibi and C. Charalambous, "2-D FIR Eigenfilters", *Proc. IEEE Int. Symp. Circuits and Systems*, Finland, pp. 1037-1040, June 1988.

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[17] C. Zhang and T. Chen, "Spectral Analysis for Sampling Image-Based Rendering Data", *Carnegie Mellon Technical Report: AMP03-01*.

Tables

Table I Rendered images qualities for different interpolation filters and different down-sampling methods in Approach B

	Synthetic Scene	Duck	Containers
D1 – W(10,2)	20.31 dB	29.87 dB	18.46 dB
D1 – W(10,2) – T	21.29 dB	32.87 dB	20.42 dB
D2 – W(10,2)	23.02 dB	33.58 dB	20.90 dB
D1 – W(8,4)	-	-	-
D1 – W(8,4) – T	24.86 dB	33.99 dB	20.44 dB
D2 – W(8,4)	27.30 dB	35.68 dB	22.04 dB

Table II Rendered images qualities for different sampling methods and reconstruction filters

(PSNR: dB)

	Duck	Containers
D1, regular eigenfilter	33.99	20.44
D1, optimal reconstruction filter	35.89	21.81
D2, regular eigenfilter	35.68	22.04
D2, optimal reconstruction filter	36.06	22.21
D1, bilinear interpolation	36.10	21.67

Figures
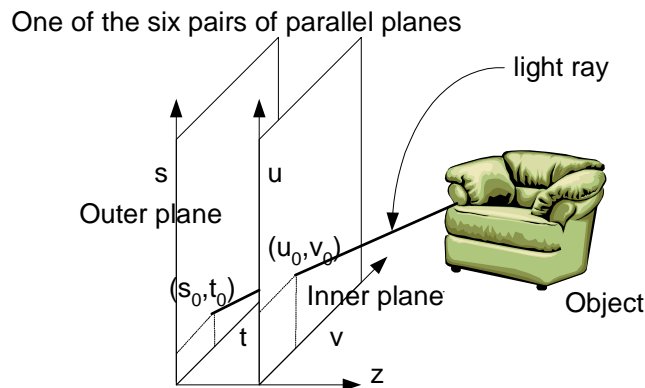
Figure 1 The parameterization of the Lightfield.
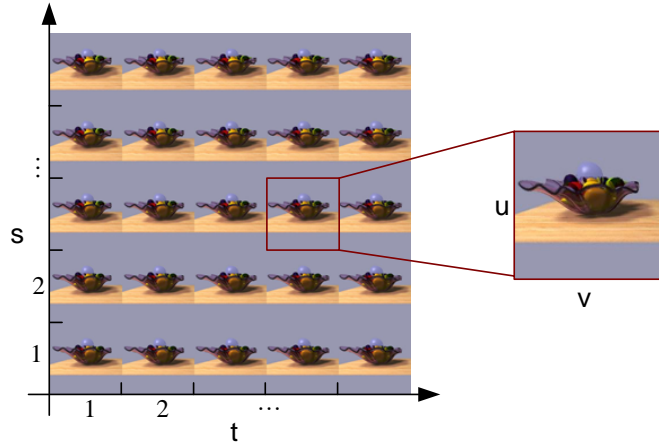
Figure 2 A sample Lightfield image array: fruit plate.

(a)

(b)
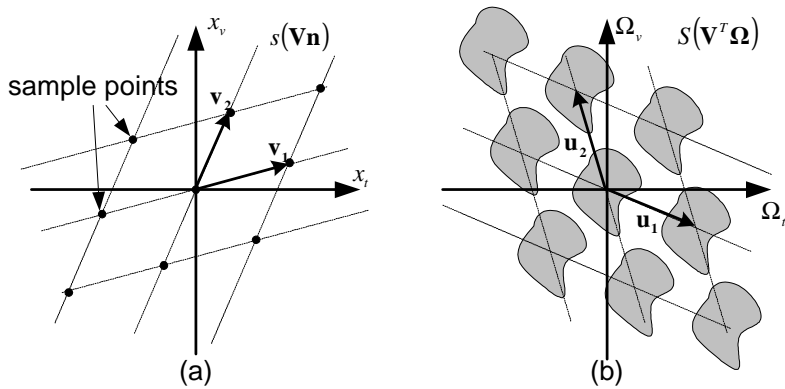
Figure 3 Generalized sampling theorem in 2D space. (a) Spatial sampling on a lattice with arbitrary geometry. (b) Frequency domain replication along a corresponding lattice.
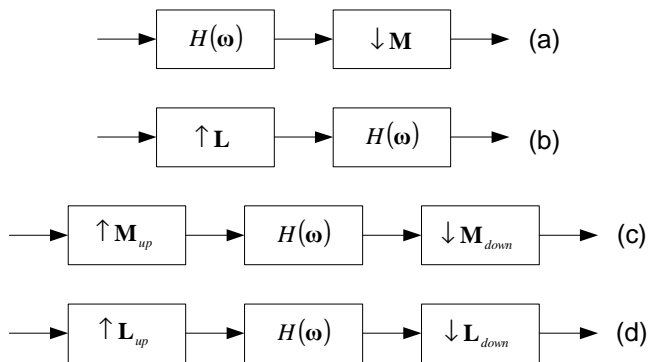
Figure 4 (a) the down-sampling system (b) the up-sampling system (c) down-sampling with rational matrix (d) up-sampling with rational matrix.
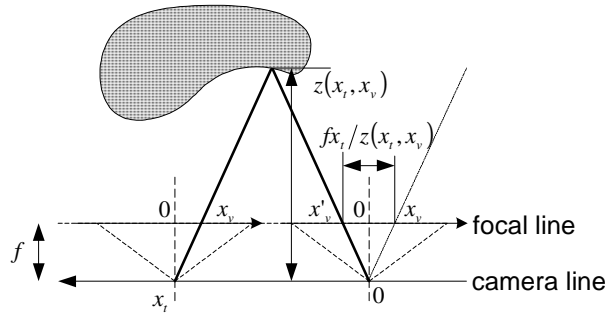
Figure 5 The light ray correspondence.
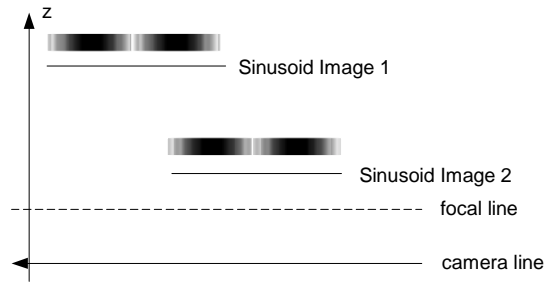
(a)
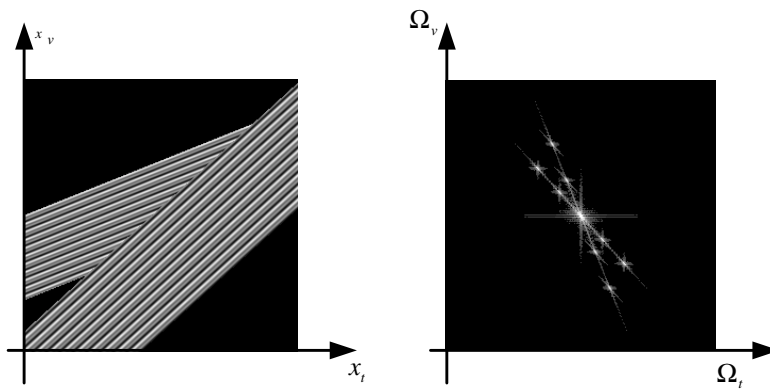
(b)

Figure 6 A two layer scene, its EPI and the Fourier transform of the EPI. (a) A scene with two layers. (b) The EPI (left) corresponding to the scene in a) and its Fourier transform (right).
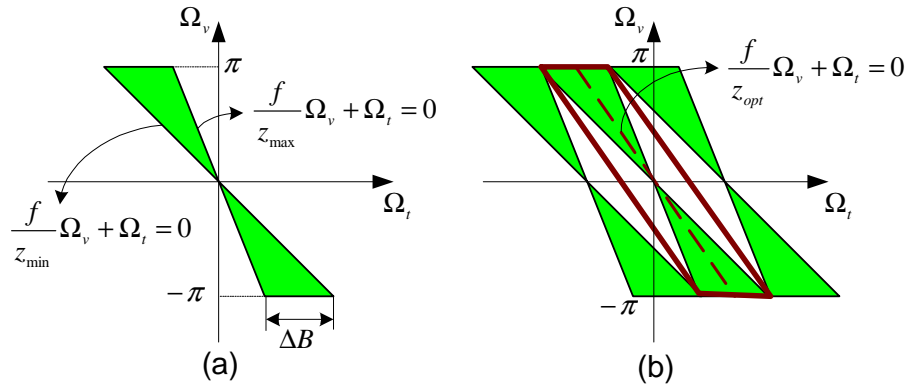
Figure 7 Frequency response of a no occlusion EPI and the "optimal" sampling in [6]. (a) Frequency support of a scene with no occlusion and Lambertian model. (b) The "Optimal" compacting and the reconstruction filter in [6].
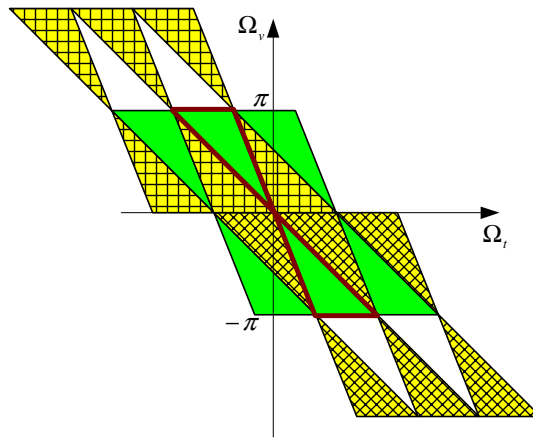
Figure 8 Optimal sampling for image-based rendering.
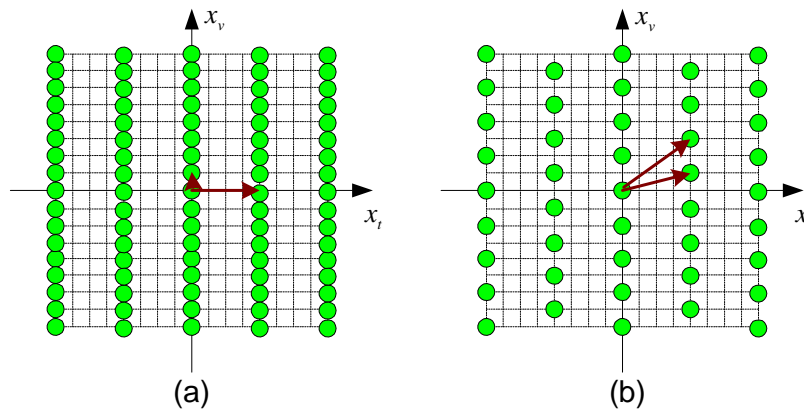
Figure 9 Comparison of the rectangular sampling approach with our new approach. (a) the new approach (b) the proposed approach
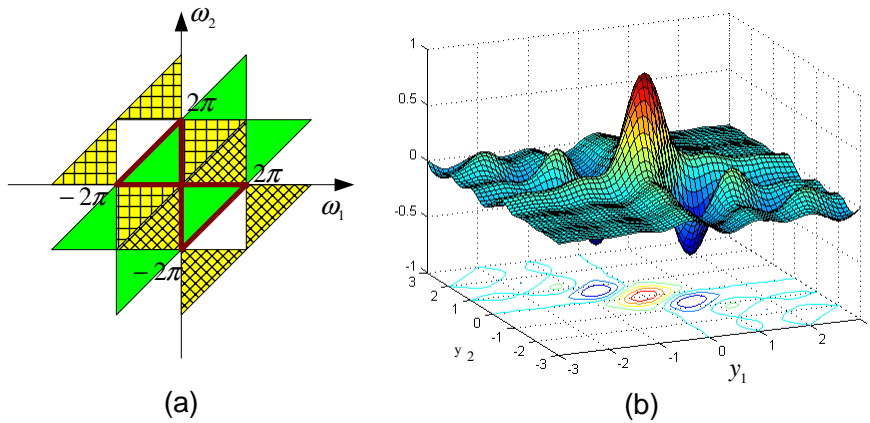
Figure 10 The reconstruction filter. (a) The frequency support and optimal compacting in the transformed frequency space. (b) The standard reconstruction filter.
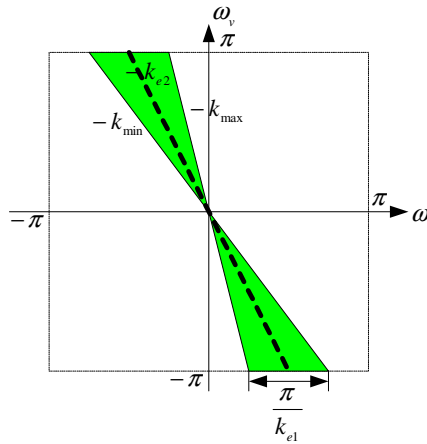
Figure 11 The Fourier transform of the original discrete signal.
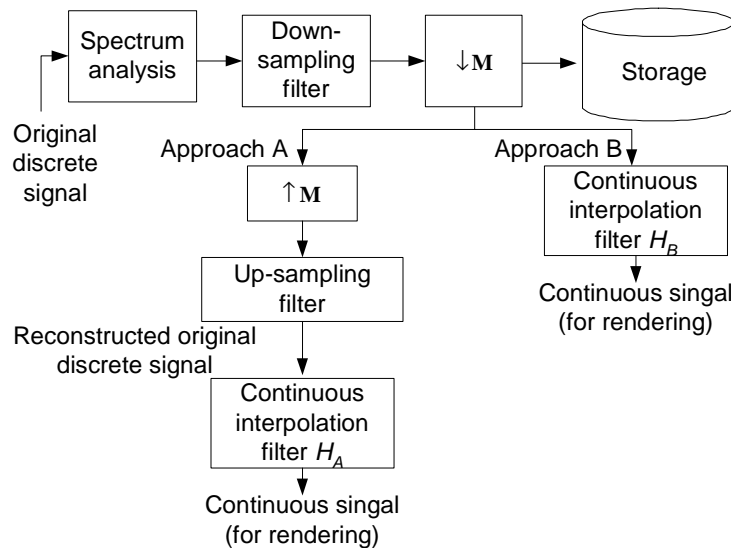
Figure 12 The discrete spatial-domain sampling framework. There are two approaches to reconstruct the continuous signal for the rendering of IBR.
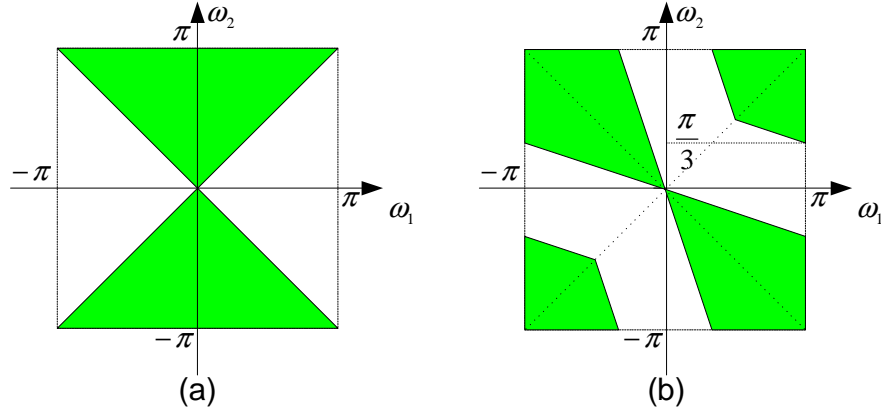
Figure 13 The Fourier transforms of the down-sampled data. (a) With the previous approach in [6]. (b) With the proposed approach.
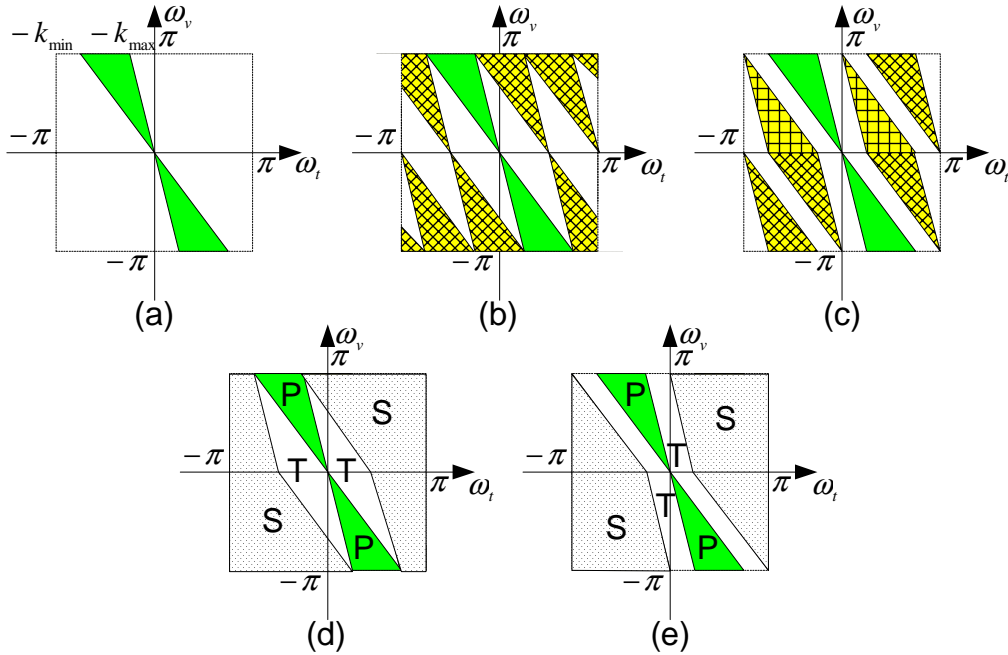
Figure 14 (a) The Fourier transform of the original discrete signal. (b) The Fourier transform of the signal after the down-sampling as in [6]. (c) The Fourier transform of the signal after down-sampling with the new approach. Note: (b) and (c) have the same down-sampling density. (d) Up-sampling filter specification for the previous approach. (e) Up-sampling filter specification for the proposed approach.
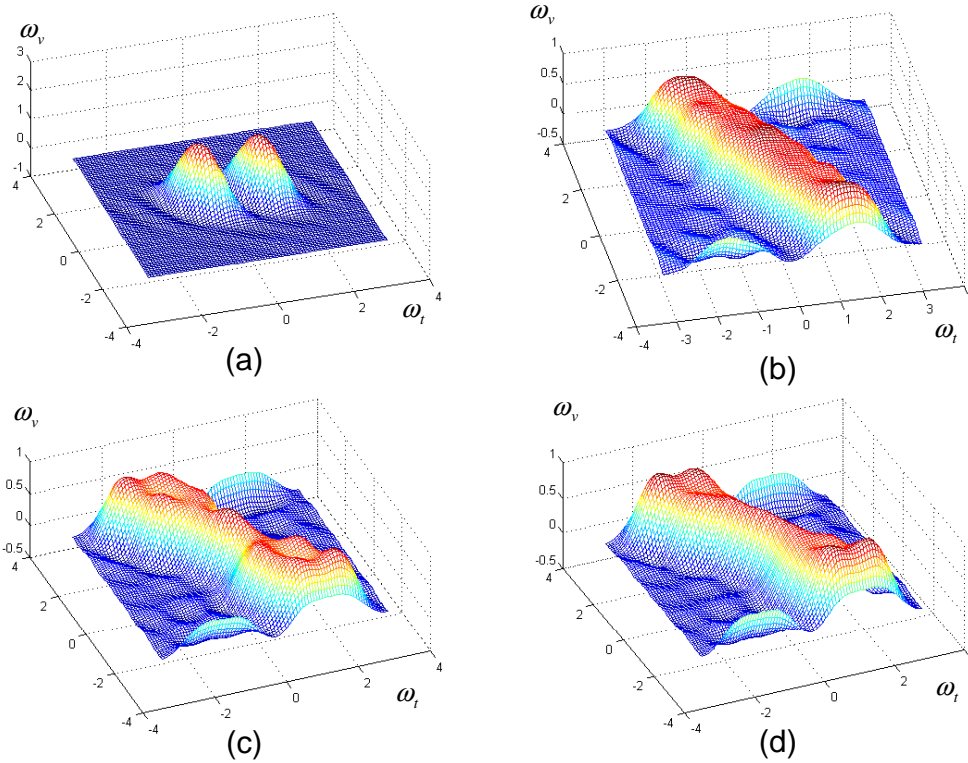
Figure 15 Filter design results for different specifications and methods. (a) Specification Figure 14 (d), no transitionband constrain. (b) Specification Figure 14 (d), with transitionband constrain. (c) Specification Figure 14 (e), no transitionband constrain. (d) Specification Figure 14 (e), with transitionband constrain.
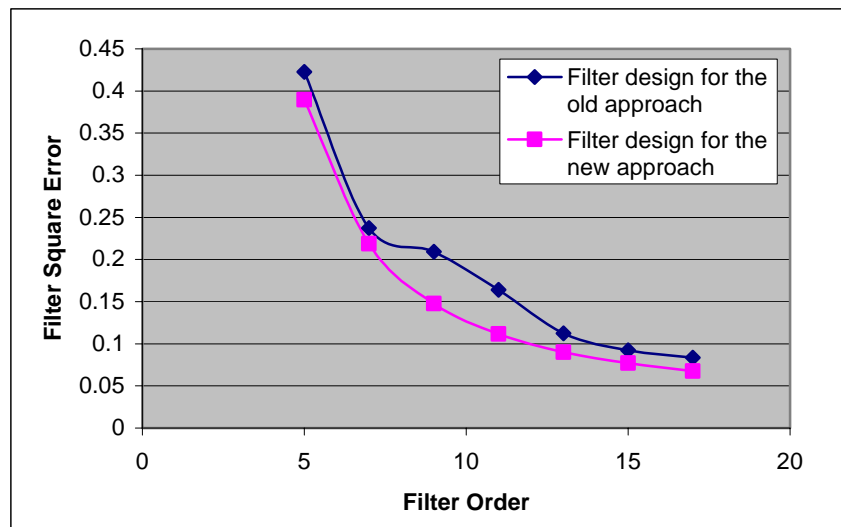
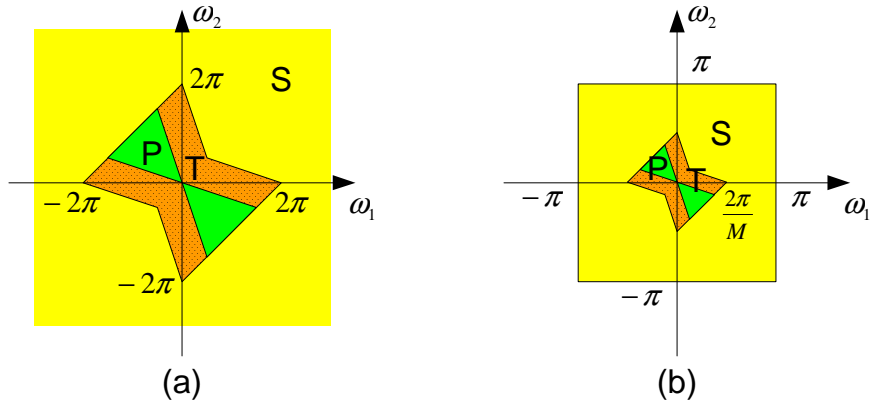**Figure 16 The comparison of the filter design for the old approach and the new approach.**

Figure 17 The design of the reconstruction filter. (a) The passband (P), the stopband (S) and the transition-band (T). (b) The corresponding digital filter has frequency response on $[-\pi, \pi] \times [-\pi, \pi]$. After up-sample, the passband and the transition-band are squeezed.
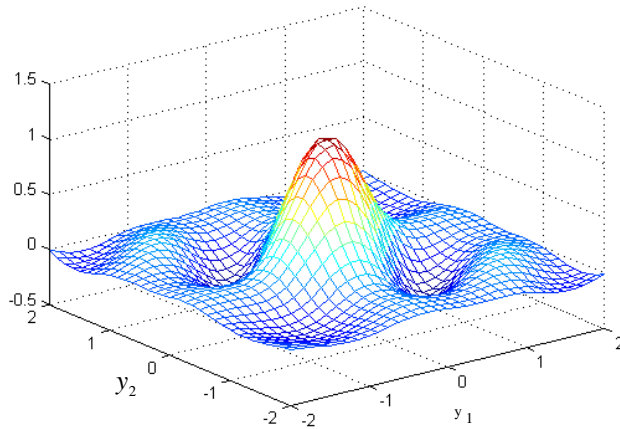
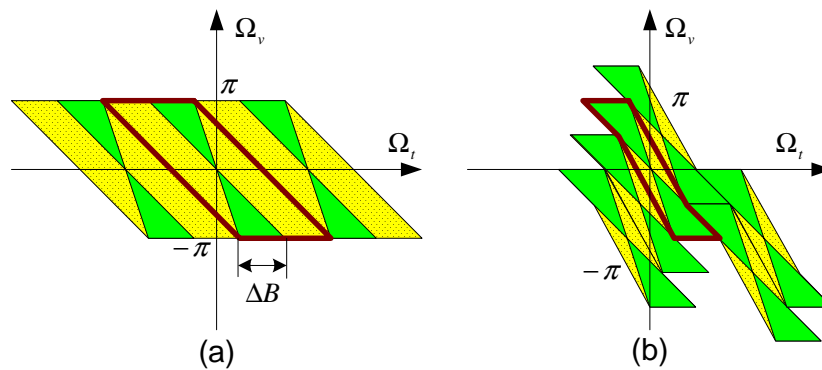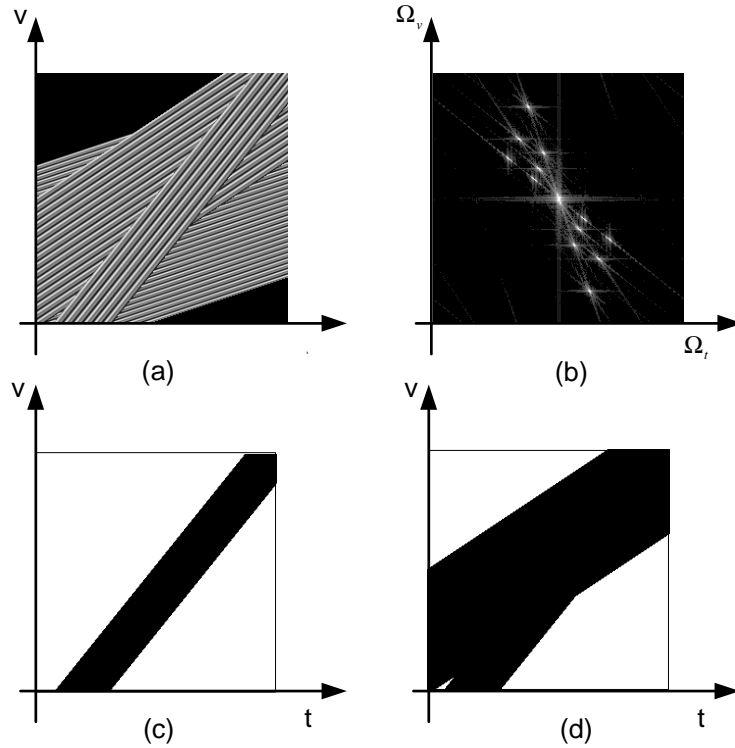**Figure 18 The optimal digital filter designed by Eigenfilter, with 4x4 window size, up-sampled by 8x8.**
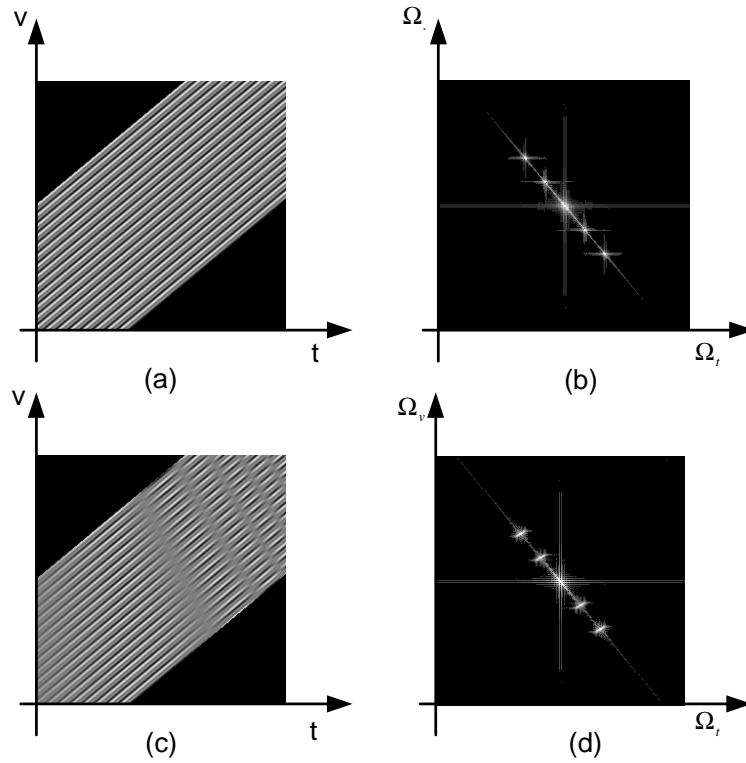

Figure 19 Optimal sampling when there are occlusions. (a) The most conservative sampling strategy. (b) The optimal sampling strategy when the occlusions are weak.

Figure 20 Epipolar image and epipolar mask. (a) A three-layer scene EPI. (b) The Fourier transform of the EPI in (a). (c) The EPM of the first layer. (d) The mulplication of the EPMs of the first and the second layers.

Figure 21 One-layer Lambertian and non-Lambertian scenes. (a) A Lambertian scene. (b) The Fourier transform of (a). (c) A non-Lambertian scene. (d) The Fourier transform of (c).
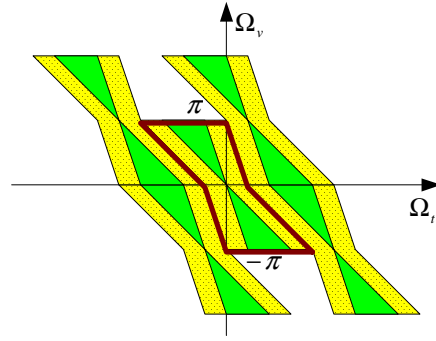
Figure 22 Optimal sampling when the scene is non-Lambertian.

(a1)   (a2)   (a3)

(b1)   (b2)   (b3)

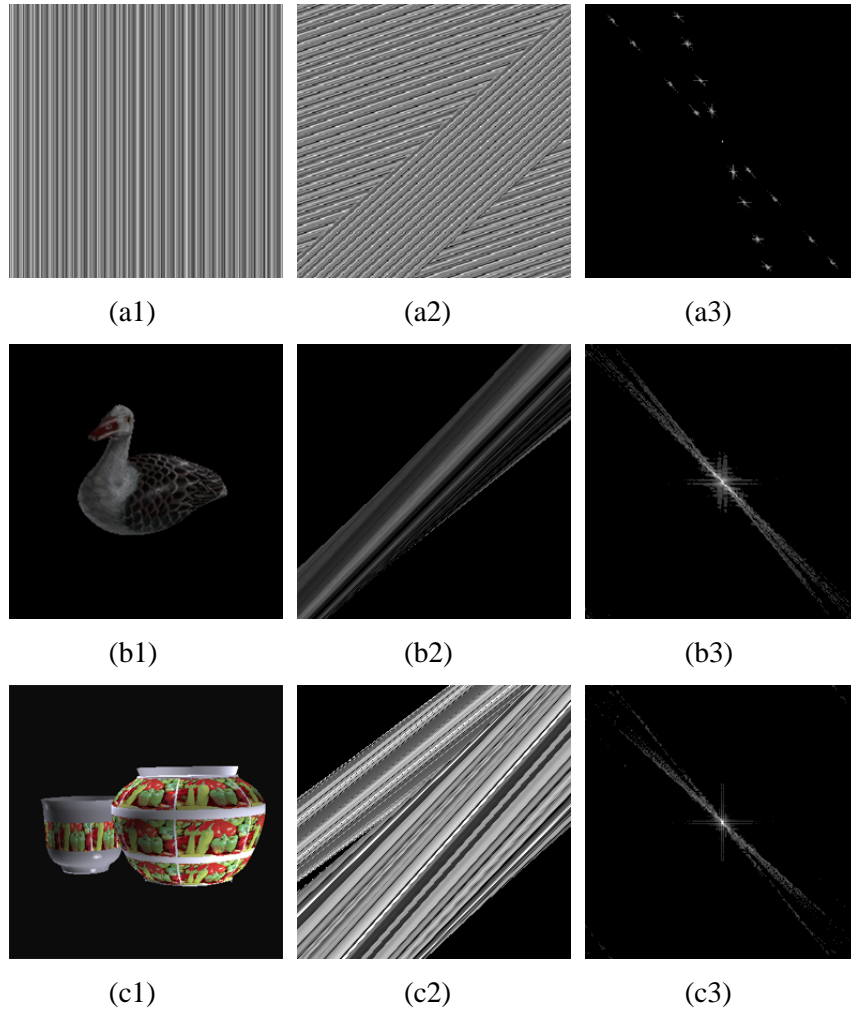(c1)   (c2)   (c3)

Figure 23 The testing scene EPIs and their Fourier transforms. (a) The synthetic scene. (b) The scene *Duck*. (c) The scene *Containers*.