CrossMark

REGULAR PAPER

Spectral analysis of image-based rendering data with scene geometry

Chang-Jian Zhu[1] · Li Yu[1]

Abstract The plenoptic function is commonly used to model image-based rendering (IBR) in terms of sampling and reconstruction. In this paper, we study the influence of scene geometry on the spectrum of a plenoptic function using spectral analysis. As the scene geometry becomes more complicated, the spectrum of the plenoptic function will broaden. Therefore, the bandwidth of the plenoptic function is closely related to the scene geometry, and the end goal of the bandwidth is to determine the minimum sampling rate (i.e., the minimum number of cameras) for IBR. Under the spectral analysis, we mathematically derive an analytical function to determine the spectral support of the plenoptic function. Based on the support of the plenoptic spectrum, the minimum number of cameras is determined to reconstruct the plenoptic function up to a certain frequency that allows a certain reconstruction quality when performing IBR. Furthermore, we derive a reconstruction filter for the plenoptic sampling for irregularly shaped scenes. Last, the experimental results are presented and compared with the theory.

Keywords Plenoptic function · Sampling · Spectral analysis · Image-based rendering · Scene geometry

✉ Chang-Jian Zhu
  changjianzhu@hust.edu.cn

[1] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China

1 Introduction

Image-based rendering (IBR) has been presented as an alternative to conventional 3D computer graphics [1–3]. Instead of representing a scene via geometric models, the scene can be represented using a set of multi-view images that compensate for the lack of geometric information as depicted in Fig. 1a. Then, arbitrary novel views can be rendered from these multi-view images as indicated in Fig. 1b. Clearly, sampling the scene using multi-view images is a fundamental step in IBR [4–6]. However, the sampling problems of the IBR remain widely unresolved in theory and practice [7, 8].

For the purposes of studying the sampling problem, we can determine the minimum number of capturing cameras needed whilst still achieving anti-aliased IBR [4–10]. The answers to the sampling problem are related to the manner in which the acquired multi-view information is described [7, 8]. Generally, a plenoptic function (POF) [11, 12] is used to describe the multi-view acquisition. More precisely, the POF describes the light intensity passing through each viewpoint, in each direction, at each moment and for each wavelength in the 3D space. Therefore, the minimum IBR sampling rate can be determined using the results of the POF sampling [13–16].

In previous studies, the authors used methods such as spectral analysis [13], reconstruction filter [23, 24], casuality requirement in scale-space theory [25], and layered depth [26–29] to study plenoptic sampling problems. The most well-known method is performing a spectral analysis of the POF [7, 8]. For example, the minimum sampling rate is determined by the spectrum of the POF [13, 14] and the bandwidth of the POF [15, 16]. In this paper, we will follow the spectral analysis method. Without the loss of generality,

Fig. 1 An illustration of the IBR. The procedures of the IBR consists of image capturing and novel view rendering. **a** A set of images is captured by some cameras along a capturing path. Diagram **b** illustrates the reconstruction of novel views from the captured images
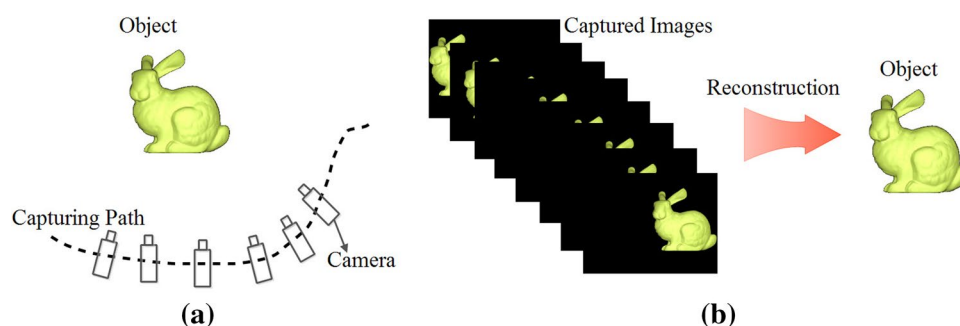
the spectral support of the POF depends on the properties and the scene depths [14]. Currently, existing techniques have studied the influence of the following factors on the plenoptic spectrum: the minimum depth and the maximum depth [13], non-Lambertian reflections and occlusions [14], whether the scene surfaces are flat [15], maximum frequency of painted signals and the angle of a slanted plane [16]. Additionally, in [15], Do et al. also analyzed the relationship between a regular curvature of the scene surface and the spectrum of the POF. Then they derived the essential bandwidth of the plenoptic spectrum based on the maximum curvature of the scene surface. Nevertheless, we determine that the scene geometry for arbitrarily irregular shapes is not completely described only by the regular curvature of the scene surface. Generally, the scene geometry for an irregular shape is unknown, its quantification is extremely complicated [14], such as the leaves of a tree, and its mathematical analysis appears futile. Therefore, compared to these prior studies, we examine the spectral properties of the POF for arbitrary irregularly shaped scenes.

Our contribution to the existing theory is a detailed study on the influence of the surface geometry shape (e.g., irregular curvatures) on the plenoptic spectrum. An extension of the study and the results of Do et al. [15] in the case of arbitrary irregular shapes can be easily derived using simple mathematics. Our analysis reveals that as the surface geometry shape becomes more complicated, the plenoptic spectrum will broaden. Based on this characterization, the plenoptic sampling theorem can be derived. By assuming that the POF is bandlimited, similar to [16], we can introduce a filter for the reconstruction of the continuous POF. The theoretical results are verified through experiments.

In addition to the fundamental interest of obtaining the sampling rate of the IBR, the results are useful in precisely characterizing the relationship between the scene and the camera pose (e.g., position and direction), e.g., what should be captured in the scene as the surface complexity of the scene becomes more complicated.

The outline of the paper is the following: Related work is introduced in Sect. 2. In Sect. 3, the POF, the SPF, and standard parameterization and notations are presented. Section 4 studies the expression of the POF spectrum on a line in space and analyzes the plenoptic spectral. Section 5 studies the sampling of the POF and a reconstruction filter. The theory presented in this paper is then verified using experiments in Sect. 6. The conclusions are drawn in Sect. 7.

2 Related work

The IBR sampling problem was mentioned for the first time in [17]. Furthermore, it was analyzed in detail by Gortler et al. [18], Emilio et al. [19], and Shum et al. [20]. The first analysis of this problem using spectral analysis was presented by Chai et al. [13]. They studied the uniform sampling of the POF using spectral analysis to determine the minimum sampling rate for IBR. By assuming a Lambertian surface without occlusions, the spectral support of a POF was approximately bounded by only the minimum and the maximum depths of a scene. The minimum sampling rate was obtained by compacting the replicas of the spectral support of the sampled light field within the smallest interval. Similarly, in [14], Zhang et al. presented a method, known as surface plenoptic function (SPF), to analyze the plenoptic spectrum. It was noted that the non-Lambertian reflections, depth variations, and occlusions could broaden the plenoptic spectrum. As previously mentioned, Do et al. [15] studied the bandwidth of the POF in detail. It was noted that unless the scene surface was flat, the continuous plenoptic spectrum was not bandlimited. Recently, and similar to the work of Do et al. [15], Gilliam et al. [16] analyzed the plenoptic spectrum of a slanted plane. They presented an exact closed-form expression of the plenoptic spectrum and derived a new Nyquist sampling rate of the POF. Similar to the spectral analysis of the POF, a unified mathematical framework [21] was presented. This framework consisted of three different parameterizations (e.g., plane-sphere, two-plane, and in-camera) to represent the light field. The mathematical framework was applied to model the image formation process in the frequency domain. Similar to the POF, a technique known as plenacoustic function [22] was presented by Ajdler et al. to study the spatialization of the sound field. Then, the

Table 1 Notations and definitions

t, ξ	The camera position on the camera plane
u, v	The pixel location on the image plane
θ	The angle between a light ray and the principal axis of camera
f	The focal length of camera
s	Curvilinear surface coordinate
S	The scene surface function
$\Psi(s)$	The function of x with respect to s
$p(t, v)$	The function of the EPI
v_m	The maximum of a FOV
Δt	The spacing between the cameras along the spatial axis
Δv	The spacing between the pixels on the image plane
ω_s	The maximum frequency of a texture signal
Ω_t	The indexing of the frequency along the t-axis
Ω_v	The indexing of the frequency along the v-axis
X	The maximum of x
T	The maximum of s
$l(s, \theta)$	The intensity of a light ray emitted from s on the scene surface at direction θ

spectral analysis was used to analyze the sampling of the plenacoustic function. Furthermore, reconstruction aspects such as the reconstruction filter were discussed by Stewart et al. [23]. A linear, spatially invariant reconstruction filter was presented that reduced the ghosting artifacts of under-sampled light fields. Using this reconstruction filter, the number of necessary camera-plane samples for light field rendering can be significantly reduced while avoiding aliasing in the light field reconstruction. Similarly, Hoshino et al. [24] studied multi-viewpoint image acquisition using an appropriate prefilter to reduce the aliasing of the reconstituted images.

The above methods were based on the spectral analysis point of view. There were also other methods used to study the plenoptic sampling. For example, Lin et al. [25] presented a method known as the lower bound of the number of samples required in light field rendering [17, 18]. Their analysis was based on the casuality requirement in scale-space theory. Furthermore, a criterion was defined to determine the lower bound of the number of samples to avoid horizontal and vertical double images. Using this criterion, closed-form solutions of lower bounds were obtained. Additionally, recent techniques have shown interesting results when using the depth information to analyze the rendering error of the IBR [26]. In [27], Shade et al. presented a new method which was known as layered depth images. The results of the depth information for a complicated scene were obtained using different layered depth. This method was also presented in [28]. Using the layer-based method, a trade off between the amount of geometric information and the number of images necessary can be determined. Similarly, certain results used the plenoptic layer-based modeling [29] to analyze how to reduce

the number of images necessary for alias-free rendering. Recently, in [30], Shidanshidi et al. extended the concept of effective sampling density (ESD) [31] to study the acquisition and rendering of the IBR. It was noted that the ESD was an effective indicator determined by system parameters and could be used to directly estimate the output video distortion without access to the ground truth.

Finally, some techniques have shown interesting parallel cameras that were not restricted to lie on a plane corresponding to a camera array [32]. The theory was adapted to consider unstructured camera systems for analyzing the plenoptic sampling for the cameras can reside at arbitrary locations and orientations. Similarly, dynamically reparameterized light fields [33] and unstructured lumigraph rendering [34] have also been presented. Recently, a spherical 4D light field sampling scheme [35] was presented. Using this sampling scheme, a spherical light field without aliasing can be reconstructed.

3 Standard parameterization and notation

For clarity, we have listed the important notations and associated definitions used throughout the paper in Table 1. We begin by briefly reviewing the properties of light ray representation. In this paper, we will follow the notations used in [14–16, 36, 37]. Using a pin-hole camera model, the POF introduced by Adelson and Bergen [11] is applied to represent a light ray. As previously mentioned, the POF is given by a seven-dimensional function. It should be noted that it is extremely complicated to address the seven dimensions of the POF. Several assumptions can be made to reduce the dimensionality [10]. For

Fig. 2 Diagram of the 4D light field and the SPF [15]. A light ray can be represented using a camera plane $t\xi$ and an image plane uv. The separation between the camera plane and the image plane is equivalent to the focal length f. The length of the object is T. Here, a light ray can also be represented using the SPF. The position of the light ray on an object plane is (s, r), and its direction is (θ_s, θ_r)

Fig. 4 An illustration of the 2D light field $p(t, v)$, in which the intensity of a light ray is defined by its captured position of the camera at t and viewing direction v on the image plane. Furthermore, $p(t, v)$ describes the relationship between the camera and the scene geometry $z(x)$

example, Gortler et al. [18] presented a 4D light field to represent each light ray in a static scene. Each light ray is parameterized by two intersections with two parallel planes. One of the planes is denoted as the camera plane, and its coordinate is (t, ξ). Another plane is denoted as the image plane, and its coordinate is (u, v). The separation of the two parallel planes is equivalent to the focal length f, as indicated in Fig. 2. Based on the above results of the light field, we can use a function $p(t, \xi, u, v)$ to represent the intensity of the light ray at a camera location (t, ξ) and a pixel location (u, v).

For simplicity of exposition, and similar to [14–16], we consider a 2D version of the POF, $p(t, v)$, by fixing ξ and u. Then, Fig. 2 can be simplified as Fig. 3a. A visual representation of the 2D light field can be constructed if we plot its intensity at all possible (t, v) coordinates. This visual representation is known as the epipolar-plane image (EPI) [38]. Next, we define how the camera and image relate to a light ray leaving the surface using the EPI.

Definition 1 Given a camera position t, image plane coordinate v, and scene depth $z(x)$ with respect to x, as indicated

in Fig. 4, the mapping between the scene and the EPI along the x-axis can be expressed as follows:

$$v = f \tan(\theta), \tag{1}$$

$$t = x - z(x)\frac{v}{f}, \tag{2}$$

where θ is the angle between a light ray and the principal axis of a camera.

Additionally, the above form for light ray parametrization can be represented by the SPF [14–16]. A light ray with the SPF is parameterized by the point of the origin on the scene surface S. Generally, we use two curvilinear surface coordinates (s, r) to parameterize the scene surface as depicted in Fig. 2 [16]. We define the variables $[x, y, z]^T$ to represent a point's Cartesian coordinates [26]. Then, using a function $S(s, r) = [x(s, r), y(s, r), z(s, r)]^T$, the location of a point on the scene surface can be defined. Additionally, the direction of the light ray is defined by

Fig. 3 **a** Choice of camera spacing on the camera plane t and the image plane v for the 2D light field, where Δt is the spacing between the cameras and Δv is the spacing between the pixels. **b** A light ray representation using the SPF, where the position of the light ray is s, and its direction is θ_s

(a)

(b)

the viewing angle (θ_s, θ_r), where θ_s and θ_r are measured relative to the z-axis. Therefore, we can use the function $l(s, r, \theta_s, \theta_r)$ to represent the intensity of the light ray emitted from a point (s, r) on the scene surface at a viewing direction (θ_s, θ_r).

Additionally, similar to the 2D version of the POF $p(t, v)$, we can simplify $l(s, r, \theta_s, \theta_r)$ into $l(s, \theta_s)$ by fixing r and θ_r. Now, we use a curvilinear surface coordinate s to parameterize the scene surface in the 2D space as shown in Fig. 3b. We define the parametrization of scene surface with the geometry shape as follows:

Definition 2 The location of a point on the scene surface can be described using the function S as follows:

$$S = \begin{cases} z(x) = f_s(x), & x \in [0, X] \\ x(s) = \Psi(s), & s \in [0, T], \end{cases} \tag{3}$$

where $\Psi(s)$ is the function of x with respect to s. X is the maximum of x, and T is the maximum of s.

It should be noted that the surface geometry shape can be approximately represented using $\Psi(s)$. Based on the actual situation, we can adjust the function $\Psi(s)$ to approximately obtain the expression of the surface geometry shape. Therefore, we can use the function $l(s, \theta)$ to represent the intensity of the light ray emitted from a point s on the scene surface at a viewing direction θ (let $\theta = \theta_s$). Then, by substituting (3) to (1) and (2) a new expression for the mapping between t, s and θ can be obtained as follows:

$$t = \Psi(s) - z(\Psi(s))\frac{v}{f} = \Psi(s) - z(\Psi(s))\tan(\theta). \tag{4}$$

Finally, we can obtain the expression using $l(s, \theta)$ for a light ray as follows:

$$l(s, \theta) = p((\Psi(s) - z(\Psi(s))\tan(\theta)), \quad f\tan(\theta)). \tag{5}$$

Additionally, a texture signal is represented by $g(s)$. The maximum frequency of the texture signal is denoted by ω_s. Let Ω_t be the indexing of the frequency along the camera position's t-axis, and let Ω_v be the indexing of the frequency along the image pixel's v-axis.

4 Spectral analysis of the POF

In this section, an analytical expression of the 2D Fourier transform (2D-FT) of the POF is given. The relationship between the scene geometry and the spectral support of the POF is analyzed.

4.1 Spectrum parametrization of the POF

We analyze the spectrum for the uniform sampling of the POF along the spatial axis. Firstly, we assume that the scene surface does not contain occlusions. In this case, the light rays within the field-of-view of cameras can intersect with at the scene object surface z(x). By [15–17], the scene object surface z(x) is constrained such that

$$\left| z'(x) \right| < \frac{f}{v_m}, \tag{6}$$

where v_m is the maximum of a field-of-view (FOV) for a camera with the FOV in reference to [16], and $z'(x)$ is the derivative of the scene depth z with respect to the coordinate x.

After that, we state the spectral characteristics of the POF in the following theorem:

Theorem 1 *The spectral characteristics of the POF in reference to [15] and [16] can be described as follows:*

$$P(\Omega_t, \Omega_v) = \begin{cases} \int_{-\infty}^{\infty} H(s, \Omega_v f - \Omega_t z(\Psi(s)))l(s)e^{-j\Omega_t\Psi(s)}ds, \\ \quad \text{when} \quad \Omega_t = \Omega_v \frac{f}{z(\Psi(s))}, \quad \Omega_t = 0. \\ 0, \quad \text{when} \quad \Omega_t \neq \Omega_v \cdot \frac{f}{z(\Psi(s))}, \quad \Omega_t \neq 0. \end{cases} \tag{7}$$

where $H(s, \Omega_v f - \Omega_t z(\Psi(s)))$ is an intermediate variable. Its expression is given by

$$H(s, \Omega_A) = 2\pi f\Psi'(s)\delta(\Omega_A) + j2\pi f \cdot z'(\Psi(s))\Psi'(s)\frac{d}{d\Omega_A}\delta(\Omega_A), \tag{8}$$

where $\Omega_A = \Omega_v f - \Omega_t z(\Psi(s))$, $z'(\Psi(s))$ is the derivative of z with respect to $\Psi(s)$, $\Psi'(s)$ is the derivative of Ψ with respect to s, $z(\Psi(s))$ is the scene depth with respect to s, and $\delta(\cdot)$ is a Dirac function.

Proof The full proof can be found in Appendix A. A brief outline of the proof is as follows. By taking the Fourier transform of the EPI function $p(t, v)$, we can define the expression of the plenoptic spectrum as follows:

$$P(\Omega_t, \Omega_v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t, v)e^{-j(\Omega_v v + \Omega_t t)}dtdv. \tag{9}$$

By substituting (5) with (9), a new expression of the plenoptic spectrum with s and θ can be obtained as follows:

$$P(\Omega_t, \Omega_v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f \cdot \frac{l(s, \theta) \cdot \Lambda}{\cos^2(\theta)} \cdot e^{-j(\Omega_v f\tan(\theta) + \Omega_t(\Psi(s) - z(\Psi(s))\tan(\theta)))}d\theta ds, \tag{10}$$

where $\Lambda = \Psi'(s) - z'(\Psi(s))\Psi'(s)\tan(\theta)$. In the extreme case, the scene surface is often assumed to be Lambertian [13–16] which signifies that $l(s,\theta) = l(s)$ for all θ. Thus, the integration variables in (10) are changed, and the following expression can then be obtained as

$$P(\Omega_t, \Omega_v) = \int_{-\infty}^{\infty} H(s, \Omega_A) l(s) e^{-j\Omega_t \Psi(s)} ds. \tag{11}$$

Using (11), the spectrum of the POF is nonzero if $\Omega_A = \Omega_v f - \Omega_t z(\Psi(s)) = 0$. Then, we can obtain the spectral characteristics of the POF, as provided in (7). □

It is apparent that the spectrum of the POF is related to the $z(\Psi(s))$. It should be noted that $z(\Psi(s))$ depends on the scene depth and the scene surface geometry. Furthermore, the spectral support consists of two regions, which are $\Omega_t = \Omega_v \frac{f}{z(\Psi(s))}$ and $\Omega_t = 0$. Using the spectral characteristics, the sampling rate for the plenoptic sampling can be approximately obtained, such as [13, 16].

4.2 Plenoptic spectral analysis with the scene geometry

4.2.1 Plenoptic spectrum

We attempt to analyze the spectrum estimation of the POF with the scene geometry. As previously mentioned, for a Lambertian scene, we can assume that $l(s,\theta)$ is a bandlimited function in the variable θ. Thus, the scene geometry can be described using scene texture information. The spectrum of the POF with respect to the scene texture information can be given by the following theorem.

Theorem 2 *The spectrum of the POF with a sinusoid pasted as the texture, e.g.,$l(s) = \cos(\omega_s s)$ can be calculated as follows:*

$$
\begin{aligned}
P(\Omega_v, \Omega_t) = {} & \frac{2\pi f}{|\Omega_t|} \cos\left(\omega_s \Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right) \\
& \times \Psi'\left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right) \\
& \times \left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)' \cdot e^{-j\Omega_t z^{-1}(\Omega_v f/\Omega_t)} \\
& + \frac{j2\pi f}{(\Omega_t)^2}\left(\cos\left(\omega_s \Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)\right. \\
& \times \Psi'\left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right) \cdot \left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)' \\
& \times z'\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right) \cdot \left. e^{-j\Omega_t z^{-1}(\Omega_v f/\Omega_t)}\right)',
\end{aligned}
\tag{12}
$$

where $z^{-1}(\cdot)$ is the inverse function of $z(\cdot)$, and $\Psi^{-1}(\cdot)$ is the inverse function of $\Psi(\cdot)$.

Proof The full proof can be found in Appendix B. A brief outline of the proof has been given as follows: Using (11), if we have a sinusoid pasted as the texture, e.g.,
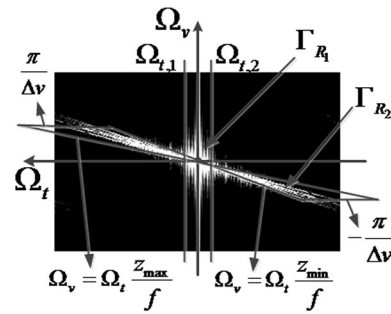

Fig. 5 Diagram of the spectral support of the plenoptic spectrum with a texture signal, which only illustrates the primary characteristics of the superimposed spectrum

$l(s) = \cos(\omega_s s)$, then the spectral expression of the POF with respect to the scene texture is given as

$$(\Omega_t, \Omega_v) = \int_{-\infty}^{\infty} H(s, \Omega_A)\cos(\omega_s s) e^{-j\Omega_t \Psi(s)} ds. \tag{13}$$

Based on the properties of the Dirac function $\delta(\cdot)$ for its integral, the derivation results of (13) are illustrated in (12). □

We can see from Theorem 2 that $P(\Omega_t, \Omega_v)$ is determined by the scene depth $z(x)$, texture information ω_s, and scene surface geometry $\Psi(s)$.

4.2.2 Spectral support of the plenoptic spectrum

Using Theorem 2, we can model the spectral support with the scene geometry. This model is illustrated in Fig. 5. The spectral support consists of two regions, Γ_{R_1} and Γ_{R_2}, which contain the majority of the spectral energy. First, the expression of the relationship between Ω_t and Ω_v based on (7) can be given as follows:

$$\Omega_v = \frac{\Omega_t z(\Psi(s))}{f}. \tag{14}$$

After that, we derive the expressions of the two regions for the spectral support. The first region is defined by four lines. It consists of two diagonal lines and two parallel lines. The two diagonal lines are related to the maximum and the minimum depths of the scene, respectively, i.e.,

$$\Omega_{v,1} = \frac{\Omega_t z_{\min}}{f} \quad \text{and} \quad \Omega_{v,2} = \frac{\Omega_t z_{\max}}{f}. \tag{15}$$

The two parallel lines depend on the camera resolution Δv, as shown in Fig. 1a, where Δv is a fixed parameter of the acquisition device used to obtain the images. Here, we derive the expression of the two parallel lines for the first region as follows:

$$\Omega_{v,3} = -\frac{\pi}{\Delta v} \quad \text{and} \quad \Omega_{v,4} = \frac{\pi}{\Delta v}. \tag{16}$$

Fig. 6 Rectangular sampling grid for the plenoptic sampling on a plane. **a** Placement of the cameras on the plane. The cameras are uniformly placed on the plane. **b** Rectangular sampling grid for plenoptic sampling. The rectangles represent cameras. The variable Δt is the camera spacing between the cameras in the t-axis, and $\Delta \xi$ is the camera spacing in the ξ-axis

(a)   **(b)**
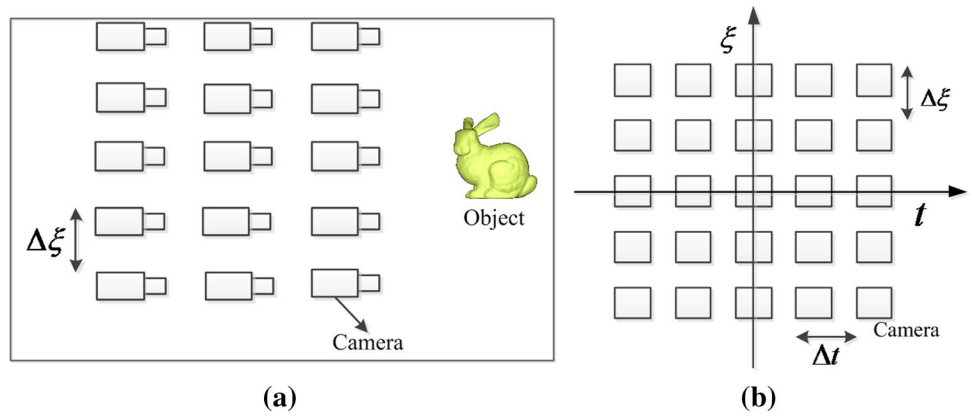
The second region is defined by two parallel lines. Combining the results with the expressions of the two lines in [14–16] and considering the characteristics of (12), the expressions of the two parallel lines can be defined as follows:

$$\Omega_{t,1} = \max \left(\left| \frac{1}{\Psi'(s) + z'(\Psi(s))\Psi'(s)\max(\theta)/f} \right| \right) \cdot \omega_s, \quad \text{and}$$
$$\Omega_{t,2} = \max \left(\left| \frac{1}{\Psi'(s) - z'(\Psi(s))\Psi'(s)\max(\theta)/f} \right| \right) \cdot \omega_s, \quad (17)$$

where $\max \left(\left| \frac{1}{\Psi'(s) + z'(\Psi(s))\Psi'(s)\max(\theta)/f} \right| \right)$ and

$\max \left(\left| \frac{1}{\Psi'(s) - z'(\Psi(s))\Psi'(s)\max(\theta)/f} \right| \right)$ denote the maximum

complexity of the scene. It should be noted that the expressions of the six lines are closely related to the scene geometry $z(\Psi(s))$. Additionally, they depend on the focal length f and the frequency of the texture signal ω_s. From this conclusion, we can determine that the scene geometry also influences the POF spectrum.

5 Sampling and reconstruction

In the previous sections, the influence of the scene geometry on the POF spectrum has been studied along both the image plane and the camera plane for the 2D light field. In should be noted that the spectrum of the POF lies on a support that broadens as the scene geometry becomes more complicated, and it is nearly bandlimited. This result verifies that the scene surfaces are flat and Lambertian [15]. Because the IBR uses many images and does not require any geometrical information, the number and the spacing between the cameras are key factors for the quality of the reconstruction for the IBR [13, 16]. On the other hand, along the camera plane direction, the plenoptic signal can be filtered to avoid aliasing.

In this section, the sampling theorem of the POF is presented. The frequency estimation for a plenoptic signal is studied when sampling the 2D light field. Furthermore, a reconstruction filter is discussed to reconstruct
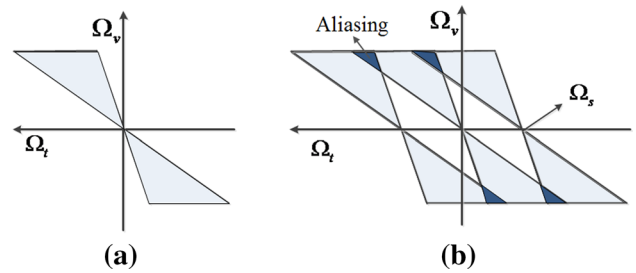

(a)   **(b)**

Fig. 7 Spectrum of the POF [13]. **a** *Top view* of the plenoptic spectrum and **b** *top view* of the plenoptic spectrum with its repetitions due to spatial sampling

the continuous plenoptic signal from the available image samples.

5.1 Frequency estimation of the plenoptic signal

To uniformly sample the POF along the camera plane, finite scene widths are considered. Now, let us consider the spectrum of the POF given by (12) at the particular frequencies Ω_t and Ω_v. Its approximate shape is given in Fig. 5. Using (17), the frequency estimation for Ω_t can be written as

$$\Omega_t \approx \max \left(\left| \frac{1}{\Psi'(s) - z'(\Psi(s))\Psi'(s)\max(\theta)/f} \right| \right) \cdot k_t \omega_s, \quad (18)$$

where k_t is a scaling factor. From the above expression, Ω_t is determined by the scene geometry (i.e., $z'(\cdot)$ and $\Psi'(s)$), as well as the texture information, focal length, and FOV of the camera.

With regard to Ω_v for the frequency along the v-axis, it is related to the camera resolution and the rendering resolution. However, we assume that the rendering resolution is equivalent to the camera resolution. Substituting (18) to (14) to obtain the follow expression for Ω_v

$$\Omega_v \approx \max \left(\left| \frac{1}{\Psi'(s) - z'(\Psi(s))\Psi'(s)\max(\theta)/f} \right| \right) \cdot k_t \omega_s z(\Psi(s)). \quad (19)$$

Fig. 8 Essential bandwidth of the POF. **a** Parametric essential bandwidth on the plenoptic spectrum. **b** Cut of the POF spectrum for a particular Ω_v frequency, when $\Omega_v = \frac{\Omega_t z_{\max}}{f}$. **c** Cut of the POF spectrum for Ω_t, when $\Omega_t = \max \left(\left| \frac{1}{\Psi'(s) - z'(\Psi(s))\Psi'(s)\frac{\max(\theta)}{f}} \right| \right) \cdot \omega_s$
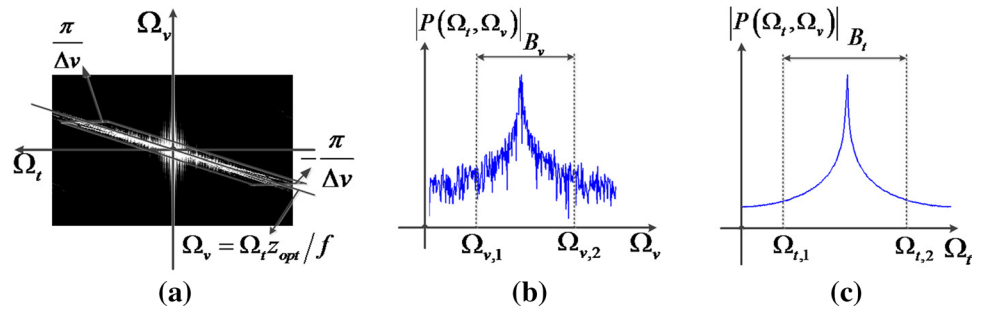
Using (18) and (19), we can determine the frequency of the POF to analyze the plenoptic sampling.

5.2 Sampling theorem of the POF

In this sub-section, the effect of the sampling on the camera plane is represented in the 2D light field. This representation considers the results of the previous sections. As depicted in Fig. 6, a set of cameras are uniformly placed on the camera plane (t, ξ). However, for simplicity, we now discuss the sampling of the POF only along the t-axis. In this paper, the sampling theorem of the POF primarily determines the number and the spacing between the cameras. Thus, we can obtain the optimal balance between the number of cameras and the quality of the reconstructed images from the image samples. Concurrently, the sampling of the POF is shown in detail for a rectangular sampling pattern [13].

The schematic top view of the spectrum of the POF is illustrated in Fig. 7a. When sampling the POF along the t-axis with a spatial sampling frequency of Ω_s, repetitions of the spectrum aliasing occur, as depicted in Fig. 7b [13]. According to Shannon's uniform-sampling theorem [39], the relationship between the sampling angular frequency Ω_s and the camera spacing Δt is $\Omega_s = 2\pi / \Delta t$. The spatial sampling frequency needed to reconstruct the POF up to a temporal frequency Ω_t is given by $\Omega_s \geq 2\Omega_t$. Nevertheless, it does not consider the shape of the plenoptic spectrum and the corresponding slanted reconstruction filter. Based on [16], to avoid aliasing in the reconstruction of the POF, the camera spacing Δt along the spatial axis can be given as follows:

$$\Delta t = \frac{2\pi z_a}{\Omega_t (z_{\max} - z_{\min})}, \tag{20}$$

where $z_{\min}$ is the minimum depth of the scene, $z_{\max}$ is the maximum depth of the scene, and z_a is the average depth. Then, by substituting (18) into (20), the camera spacing for plenoptic sampling can be determined using as following expression:

$$\Delta t = \frac{2\pi z_a}{\max \left(\left| \frac{1}{\Psi'(s) - z'(\Psi(s))\Psi'(s)\max(\theta)/f} \right| \right) \cdot (z_{\max} - z_{\min}) \cdot k_t \omega_s}. \tag{21}$$

In this expression, $z(\cdot)$ denotes the depth variation. The scene geometry complexity for the surfacecoce geometry shape (e.g., curvatures) is represented using $\Psi(\cdot)$.

5.3 Reconstruction of the POF

From (7), it can be observed that the plenoptic spectrum is band-unlimited. However, a large fraction energy of the plenoptic signal consists of two regions, $[\Omega_{v,1}, \Omega_{v,2}]$, $[\Omega_{v,3}, \Omega_{v,4}]$, and $[\Omega_{t,1}, \Omega_{t,2}]$ as indicated in Fig. 8. We can then assume that the plenoptic spectrum is approximately bandlimited to its essential bandwidth. Here, the needed essential bandwidth consists of the area of the two regions in Fig. 8a. Because the discrete light field at each point of the sampling grid is known, the usual typical interpolation techniques [34, 40, 41] are applied to reconstruct the continuous light field at any location of the camera plane. The interpolation filter depends on the sampling grid and can be separated for the camera plane and the image plane. When the samples have been obtained by rectangular sampling, the interpolation filter can be written as a skewed filter with support as follows:

$$\left\{ \Omega_t, \Omega_v : \Omega_t \in \left[\frac{\Omega_{t,1}}{\eta_s} - \frac{\pi}{\Delta t}, \frac{\Omega_{t,2}}{\eta_s} + \frac{\pi}{\Delta t} \right], \Omega_v \in [\Omega_{v,1}, \Omega_{v,2}] \right\}. \tag{22}$$

The support of the filter is depicted by the bold lines in Fig. 8a. The two parallel lines for $\Omega_{v,1}$ and $\Omega_{v,2}$ are related to the camera resolution Δv, as indicated in Fig. 3a. Moreover, the reconstruction filter is created by setting a skewed filter [15, 16]. Here, the expression of the skew in reference to [16] can be shown as follows:

$$\eta_s = \frac{z_{\min} + z_{\max}}{2f}. \tag{23}$$

Finally, the z_{opt} in Fig. 8a is the optimal constant depth of the scene, and it is calculated using (23).

Based on the above analysis, we can see that the sampling rate for plenoptic sampling and the reconstruction filter design primarily depend on the frequency of the plenoptic signal. Therefore, the essential bandwidth is a key factor for the plenoptic sampling.
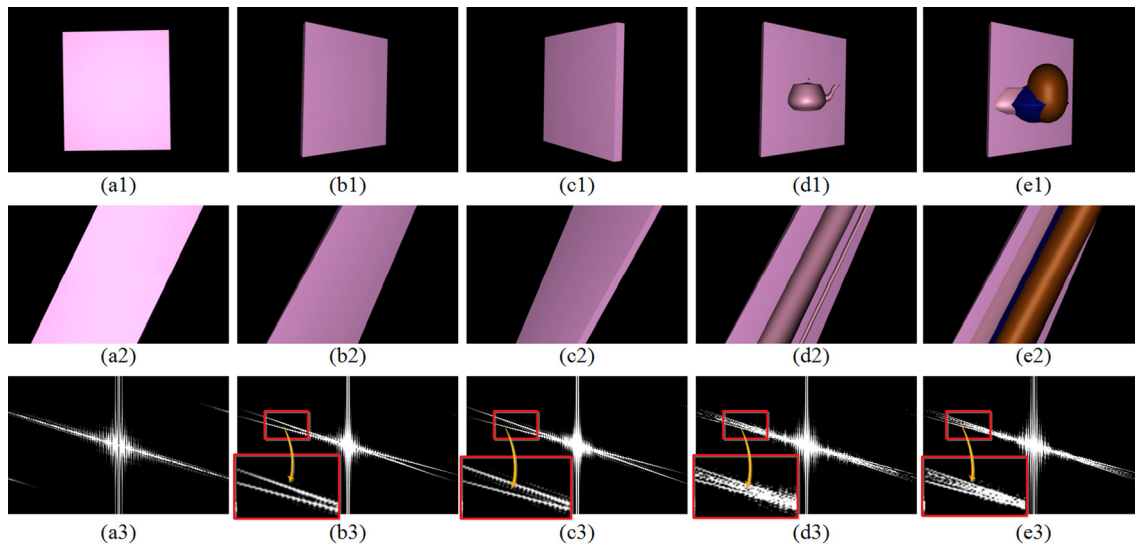
Fig. 9 Spectral supports of the POF with respect to the scene geometrical variations. Diagrams (**a1**)–(**e1**) depict walls with different surface complexities. Diagram (**a1**) depicts a constant depth wall, which is named the first wall; Diagram (**b1**) depicts the second wall, in which the slant angle between the plane of the wall and the x-axis is $\frac{\pi}{4}$ rad; Diagram (**c1**) depicts the third wall, with a slant angle is $\frac{3\pi}{4}$ rad; Diagram (**d1**) depicts the fourth wall; and Diagram (**e1**) depicts the fifth wall. The minimum and maximum depths of these walls are the same. Diagrams (**a2**)–(**e2**) depict the EPI-volumes for the corresponding walls. Diagrams (**a3**)–(**e3**) illustrate the spectrums for different walls, which are calculated using the corresponding EPI-volumes

6 Experimental results

In this section, experimental results are presented for the spectral supports of the POF and the reconstruction results of a plenoptic signal. It should be noted that the interpolation filter (IPF) described by (22) is applied to perform the reconstruction. The reconstruction results are then compared to a reconstruction result presented by Stewart et al. [23].

6.1 Spectral support of the POF

6.1.1 Synthetic scenes

The spectral supports of the POF are simulated using five different wall models. In the simulations, the five walls are all rendered using 3Ds Max. For simulation, the characteristics of the plenoptic spectrum with respect to the scene geometrical variations can be revealed. More precisely, with the exception of Fig. 9(a1), which has no depth variation, the minimum and maximum depths of these scenes are the same. However, the surface complexity of the scenes varies as indicated in Fig. 9(a1)–(e1). Additionally, it should be noted that the size is the same for all of the walls. Here, the width of the wall is 127.0 cm, and the height is 186.0 cm. In the presented simulations, all of the walls are fixed at the point $(0, -40, 400)$ cm. Additionally, the rectangular sampling [13] of the scenes is considered. More precisely, each wall is captured by a set of cameras. The

cameras are uniformly placed along a line $L = \{(x, 0, 0)\}$. Here, $x \in [-100.0, 100.0]$cm and the images are captured every 1.0 cm along the straight line. The camera resolution is 240×320 pixels. Finally, from the captured images, each EPI-volume can be obtained, and the Fourier transform operation of the POF is performed using (9). For this analysis, each EPI-volume consists of 200 captured images.

From the simulations, the EPI-volumes and the corresponding spectrums are depicted in Fig. 9. It can be observed in Fig. 9(a2) that the EPI-volume consists of a set of parallel lines because the slope of the EPI depends on the scene depth. The depths of the first wall are constant; thus the slopes of the EPIs for the first wall are the same. Furthermore, as the depths of the wall change, the slopes of the EPIs also change as illustrated in Fig. 9(b2)–(e2). For example, the EPI-volume in Fig. 9(b2) consists of a set of slanted lines, and their slopes are different. It should be noted that the slope of the EPI increases as the depth increases. On the other hand, the EPI-volumes in Fig. 9(b2) and (c2) are different, even though the minimum and maximum depths are the same for the second wall and the third wall. This phenomenon occurs because the slant angles of the two walls are different. Furthermore, there are certain differences between the EPI-volumes of the second wall and that of the fourth wall. The reason is that the surface complexity between the two walls is different. Additionally, the EPI-volume for the fourth wall is different compared to that of the fifth wall. Intuitively, the EPI-volumes will be more complicated as the surface complexity becomes more
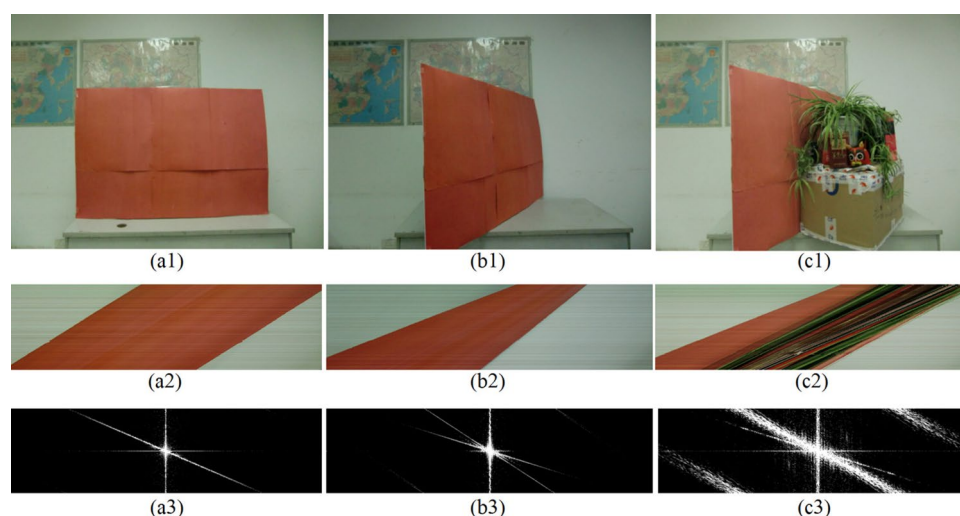
Fig. 10 Example of the spectral support of the POF from actual scenes. The first plane is depicted in (**a1**), in which the surface plane is parallel to the camera plane, and the distance between the plane and the camera plane is 134.0 cm. Its EPI-volume is depicted in (**a2**), and the spectral support is illustrated in (**a3**). Diagram (**b1**) depicts the second plane, which has a slant angle of $\frac{\pi}{4}$ rad. The minimum depth is 91.0 cm, and the maximum depth is 188.0 cm. Its EPI-volume is depicted in (**a2**), and the spectral support is illustrated in (**a3**). Diagram (**c1**) depicts the third plane and a miniascape, and its parameters are the same as that of the second plane. Its EPI-volume is illustrated in (**c2**), and the spectral support is depicted in (**c3**)

complicated as depicted in Fig. 9(d2) and (e2). From the above analysis, we can find that the EPIs are related to the depth, slant angle of the wall, and surface complexity of the wall.

The spectrums of the corresponding walls are illustrated in Fig. 9(a3)–(e3). It can be observed in Fig. 9(a3) that the spectrum support of the first wall consists of two regions. Intuitively, the two regions are comprised of two intersecting lines. The two regions are the same as the spectral support in Fig. 5. The first region is a vertical molding, and it can be represented by Γ_{R_1}. Additionally, it depends on the characteristics of the scene surface. Furthermore, the second region depends on the minimum and the maximum depths of the wall as well as the camera resolution and the scene complexity. Because the first wall is parallel to the camera line, i.e., no depth variation, the second region only comprised of a diagonal line (i.e., Γ_{R_2} in Fig. 5). Nevertheless, the spectrum will be broadened if the depths change. In particular, the second region of the spectral support is split into two diagonal lines. Note that the two diagonal lines depend on the minimum and the maximum depths of the wall. For example, when the scene is a slant wall as indicated in Fig. 9(b1), the corresponding spectrum consists of three intersecting lines (i.e., the second region is split into two diagonal lines) as depicted in Fig. 9(b3). The reason for this phenomenon is that the scene depth varies. Similarly, the spectrum in Fig. 9(c3) also consists of three intersecting lines. Although the slant angle is different between the second wall and the third wall in Fig. 9(b1) and (c1), their spectrum supports are

the same as illustrated in Fig. 9(b3) and (c3), respectively. This result occurs because the minimum and the maximum depths are the same between the two walls. Additionally, an interesting phenomenon is that the area between the two diagonal lines of the spectrum in Fig. 9(b3) and (c3) is not padded. However, the area between the two diagonal lines of the spectrum in Fig. 9(d3) and (e3) is padded. It can be observed that these spectra are more complicated and wider than those in Fig. 9(b3) and (c3). Additionally, the spectrums in Fig. 9(d3) and (e3) are broadened compared to those in Fig. 9(b3) and (c3). This phenomenon is caused by the surface complexity of the walls in Fig. 9(d1) and (e1) which are more complicated than the second wall and the third wall in Fig. 9(b3) and (c3), respectively. Therefore, the spectral support will be larger as the scene complexity becomes more complicated.

The above example provides a relatively good idea of the POF spectrum and how it is influenced by the scene geometry. We can consider that the spectrum of the POF will be broadened as the surface complexity becomes more complicated. Based on the characteristics of the example, we can determine the relationship between the scene geometry and spectrum to derive the sampling theory of the POF.

6.1.2 Actual scenes

To study the spectral support of the POF with the scene geometry variations for actual scenes, three different scenes are captured by a set of cameras on a line. For simplicity, the actual scenes are three different planes. More
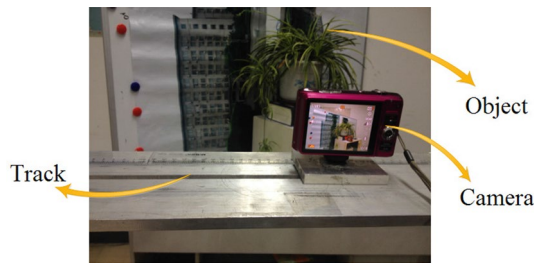
Fig. 11 Data acquisition set-up. The captured object is a blackboard and a miniascape. The camera is uniformly moved in increments of 0.5 cm along a track, and the scene is captured

precisely, the color of the first plane is single red and has no depth variation, as indicated in Fig. 10(a1). The color of the second plane is also single red, and it is a slant plane, as depicted in Fig. 10(b1). The third plane consists of a slant plane and a miniascape, as indicated in Fig. 10(c1). The size is the same for all of the planes, and only the surface geometry is different. Here, the width of the plane is 122.0 cm, and the height is 82.0 cm. Moreover, the minimum and the maximum depths of the second and the third planes are the same, respectively. In the actual experiments, a camera is uniformly moved in increments of 0.5 cm to capture each plane, as indicated in Fig. 11. For each plane, 175 images are captured at different positions along a line. For this analysis, the EPI-volume consists of 175 captured images. This parameter is randomly selected and does not affect the experimental results. Last, the camera resolution is 480×640 pixels.

Similar to the EPI-volume in Fig. 9(a2) for the simulations, Fig. 10(a2) indicates that the EPI-volume consists of a set of parallel lines, which occurs because the depths of the first plane have no variations. Additionally, the EPI-volume in Fig. 10(b2) and 10(c2) consists of a set of tilted lines. The slopes of the tilted lines are different. Similar to Fig. 9(b2)–(e2), the slope of the EPI increases as the depth increases. Another interesting phenomenon is that the complexity of the EPI-volume in Fig. 10(c2) is more complicated than that of the EPI-volume in Fig. 10(b2). This phenomenon is caused by the fact that the third plane is more complicated than the second plane. Therefore, the EPI-volume will be more complicated as the object surface becomes more complicated.

For the spectrums of these planes, similar to the spectral support presented in Fig. 9(a3) for the simulations, Fig. 10(a3) demonstrates that the spectral support also consists of two intersection lines. When the depths of the plane change, the spectrum will be broadened. More precisely, the diagonal line to the first region of the spectrum in Fig. 10(a3) is split into two diagonal lines, as indicated in Fig. 10(b3). It can be observed in Fig. 10(b3) that the spectral support consists of three intersection lines. The

two diagonal lines depend on the minimum depth and the maximum depth. Furthermore, similar to Fig. 9(b3) and (c3), the area between the two diagonal lines in Fig. 10(b3) is not padded. However, the area between the spectrum of the minimum depth and the maximum depth for the third plane in Fig. 10(c3) is padded. We can see that the spectrum is more complicated and wider than the spectrum of the second plane in Fig. 10(b3). The reason for this phenomenon is shown by the surface complexity of the third plane in Fig. 10(c1), which is more complicated than the second plane in Fig. 10(b1). Therefore, the results from the actual experiment are the same as the simulations in Sect. 6.1.1.

6.2 Reconstructions using different filters

To evaluate the performance of the new proposed reconstruction filter, we reconstruct a set of virtual views using different synthetic scenes. Additionally, the virtual views are rendered by interpolating nearby images [34, 42]. In the presented experiments, the synthetic scenes for the second wall in Fig. 9(b1), the fifth wall in Fig. 9(e1), and the bunny and dragon scenes [43] are used. Here, a virtual view, which is one of 450 rendered virtual views, is used to analyze the performance of the proposed reconstruction filter. The 450 virtual views are rendered using 325 captured images for each scene. The rendered virtual views (i.e., ground truth) are depicted in Fig. 12(a1)–(d1) for the corresponding scenes. Additionally, to compare the reconstruction for using the IPF, a linear and spatially invariant reconstruction filter (LSIF), which is presented by Stewart et al. [23], is used. They apply the low-pass filtering method to reduce the aliasing and eliminate the spectral overlap. The appropriate width for the low-pass filter can be given as follows:

$$w = 2\pi / f \left(\frac{\Delta t}{z_{\min}} - \frac{\Delta t}{z_{\max}} \right). \tag{24}$$

The rendering results for the corresponding scenes are depicted in Fig. 12(a2)–(d2) when using the IPF to render the virtual views. Figure 12(a3)–(d3) illustrates the rendering results when using LSIF. For the second wall, the virtual view can be perfectly rendered using either the IPF or the LSIF, as indicated in Fig. 12(a2) and (a3). However, considerable aliasing exists in Fig. 12(b2)–(d2) and (b3)–(d3) for other scenes. The aliasing is labeled by red rectangles. The reason for this phenomenon is because the surface complexity of the three scenes is more complicated than that of the second wall. Additionally, because the number of captured images is too small (i.e., under-sampled), considerable aliasing exists in the rendered results. Apparently, it can be
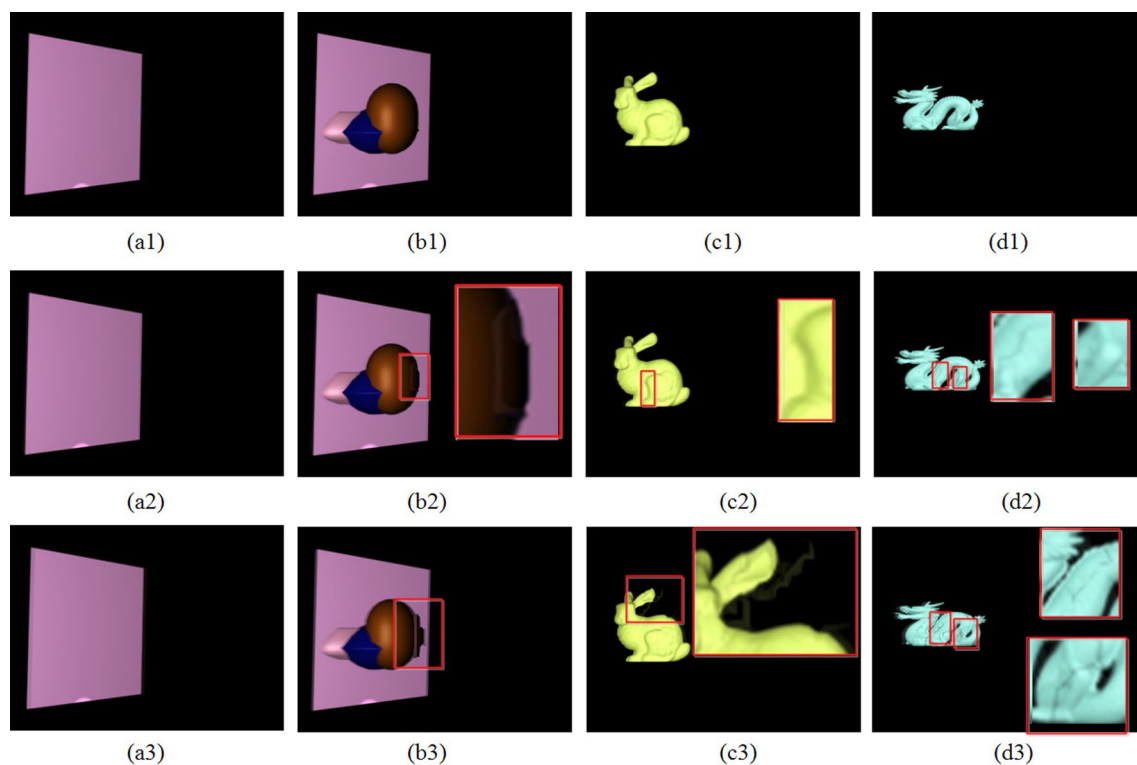
Fig. 12 Virtual views to be rendered using different reconstruction filters for different scenes. Diagrams (a1)–(d1) depict the virtual views to be rendered (ground truth): (a1) the second wall scene; (b1) the fifth wall scene; (c1) the bunny scene; and (d1) the dragon scene. Diagrams (a2)–(d2) depict the rendering results for the corresponding scenes using the IPF. Diagrams (a3)–(d3) illustrate the rendering results of the corresponding scenes using the LSIF

observed in Fig. 12(a2)–(d2) and 12(a3)–(d3) that the rendering quality when using the IPF is better than the rendering quality when using the LSIF. For example, the aliasing in Fig. 12(b3) is more considerable than that in Fig. 12(b2). A similar phenomenon can be observed in Fig. 12(c2), (c3), (d2), and (d3). Therefore, we can assume that the IPF can be applied to reconstruct the virtual views by a number of captured images. Furthermore, the performance of the reconstruction results precedes using the LSIF.

To enable a comparison of accuracy with alternative solutions, a public available dataset, which is described in [44], is also applied to perform the experimental evaluation. The scenes are named aquarium-20, street-10, tree-18, and yellowhouse-12, which are depicted in Fig. 13(a1)–(d1), respectively. For each scene, a novel view is rendered by 12 captured images using different reconstruction filters (i.e., the IPF and LSIF).

We present a set of rendered views to describe the results. The results rendered by the IPF are depicted in Fig. 13(a2)–(d2). Figure 13(a3)–(d3) illustrates the corresponding rendering results when using the LSIF. It can be observed in Fig. 13(a2)–(d2) or 13(a3)–(d3) that considerable aliasing exists in the rendered views of these scenes due to the number of captured images being too small (i.e.,

under-sampled). The red rectangles in the figures denote the areas of aliasing. Apparently, the aliasing in Fig. 13(a3)–(d3) is more considerable than that in Fig. 13(a2)–(d2). Furthermore, Table 2 lists the PSNRs for the rendered images. For example, the PSNRs in Fig. 13(a2) and (a3) are 24.68 and 23.74 dB, respectively. Based on the above analysis, the results of this experiment also indicates that the performance of the reconstruction result precedes that of the LSIF.

6.3 Reconstructions of the synthetic EPI volumes

Now, we can analyze the reconstruction of the EPI-volumes generated from different numbers of captured images using four synthetic scenes, as depicted in Fig. 12(a1)–(d1). For experimental purposes, the influence of the number of captured images on the rendering quality can be derived, and our proposed sampling results for the POF can be evaluated. The experimental results are conducted in static scene, in which the images are captured at different spatial positions along a straight line. The reconstruction filter is the skewed reconstruction filter presented in Sect. 5.3.

Generally, the number of virtual views is larger than the number of captured images []. In this analysis, the number of rendered virtual views is 450, and the rendered virtual
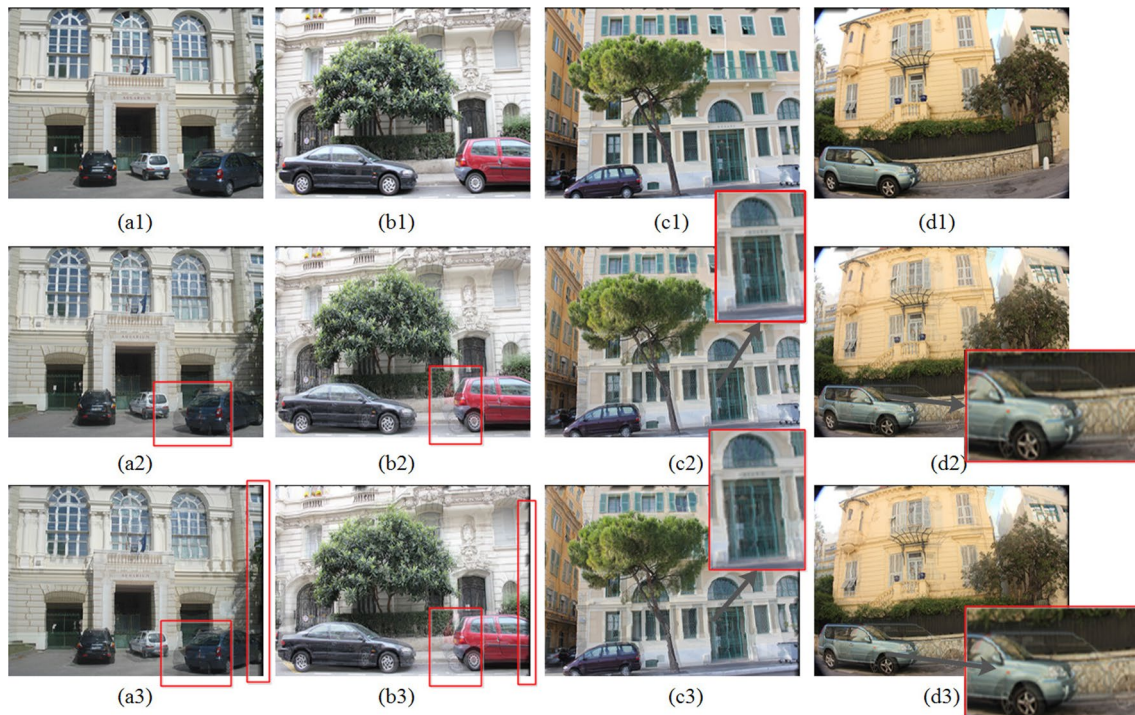
Fig. 13 New view rendered using different reconstruction filters for the aquarium-20, street-10, tree-18, and yellowhouse-12 scenes [44]. Diagrams (**a1**)–(**d1**) illustrate the ground truths: (**a1**) aquarium-20 scene, (**b1**) street-10 scene, (**c1**) tree-18 scene, and (**d1**) *yellow-house*-12 scene. Diagrams (**a2**)–(**d2**) depict the rendering results of the corresponding scenes using the IPF. Diagrams (**a3**)–(**d3**) depict the rendering results of the corresponding scenes using the LSIF

Table 2 The PSNRs (dB) using our proposed method and LSIF for different scenes

Methods	Aquarium-20	Street-10	Tree-18	Yellowhouse-12
LSIF	23.74	23.02	24.02	23.35
IPF	24.68	25.14	27.76	23.75

views are uniform along the line $L = \{(x, 0, 0)\}$. Here, $x \in [-100.0, 100.0]$ cm. Different numbers of captured images in the range of [10, 425] are tested, and the number is increased by increments of 25 to analyze the relationship between the rendering quality and the number of captured images. For all of the cases, the four scenes are fixed at the point $(0, -40, 400)$ cm. The camera is uniformly moved to capture images along the straight line L.

The rendering results are illustrated in Figs. 14 and 15. It can be observed in Fig. 14 that considerable aliasing exists in the rendered EPI-volumes when the number of captured images is 25. Furthermore, when the number of captured image is 25, the aliasing of the fifth wall in Fig. 14b is more significant than that in Fig. 14a. This phenomenon can also be observed by the rendered EPI-volumes when the number of captured images is 75. The reason for this phenomenon is that the scene complexity of the fifth wall is more complicated

than that of the second wall. Additionally, because it is undersampled, a significant aliasing exists in the rendered results. From this experiment, we can see that the rendering quality reduces as the scene complexity becomes more complicated.

Intuitively, the aliasing will be reduced as the number of captured images increases, as depicted in Fig. 14. For the second wall, minimal aliasing occurs when the number of captured images is 300, as indicated in Fig. 14a. When the number of captured images is 325, no aliasing can be observed. For this wall, the angle between the plane and the x-axis is $\frac{\pi}{4}$ rad, and $\omega_s = 125$ rad/m. Using the width and the height for the wall scene, we can calculate that the maximum depth is 444.9013 cm. Additionally, the focal length is 1.0 cm. Using (21), we can then calculate that the sampling spacing between cameras is approximately 0.6151 cm. Additionally, when the number of captured images is 325, the spacing between the captured images is approximately 0.6154 cm. This result demonstrates that the relationship between the sampling spacing from (21) and the necessary spacing from the experiment is approximately equal. Therefore, we can consider that the results of our proposed method are satisfactory.

For other scenes, the aliasing in the rendered images will also be reduced as the number of captured images increases. For example, the aliasing can be clearly observed when the number of captured images is 25, as indicated in

Fig. 14 The reconstructions of the synthetic EPI using different numbers of captured images for different scenes. In this analysis, the EPI-volume consists of 450 rendered images. In **a**, the rendered EPI-volumes of the second wall are presented. In **b**, the rendered EPI-volumes of the fifth wall are presented. In **c**, the rendered EPI-volumes of the bunny scene are presented. In **d**, the rendered EPI-volumes of the dragon scene are presented
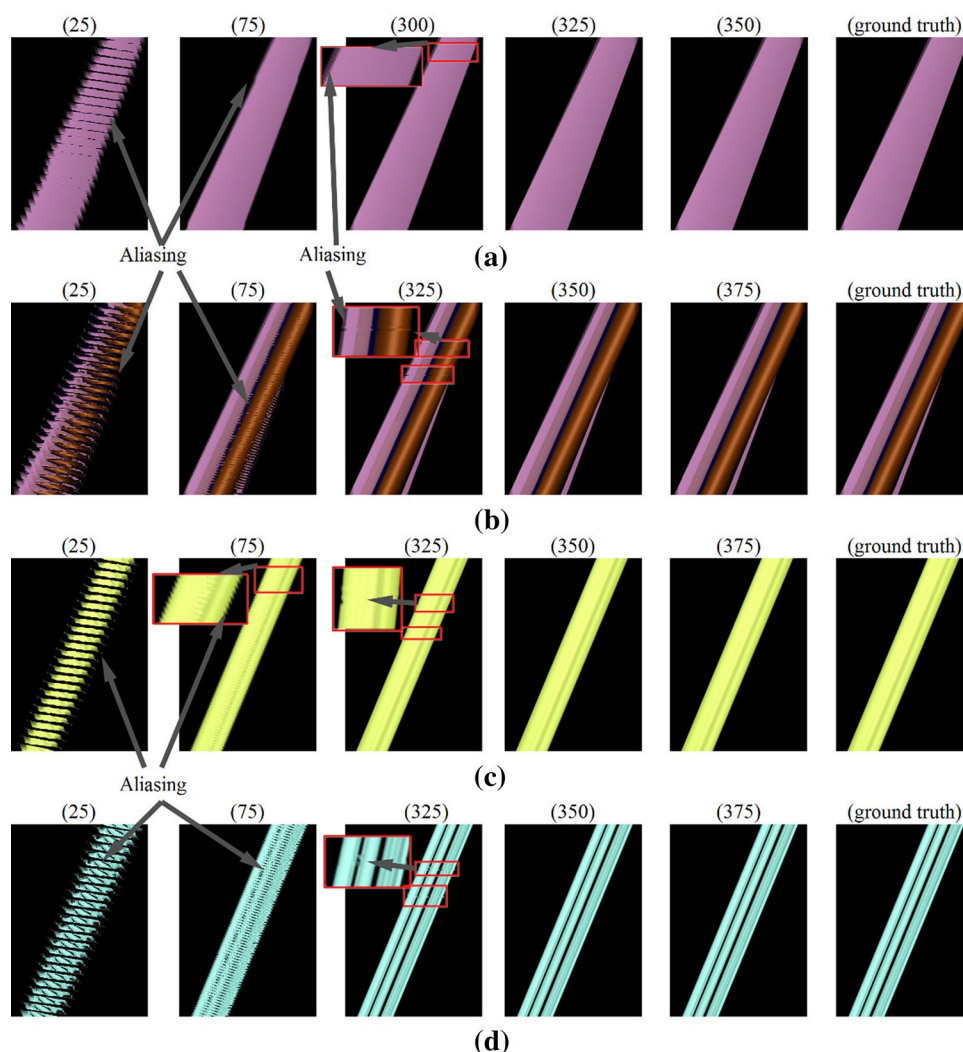

Fig. 14b. In this figure, the aliasing will be reduced when the number of captured images is 75. Furthermore, when the number of captured images is 325, the minimal aliasing would be barely visible. No aliasing is observed when the number of captured images is 350. Furthermore, the aliasing is labeled using red rectangles in the figures. In addition, this phenomenon is shown in Fig. 14c, d for the bunny and dragon scenes, respectively.

Furthermore, the corresponding PSNRs increase as the number of captured images increases, as indicated in Fig. 15. For the second wall, the variations in the PSNR are extremely small when the number of captured images is larger than 325, as depicted in Fig. 15a. This conclusion corresponds to the phenomenon in Fig. 14a, in which no aliasing is observed when the number of captured images is 325. For the other scenes, the variations in the corresponding PSNRs are extremely small when the number of captured images is larger than 350, as illustrated in Fig. 15b–d. Additionally, this conclusion corresponds with phenomenon in Fig. 14b–d, in which no aliasing is observed when the number of captured images is 350. More interestingly,

the PSNRs improve dramatically when the number of captured images increases; however, they improve extremely slowly when the number of captured images is greater than 350. This phenomenon occurs because the influence of the number of captured images on the rendering quality is extremely small when the number of captured images increases to a certain amount. However, there are different variation tendencies for different scenes. The variation tendency of the PSNR for the second wall with the number of captured images is gentler compared to that of the fifth wall. Additionally, the variation of the PSNR in Fig. 15c for the bunny scene is gentler than the dragon scene in Fig. 15d, which is because the surface of the dragon scene is more complicated than the surface of the bunny scene. From the above experiments, the number of captured images needed for the second wall reconstitution in Fig. 14a is 325. Nevertheless, the number of captured images needed for the fifth wall and bunny and dragon is 350.

From the above analysis, the number of captured images needed increases as the surface complexity of scene increases. This conclusion is also described in Sects. 5.2 and

Fig. 15 The PSNR of the rendering images with respect to the number of captured images for different scenes. **a** The PSNR for the second wall scene; **b** the PSNR for the fifth wall scene; **c** the PSNR for the bunny scene; **d** the PSNR for the dragon scene
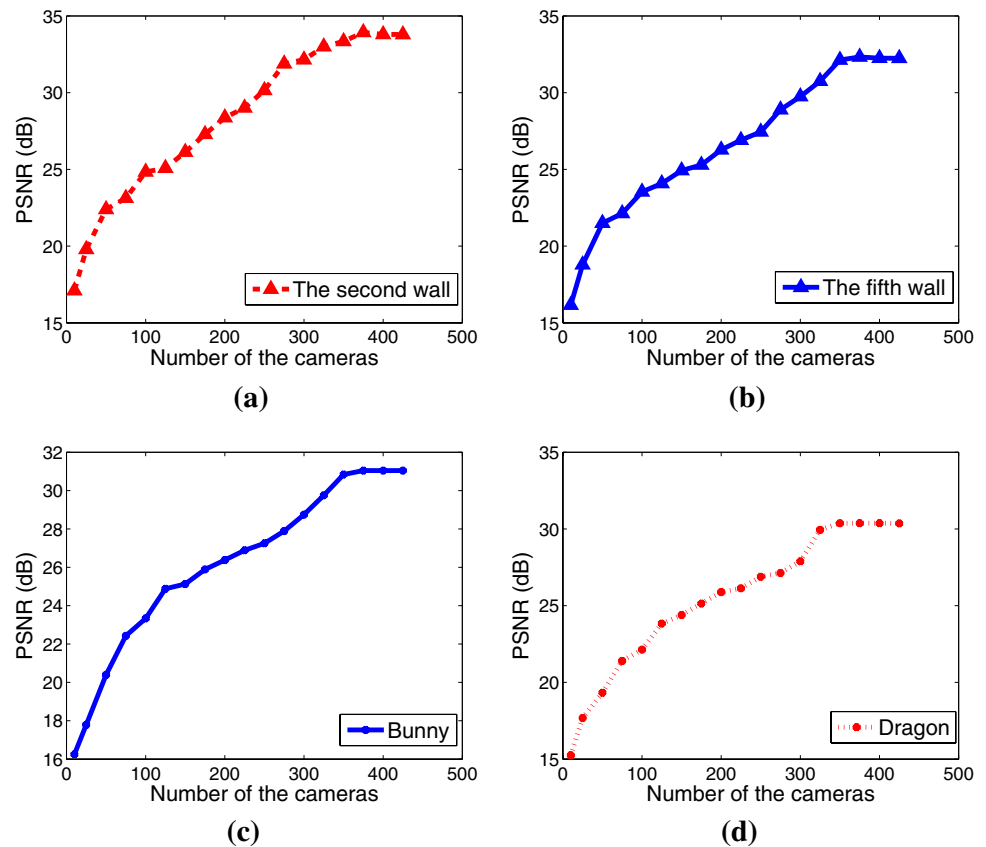
6.1. We can determine that the spectral support of the POF will be broadened as the surface complexity becomes more complicated. Furthermore, the frequency of the POF will be larger as the spectral support broadens from (18). Then, the number of captured images needed increases as the frequency of the POF becomes larger, as indicated by (21).

7 Conclusion

In this paper, we introduced and studied the sampling of the POF. Its characterization of the scene is static at any point in space. This function has been studied and its spectrum along a linear axis with the scene surface is flat with Lambertian assumptions. The influence of the scene geometry on the spectrum has been studied along both the image plane and camera plane frequency axes. Based on the support of the spectrum, the number and the spacing between the cameras needed to reconstruct the continuous light field up to a certain camera plane frequency have been determined. A quantitative sampling theorem, i.e., the minimum sampling rate for plenoptic sampling, has been presented. Finally, the experimental results are presented and compared with the theoretical results.

Appendix

Proof of Theorem 1

In this appendix, we derive the plenoptic spectrum with respect to the scene geometry. Note that this derivation references to [15] and [16]. Starting from (10), the expression of the plenoptic spectrum with respect to s and θ can be rewritten as follows:

$$P(\Omega_t, \Omega_v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f \cdot \frac{l(s, \theta) \cdot \Lambda}{\cos^2(\theta)} \cdot e^{-j(\Omega_v f \tan(\theta) + \Omega_t(\Psi(s) - z(\Psi(s)) \tan(\theta)))} d\theta ds,$$
(25)

where $\Lambda = \Psi'(s) - z'(\Psi(s))\Psi'(s)\tan(\theta)$. In the extreme case, the scene surface is often assumed to be Lambertian [13–16], which means $l(s, \theta) = l(s)$ for all θ. Thus, the following expression of the pelnoptic spectrum for a Lambertian scene is obtained as

$$P(\Omega_t, \Omega_v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f \cdot \frac{l(s,) \cdot \Lambda}{\cos^2(\theta)}$$
$$\cdot e^{-j(\Omega_v f \tan(\theta) + \Omega_t(\Psi(s) - z(\Psi(s)) \tan(\theta)))} d\theta ds, \tag{26}$$

To expediently analyze the plenoptic spectrum by (26), we define an intermediate variable $H(s, \Omega_A)$, and its expression is

$$H(s, \Omega_A) = \int_{-\infty}^{\infty} \frac{f \Psi'(s)\left(1 - z'(\Psi(s)) \tan(\theta)\right)}{\cos^2(\theta)} e^{-j\Omega_A \tan(\theta)} d\theta, \tag{27}$$

where $\Omega_A = \Omega_v f - \Omega_t z(\Psi(s))$. Then, (27) can be split into two parts as follows:

$$H(s, \Omega_A) = H_1 + H_2, \tag{28}$$

where

$$H_1 = \Psi'(s)f \int_{-\infty}^{\infty} \frac{e^{-j\Omega_A \tan(\theta)}}{\cos^2(\theta)} d\theta = 2\pi f \Psi'(s)\delta(\Omega_A). \tag{29}$$

$$\begin{aligned} H_2 &= \int_{-\infty}^{\infty} \frac{f \cdot z'(\Psi(s))\Psi'(s) \tan(\theta)}{\cos^2(\theta)} e^{-j\Omega_A \tan(\theta)} d\theta \\ &= f \cdot z'(\Psi(s))\Psi'(s) \int_{-\infty}^{\infty} \tan(\theta) e^{-j\Omega_A \tan(\theta)} d\tan(\theta) \\ &= j2\pi f \cdot z'(\Psi(s))\Psi'(s) \frac{d}{d\Omega_A}\delta(\Omega_A). \end{aligned} \tag{30}$$

where $\delta(\cdot)$ is a Dirac function. Finally, substituting (29) and (30) to (27), we obtain a new expression of $P(\Omega_t, \Omega_v)$ as follows:

$$\begin{aligned} P(\Omega_t, \Omega_v) &= \int_{-\infty}^{\infty} H(s, \Omega_A)l(s)e^{-j\Omega_t\Psi(s)} ds \\ &= \int_{-\infty}^{\infty} 2\pi f \Psi'(s)\delta(\Omega_A) \\ &\quad \times \left(1 + j \cdot z'(\Psi(s))\frac{d}{d\Omega_A}\right)l(s)e^{-j\Omega_t\Psi(s)} ds. \end{aligned} \tag{31}$$

By (31), the spectrum of the POF is nonzero if $\Omega_A = \Omega_v f - \Omega_t z(\Psi(s)) = 0$. The spectral characteristics of the POF are shown as

$$P(\Omega_t, \Omega_v) = \begin{cases} \int_{-\infty}^{\infty} H(s, \Omega_v f - \Omega_t z(\Psi(s)))l(s)e^{-j\Omega_t\Psi(s)} ds, \\ \quad \text{when} \quad \Omega_t = \Omega_v \frac{f}{z(\Psi(s))}, \quad \Omega_t = 0. \\ 0, \quad \text{when} \quad \Omega_t \neq \Omega_v \cdot \frac{f}{z(\Psi(s))}, \quad \Omega_t \neq 0. \end{cases} \tag{32}$$

Proof of Theorem 2

In this appendix, we derive the plenoptic spectrum with respect to a texture signal $l(s) = \cos(\omega_s s)$. First, substituting $l(s) = \cos(\omega_s s)$ to (11), we can obtain the following expression for the plenoptic spectrum as

$$P(\Omega_t, \Omega_v) = 2\pi f \cdot \int_{-\infty}^{\infty} \Psi'(s)\delta(\Omega_A) \cos(\omega_s s)e^{-j\Omega_t\Psi(s)} ds + j2\pi f \cdot$$
$$\times \int_{-\infty}^{\infty} z'(\Psi(s))\Psi'(s)\frac{d}{d\Omega_A}\delta(\Omega_A) \cos(\omega_s s)e^{-j\Omega_t\Psi(s)} ds, \tag{33}$$

To calculate (33), we define two intermediate variables C and D as follows:

$$C = 2\pi f \int_{-\infty}^{\infty} \Psi'(s)\delta(\Omega_A) \cos(\omega_s s)e^{-j\Omega_t\Psi(s)} ds. \tag{34}$$

$$D = j2\pi f \cdot \int_{-\infty}^{\infty} z'(\Psi(s))\Psi'(s)\frac{d}{d\Omega_A}\delta(\Omega_A) \cos(\omega_s s)e^{-j\Omega_t\Psi(s)} ds. \tag{35}$$

The derivation of the integrals C and D is based on the properties of the Dirac function as shown in the following:

$$\int_{-\infty}^{+\infty} F(x)\delta(\kappa(G-x))dx = F(G)/|\kappa|, \tag{36}$$

$$\int_{-\infty}^{+\infty} F(x)\delta'(G-x)dx = F'(G), \tag{37}$$

where $F(x)$ is a function for x, $\delta(\cdot)$ is the Dirac function with x, and κ and G are scalar parameters. Note that $F'(x)$ is the derivative of $F(x)$ with respect to x. Based on the above properties of the Dirac function, we first derive the integral for C. For Ω_A, it can be written as

$$\Omega_A = \Omega_t\left(\frac{\Omega_v f}{\Omega_t} - z(\Psi(s))\right). \tag{38}$$

To expediently perform the integral, we define a variable a, and let $a = z(\Psi(s))$. Then, $s = \Psi^{-1}\left(z^{-1}(a)\right)$. Substituting a to (34), the expression of C is rewritten as

$$C = 2\pi f \int_{-\infty}^{\infty} \delta\left(\Omega_t\left(\frac{\Omega_v f}{\Omega_t} - a\right)\right)\Phi(a)da, \tag{39}$$

where $\Phi(a)$ is an intermediate variable, and it is a function for a. That is

$$\Phi(a) = \cos\left(\omega_s\Psi^{-1}\left(z^{-1}(a)\right)\right) \cdot \Psi'$$
$$\left(\Psi^{-1}\left(z^{-1}(a)\right)\right) \cdot \left(\Psi^{-1}\left(z^{-1}(a)\right)\right)' \cdot e^{-j\Omega_t z^{-1}(a)}, \tag{40}$$

where $z^{-1}(\cdot)$ is the inverse function of $z(\cdot)$, and $\Psi^{-1}(\cdot)$ is the inverse function (i.e., (36)) of $\Psi(\cdot)$. According to the properties of the Dirac function and (39), C is calculated as

$$\begin{aligned} C &= \frac{2\pi f \Phi\left(\frac{\Omega_v f}{\Omega_t}\right)}{|\Omega_t|} = \frac{2\pi f}{|\Omega_t|}\cos\left(\omega_s\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right) \\ &\quad \times \Psi'\left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right) \cdot \\ &\quad \times \left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)' \cdot e^{-j\Omega_t z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)}. \end{aligned} \tag{41}$$

In the next, we calculate the integral for D. Similar to the expression of (39), substituting a to (35), the expression of D is rewritten as

$$D = \frac{-j2\pi f}{\Omega_t} \cdot \int_{-\infty}^{\infty} \delta'\left(\Omega_t\left(\frac{\Omega_v f}{\Omega_t} - a\right)\right)\Phi_D(a)\mathrm{d}a, \qquad (42)$$

where $\delta'(\cdot)$ is the derivative of $\delta(\cdot)$ with respect to a, where $\Phi_D(a)$ is an intermediate variable, and it is a function for a. That is

$$\Phi_D(a) = \cos\left(\omega_s\Psi^{-1}\left(z^{-1}(a)\right)\right) \cdot z'\left(z^{-1}(a)\right)$$
$$\times \Psi'\left(\Psi^{-1}\left(z^{-1}(a)\right)\right) \cdot \left(\Psi^{-1}\left(z^{-1}(a)\right)\right)' \cdot \mathrm{e}^{-j\Omega_t z^{-1}(a)}. (43)$$

According to the properties of the Dirac function (i.e., (37)) and (42), D is calculated as

$$D = \frac{j2\pi f \Phi_D'\left(\frac{\Omega_v f}{\Omega_t}\right)}{(\Omega_t)^2}$$
$$= \frac{j2\pi f}{(\Omega_t)^2}\left(\cos\left(\omega_s\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)\right.$$
$$\cdot z'\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right) \cdot \mathrm{e}^{-j\Omega_t z^{-1}(\Omega_v f/\Omega_t)}$$
$$\times \Psi'\left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)$$
$$\left.\cdot \left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)'\right)'. \qquad (44)$$

Finally, based on (41) and (44), (32) is calculated as

$$P(\Omega_v, \Omega_t) = C + D$$
$$= \frac{2\pi f}{|\Omega_t|}\cos\left(\omega_s\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)$$
$$\cdot \Psi'\left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)$$
$$\times \left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)' \cdot \mathrm{e}^{-j\Omega_t z^{-1}(\Omega_v f/\Omega_t)}$$
$$+ \frac{j2\pi f}{(\Omega_t)^2}\left(\cos\left(\omega_s\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)\right.$$
$$\times \Psi'\left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)$$
$$\cdot \left(\Psi^{-1}\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right)\right)'$$
$$\left.\cdot z'\left(z^{-1}\left(\frac{\Omega_v f}{\Omega_t}\right)\right) \cdot \mathrm{e}^{-j\Omega_t z^{-1}(\Omega_v f/\Omega_t)}\right)'. \quad (45)$$

Note that the above analysis must be based on the case of bandwidth-limited for the plenoptic spectrum. Therefore, for (45), we restrict the bandwidth of the plenoptic spectrum in Ω_t-axis. Based on the results in [14–18] and (45), Ω_t can be defined as

$$\Omega_t = \left[\max\left(\left|\frac{1}{\Psi'(s)+z'(\Psi(s))\Psi'(s)\max(\theta)/f}\right|\right)\cdot\omega_s,\right.$$
$$\left.\max\left(\left|\frac{1}{\Psi'(s)-z'(\Psi(s))\Psi'(s)\max(\theta)/f}\right|\right)\cdot\omega_s\right]. \qquad (46)$$

References

1. Kubota, A., Smolic, A., Magnor, M., Tanimoto, M., Chen, T., Zhang, C.: Multiview imaging and 3DTV. IEEE Signal Process. Mag. **24**(6), 10–21 (2007)
2. Tanimoto, M., Tehrani, M.P., Fujii, T., Yendo, T.: Free-viewpoint TV. IEEE Signal Process. Mag. **28**(1), 67–76 (2011)
3. Chen, Z., Barnes, J.F., Bodenheimer, B.: Hybrid and forward error correction transmission techniques for unreliable transport of 3D geometry. Multimed. Syst. **10**(3), 230–244 (2005)
4. Zhang, C., Chen, T.: Light field sampling. Synthesis lectures on image, video, and multimedia processing. Morgan and Claypool Publishers 2(1), 1–102 (2006)
5. Levoy, M.: Light fields and computational imaging. IEEE Comput. **39**(8), 46–55 (2006)
6. Shum, H.Y., Chan, S.C., Kang, S.B.: Image-based Rendering. Springer, New York, NY (2007)
7. Shum, H.Y., Kang, S., Chan, S.C.: Survey of image-based representations and compression techniques. IEEE Trans. Circ. Syst. Video Technol. **13**(11), 1020–1037 (2003)
8. Zhang, C., Chen, T.: A survey on image-based rendering-representation, sampling and compression. EURASIP Signal Process. Image Commun. **19**(1), 1–28 (2004)
9. Kozuch, M., Wolf, W., Wolfe, A.: An experimental analysis of digital video library servers. Multimed. Syst. **8**(2), 135–145 (2000)
10. Berent, J., Dragotti, P.L.: Plenoptic manifolds. IEEE Signal Process. Mag. **24**(6), 34–44 (2007)
11. Adelson, E. H., Bergen, J.R.: The plenoptic function and the elements of early vision. In: Proceedings of the Computational Models of Visual Processing. Cambridge, MA, USA: MIT Press, pp. 3–20 (1991)
12. McMillan, L., Bishop, G.: Plenoptic modeling: an image-based rendering system. In: Proceedings of the Computer Graphics (SIGGRAPH'95), pp. 39–46 (1995)
13. Chai, J.X., Tong, X., Chan, S.C., Shum, H.Y.: Plenoptic sampling. In: Proceedings of the SIGGRAPH. New York, NY, pp. 307–318 (2000)
14. Zhang, C., Chen, T.: Spectral analysis for sampling image-based rendering data. IEEE Trans. Circuits Syst. Video Technol. **13**(11), 1038–1050 (2003)
15. Do, M.N., Marchand-Maillet, D., Vetterli, M.: On the bandwidth of the plenoptic function. IEEE Trans. Image Process. **21**(2), 708–717 (2012)
16. Gilliam, C., Dragotti, P., Brookes, M.: On the spectrum of the plenoptic function. IEEE Trans. Image Process. **23**(2), 502–516 (2014)
17. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of the SIGGRAPH. New Orleans, LA, pp. 31–40 (1996)
18. Gortler, S., Grzeszczuk, R., Szeliski, R., Cohen, M.: The lumigraph. In: Proceedings of the SIGGRAPH, pp. 43–54 (1996)
19. Emilio, C., Lerios, A., Fussell, D.: Uniformly sampled light fields. In: Proceedings of the Rendering Techniques'98, Springer, Vienna, pp. 117–130 (1998)
20. Shum, H.Y., He, L.W.: Rendering with concentric mosaics. In: Proceedings of the Computer Graphics (SIGGRAPH'99), pp. 299–306 (1999)
21. Liang, C.K., Shih, Y.C., Chen, H.H.: Light field analysis for modeling image formation. IEEE Trans. Image Process. **20**(2), 446–460 (2011)
22. Ajdler, T., Sbaiz, L., Vetterli, M.: The plenacoustic function and its sampling. IEEE Trans. Signal Process. **54**(10), 3790–3804 (2006)
23. Stewart, J., Yu, J., Gortler, S.J., McMillan, L.: A new reconstruction filter for undersampled light fields. In: Proceedings of the ACM International Conference Proceeding Series, vol. 44, pp. 150–156 (2003)

24. Hoshino, H., Okano, F., Yuyama, I.: A study on resolution and aliasing for multi-viewpoint image acquisition. IEEE Trans. Circuit. Syst. Video Technol **10**(3), 366–375 (2000)

25. Lin, Z.C., Shum, H.Y.: On the number of samples needed in light field rendering with constant-depth assumption. In: Proceedings of the CVPR, pp. 588–595 (2000)

26. Nguyen, H.T., Do, M.N.: Error analysis for image-based rendering with depth information. IEEE Trans. Image Process. **18**(4), 703–716 (2009)

27. Shade, J., Gortler, S., He, L.W., Szeliski, R.: Layered depth images. In: Proceedings of the SIGGRAPH. Orlando, FL, pp. 231–242 (1998)

28. Zitnick, C.L., et al.: High-quality video view interpolation using a layered representation. ACM Trans. Graph. **23**(3), 600–608 (2004)

29. Pearson, J., Brookes, M., Dragotti, P.L.: Plenoptic layer-based modeling for image based rendering. IEEE Trans. Image Process. **22**(9), 3405–3419 (2013)

30. Shidanshidi, H., Safaei, F., Li, W.: Estimation of signal distortion using effective sampling density for light field-based free viewpoint video. IEEE Trans. Multimed. (2015). doi:10.1109/TMM.2015.2447274

31. Shidanshidi, H., Safaei, F., Li, W.: Objective evaluation of light field rendering methods using effective sampling density. In: Proceedings of the 2011 IEEE 13th International Workshop on Multimedia Signal Processing (MMSP). Hangzhou, pp. 1–6 (2011)

32. Chen, C., Schonfeld, D.: Geometrical plenoptic sampling. In: Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 3769–3772 (2009)

33. Isaksen, A., McMillan, L., Gortler, S.J.: Dynamically reparameterized light fields. In: Proceedings of the ACM SIGGRAPH, pp. 297–306 (2000)

34. Buehler, C., Bosse, M., McMillan, L., Gortler, S.J., Cohen, M.F.: Unstructured lumigraph rendering. In: Proceedings of the SIGGRAPH, pp. 425–432 (2001)

35. Bagnato, L., Frossard, P., Vandergheynst, P.: Plenoptic spherical sampling. In: Proceedings of 19th IEEE ICIP, pp. 357–360 (2012)

36. Li, Z., Wong, K.H., Leung, M.C., Ko, H.F., Lee, K.K., Chang, M.M.Y.: An interactive handheld spherical 3D object display system. Multimed. Syst. **17**(5), 435–447 (2011)

37. Liu, Z., Wang, Z., Ma, C., Zhang, C., Mitani, J., Fukui, Y.: Shape alignment and shape orientation analysis-based 3D shape retrieval system. Multimed. Syst. **16**(4–5), 319–333 (2010)

38. Bolles, R., Baker, H., Marimont, D.: Epipolar-plane image analysis: an approach to determining structure from motion. Int. J. Comput. Vis. **1**(1), 7–55 (1987)

39. Shannon, C.E.: Communications in the presence of noise. Proc. IREE **37**, 10–21 (1949)

40. Vetterli, M., Kovacevic, J.: Wavelets and Subband Coding. Signal Processing. Prentice-Hall, Englewood Cliffs, NJ (1995)

41. Vaidyanathan, P.: Multirate Systems and Filter Banks. Prentice-Hall, Englewood Cliffs, NJ (1992)

42. Jung, C., Wang, L., Zhu, X., Jiao, L.: 2D to 3D conversion with motion-type adaptive depth estimation. Multimed. Syst. 21(5), 451–464 (2014)

43. Stanford 3D scanning. https://graphics.stanford.edu/data/3Dscanrep/. Accessed 7 Jan 2016

44. Chaurasia, G., Sorkine-Hornung, O., Drettakis, G.: Silhouette-aware warping for image-based rendering. In: Proceedings of the Computer Graphics Forum, vol. 30(4), pp. 1223–1232 (2011)