Tomographic approach for the quantitative scene reconstruction from light field images

NICOLA VIGANÒ,[1,*] HENRI DER SARKISSIAN,[1] CHARLOTTE HERZOG,[2,3,4] OMBELINE DE LA ROCHEFOUCAULD,[2] ROBERT VAN LIERE,[1] AND KEES JOOST BATENBURG[1,5]

[1]*Centrum Wiskunde & Informatica (CWI), NWO, Amsterdam, The Netherlands*
[2]*Imagine Optic, Bordeaux, France*
[3]*Laboratoire Photonique, Numérique et Nanosciences UMR 5298: Institut d'Optique Graduate School, CNRS, Univ. Bordeaux, France*
[4]*Inria Bordeaux Sud-Ouest, Bordeaux, France*
[5]*Leiden Universiteit, Leiden, The Netherlands*
[*]N.R.Vigano@cwi.nl

Abstract: Current computational methods for light field photography model the ray-tracing geometry inside the plenoptic camera. This representation of the problem, and some common approximations, can lead to errors in the estimation of object sizes and positions. We propose a representation that leads to the correct reconstruction of object sizes and distances to the camera, by showing that light field images can be interpreted as limited angle cone-beam tomography acquisitions. We then quantitatively analyze its impact on image refocusing, depth estimation and volumetric reconstructions, comparing it against other possible representations. Finally, we validate these results with numerical and real-world examples.

1. Introduction

In recent years, plenoptic cameras and light field imaging have become trending topics by allowing new possibilities in the field of computational photography, such as the possibility to refocus on different planes of the scene [1, 2], and the ability to extract depth information regarding the photographed objects [3].

The first appearances of commercially available light field cameras have mostly been targeted at digital photography applications. The requirements for such applications are usually confined to the mentioned post-acquisition refocusing. However, a quantitative treatment of the light field data can benefit to a whole class of applications where an accurate estimation of object sizes and distances is crucial, ranging from industrial quality control and automation processes to the trendy topic of autonomous driving. Quantitative light field image processing can also benefit already established imaging techniques. An example of these relevant techniques is photogrammetry, which aims at extracting the exact position of surface points from collections of digital photographs. By doing so, it allows to reconstruct the objects' three-dimensional contour.

In this article, we address the topics of image refocusing, objects position and size estimation, and volumetric scene reconstruction. When determining the object positions, aside from their lateral offset from the optical axis of the camera, light field images allow to identify their distance from the camera along such axis. This procedure, naturally defined for opaque objects, is called *depth estimation*. A *depth map* is an image whose pixel values correspond to the distances between the physical objects imaged in them and the main lens of the light field camera. For translucent objects, it is possible to obtain the volumetric reconstruction of the scene, which presents features of both depth estimation and refocusing.

Each of these computational tasks are based on the modeling of the light rays in space. We will refer to different ray tracing models as *ray parametrizations*. Modeling and processing of the light field data is currently based on the ray tracing of chief rays going through the optical elements

inside the camera [1, 4]. Due to the characteristics of the optical elements involved, the ray paths inside the camera might not be accurately representative of the ray paths in the photographed scene, and lead to estimation errors. Moreover, it is common practice to ignore the divergence of the rays. This approximation is usually introduced to reduce the computational burden and produce images of consistent size [5]. While this is desirable for common digital photography applications, it can lead to additional errors in quantitative distance and size estimations.

Plenoptic cameras are traditionally configured in what is called the "unfocused" mode, which sacrifices a lot of the sensor spatial resolution in favor of the added angular resolution. In the past years, a new configuration has been proposed, which operates in "focused" mode and that allows to obtain higher spatial resolutions compared to the unfocused mode [6, 7].

Tomography is a well established imaging technique that allows to reconstruct a three-dimensional representation of the interior of objects from many projections. The strong similarity between tomography and light field acquisition geometries has already been demonstrated in the field of light field microscopy [8, 9].

Here, we first show that light field photography acquisitions can be transformed into limited angle cone-beam geometry tomography acquisitions. We then use the theory and background of tomography to analyze the limitations of existing light field representations and construct a new ray parametrization. Such parametrization produces the correct reconstruction of object sizes, and leads to the accurate estimation of their distances from the camera. We also compare the proposed ray parametrization against other common choices, and show the impact of each parametrization of the estimated positions and sizes. This step is a prerequisite for carrying out quantitative analysis using light field images. Moreover, the proposed ray parametrization is agnostic of the underlying camera geometry, and it can operate with both the focus and unfocused plenoptic cameras.

To be able to derive the proposed ray parametrization, we first show that ignoring dilation effects due to the different refocusing distances results in the change from a cone-beam to a parallel-beam geometry. We then analyze the differences between modeling the refocusing reconstruction based on the ray tracing inside and outside the camera: the main lens' object and image space, respectively. We demonstrate that the two geometries become increasingly different as the magnification factor increases, which suggests that a correct reconstruction of the scenes should be modelled and carried out in the object space.

Our analysis also shows that carrying out a refocusing reconstruction in the object space is equivalent to performing a simple tomographic back-projection over the refocusing stack. We analyze the achievable depth resolution for such reconstructions in light field photography. We introduce a three-dimensional representation for the light field images that is based on the tomographic three-dimensional Fourier-slice theorem [10]. Such a representation was also used to reduce the storage footprint of light field images [11]. Within this framework, we show the impact of larger aperture sizes on the depth resolution of volumetric reconstructions.

The paper is organized as follows. In section 2, we address the relationship between tomographic and light field photography acquisition geometries, along with the shortcomings of the previous parametrization and the outcomes of the newly proposed parametrization. In sections 3 and 4, we present and discuss a set of numerical examples to showcase the impact of the developed methods on the quantitative estimation of object sizes and distances. Finally, section 5 concludes the paper.

2. Computational model

To achieve a better understanding of the geometry, we first introduce the terms and conventions used throughout this article. We then proceed to the analysis of the problems of the ray parametrization presented in [1]. After presenting the transformation of light field images into limited angle cone-beam tomography acquisitions in section 2.3, we show the effects of ignoring

(a) Compound plenoptic camera.   (b) Cone-beam tomography.
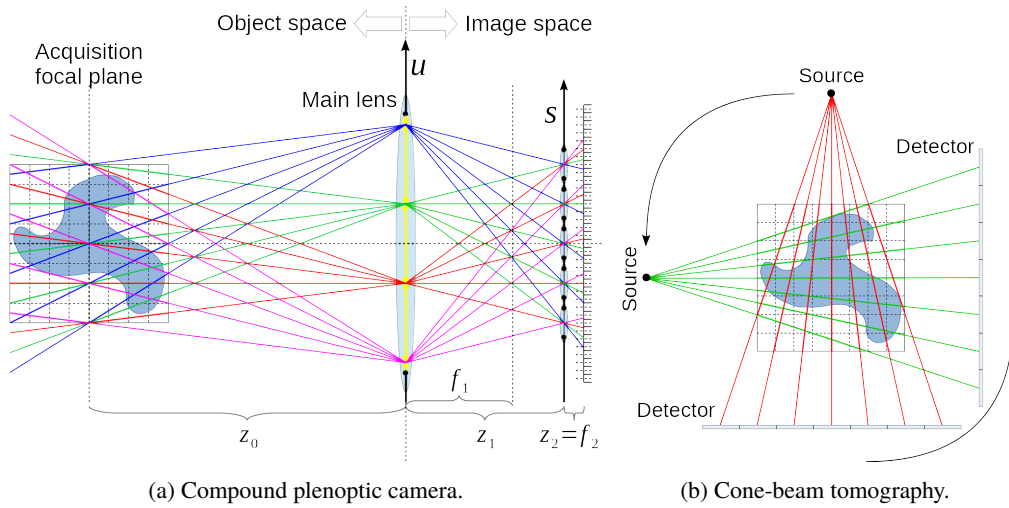
Fig. 1. Comparison of the two setups: (a) compound plenoptic camera using a MLA,
(b) cone-beam tomography acquisition.

the divergence of the rays belonging to the same sub-aperture image (section 2.4), and the effects
of the reconstruction in both image and object spaces (section 2.5).

Following the discussion of the proposed parametrization, we address the related topics of
volumetric reconstructions in section 2.6, and depth estimation accuracy in section 2.7.

2.1. Conventions

We adopt the conventions and representations derived from the usual classical chief ray model
of a generic plenoptic camera [4], depicted in Fig. 1(a). The main compound lenses have been
simplified into a single lens with an incorporated aperture. The distances between the main lens,
the microlens array (MLA), and the detector (sensor), together with their relative sizes, are not
representative of a real camera design, but they were chosen to facilitate the understanding of the
internal camera structure.

The space in Fig. 1(a) is divided into the object space and the image space, respectively on
the left and right hand side of the main lens. As mentioned earlier, the object space is the
photographed scene outside the camera, and the image space is the region inside the camera
where the formation of the acquired image takes place [12].

The distance between the natural focusing plane in the object space and the main lens is called
z_0, while the distance between the main lens and the MLA is called z_1. The focal distances of the
main lens and the lenslets in the MLA are called f_1 and f_2 respectively. The distance between
the MLA and the sensor is called z_2 and is usually set to f_2, and given the fact that $f_2 << z_1$, the
main lens is considered to be at the optical infinity of the lenslets in the MLA.

The coordinates (u, v) define the positions on the main lens, while the coordinates (s, t) define
the positions of each lenslet on the MLA. For the pixels behind each lenslet, we use coordinates
(σ, τ), but in this setup, there is a one-to-one correspondence with coordinates (u, v) on the main
lens. We refer to the collection of pixels behind a lenslet as the micro-image. The pixels sharing
the same (σ, τ) position under each lenslet form a sub-aperture — or pinhole — image. This is
due to the fact that their intensities originate from the same point (u, v), as if there were a pinhole
in that position. The corresponding pinhole is indicated by the yellow aperture in Fig. 1(a).

We now consider the tomographic setup, shown in Fig. 1(b), with a point source emitting a
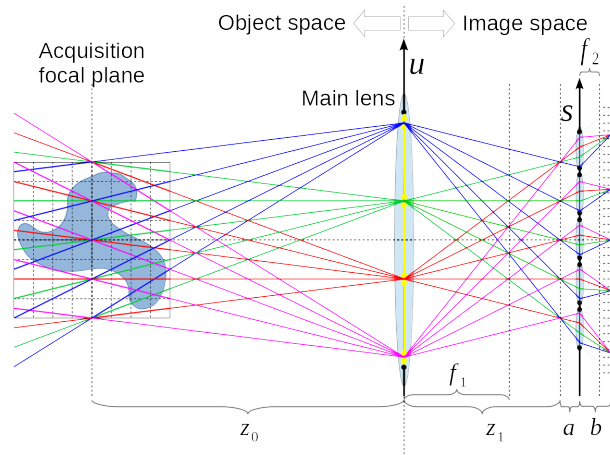beam propagating on a spherical wavefront. This type of setup is characterized by a *conical*

Fig. 2. Ray-tracing geometry for compound plenoptic camera using a focused configuration.

beam. However, in some specific setups, where the source is distant enough from the sample, the beam has a negligible divergence, hence a *parallel-beam* approximation can be made. The tomographic acquisition can be described by the point source and detector rotating around the sample, as shown in Fig. 1(b).

The images collected on the detector represent projections of the sample interiors, and they can be computed by integrating the sample volume over the lines connecting each detector pixel with the source position. The integral of a line over a scalar field does not depend on the direction of integration, and for this reason tomographic acquisitions in a parallel-beam setting are usually performed over 180 degrees, instead of the full circle. When the acquisition is performed on a smaller angular range than usual, the reconstruction problem becomes severely ill-posed. This problem is often referred to as *limited angle* tomography.

As mentioned earlier, tomography usually deals with semi-transparent objects, while when dealing with opaque objects, it gives rise to artifacts (typically called *photon starvation* artifacts). Digital photography does not suffer from this restriction, because it generally works with surface scattered light, and it renders all the collected light as if it were coming from only one plane in the photographed scene. The transmission and surface scattering reconstruction problems become very similar when restricting the support region to only one slice of the sample volume. By applying this restriction, line integrals are collapsed to point sampling over the considered slice, as in the scattering configuration. Moreover, except when dealing with volumetric reconstructions in section 2.6, for simplicity, we assume that all the objects in the scene are opaque.

2.2. Focused light field

The configuration described in section 2.1 is the so called unfocused light field (ULF). A focused light field (FLF) configuration also exists [6,7], which allows superior spatially resolved refocused images, and it is represented in Fig. 2.

The most remarkable difference between the ULF and the FLF is that in the FLF, the (s, t) coordinate axes do not reside on the MLA position, but on the focal plane of the MLA-sensor optical system. We denote as a the distance between the (s, t) coordinate axes and the MLA, and b the distance between the MLA and the sensor, which is in this case different from f_2.
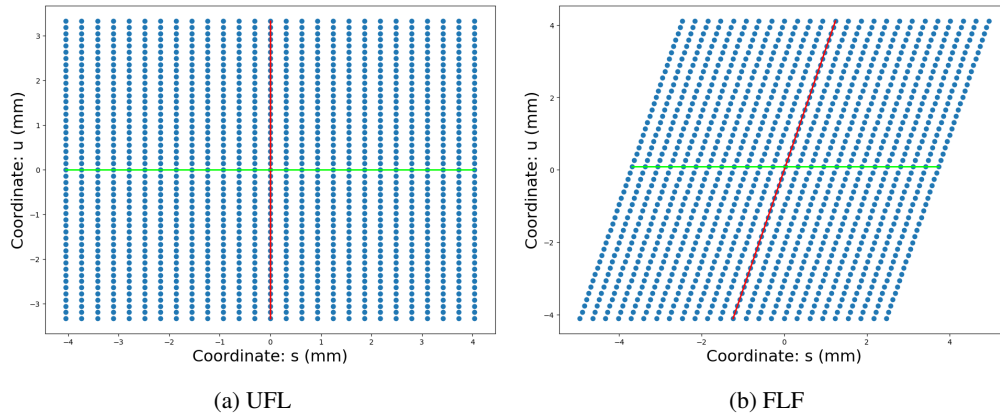
Fig. 3. Comparison of the (s, u) phase space in the: (a) unfocused light field camera, (b) focused light field camera. In each plot, the blue dots represent the sampling corresponding to each sensor pixel, the red line represents one micro image, and the green line represents one sub-aperture image.

The ULF coordinates depend on the MLA and sensor coordinates as follows:

$$s_{\text{ULF}} = s_{\text{MLA}}, \tag{1}$$

$$t_{\text{ULF}} = t_{\text{MLA}}, \tag{2}$$

$$u_{\text{ULF}} = \frac{z_1}{f_2}\sigma, \tag{3}$$

$$v_{\text{ULF}} = \frac{z_1}{f_2}\tau, \tag{4}$$

where the coordinates $(s_{\text{MLA}}, t_{\text{MLA}})$ are the centers of the lenslets. The FLF coordinates instead are as follows:

$$s_{\text{FLF}} = s_{\text{MLA}} + \frac{a}{b}\sigma, \tag{5}$$

$$t_{\text{FLF}} = t_{\text{MLA}} + \frac{a}{b}\tau, \tag{6}$$

$$u_{\text{FLF}} = \frac{z_1 + a}{b}\sigma, \tag{7}$$

$$v_{\text{FLF}} = \frac{z_1 + a}{b}\tau, \tag{8}$$

where the (s, t) coordinates now depend also on the (σ, τ) coordinates. This results in a shearing of the FLF sampling, with respect to the ULF sampling in the (s, t, u, v) phase space, as shown in Fig. 3 for the (s, u) hyperplane. The implication of the sampling shear, is that each sub-aperture image has an additional (s, t) shift on top of the lenslets positions, that depends on its position in (u, v).

All the considerations and mathematical derivations carried out in this article are only based on the effective coordinates (s, t, u, v) of the sampled points. As a consequence, they are independent of the underlying camera configuration. For this reason, we chose to focus only on the ULF version, to simplify the understanding and reduce the redundancy of the article.

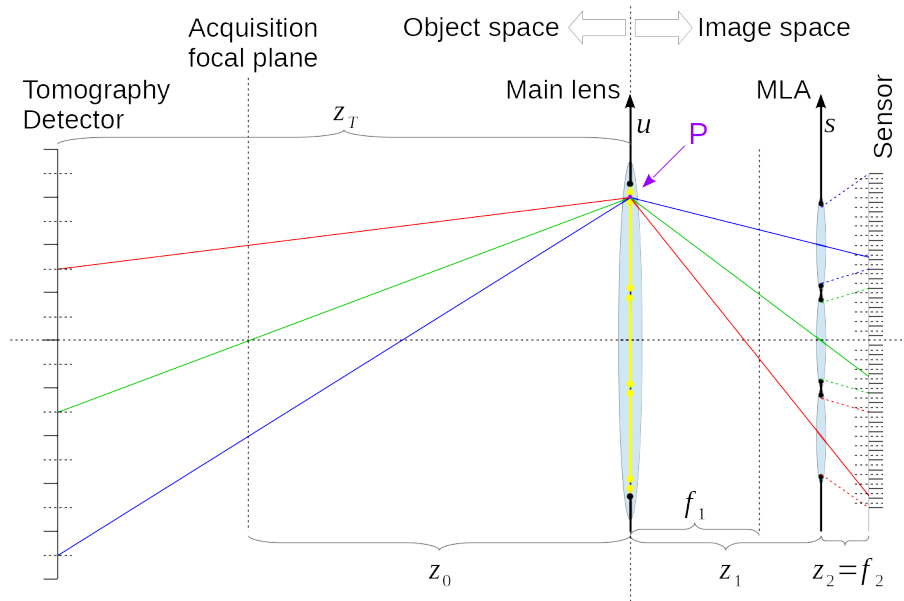For the mathematical derivation of Eqs. 1-(4) and Eqs. 5-(8), we refer to appendix B.

Fig. 4. Ray tracing geometry for one sub-aperture image, superimposed with the corresponding tomographic geometry.

2.3. Limited angle cone-beam acquisition geometry

In this section, we show the geometric equivalence of a light field acquisition to a limited angle cone-beam tomographic acquisition. In fact, one sub-aperture image of the light field corresponds to one tomographic projection. This is clearly visible in Fig. 4, where we show the chief rays belonging to a single selected sub-aperture image. In this interpretation, the light originating from the object is collected by a fictitious pinhole placed over the main lens in point P, before being refracted towards the center of the lenslets over the MLA.

In the same figure, a virtual sensor is depicted behind the object to show that the same line geometry can be interpreted as a regular tomographic projection. In this case, the light rays can be understood as emitted from point P, which acts as a point source, and traveling in the direction of the sample. After passing through the sample, they are collected by the fictitious detector behind the sample.

The only difference between the two interpretations is the direction of propagation of the light rays. This is not a problem, because the intensity deposited on the detector is the line sum over the light rays, which is independent to the direction of propagation. For an in-depth analysis of the intensity deposited in both cases, we refer to appendix A.

2.4. Parallel-beam approximation

The light field refocusing problem is generally formulated as follows: the plane of the MLA is virtually shifted along the optical axis of the main lens, and the rays previously hitting the center of the microlenses are propagated to the new MLA position. We show this ray parametrization in Fig. 5 [1].

Given the ratio α between the new refocusing distance z_1' and the acquisition distance z_1, such parametrization results in the following coordinates change:

$$s' = u + \alpha(s - u) = \alpha s + (1 - \alpha)u, \tag{9}$$

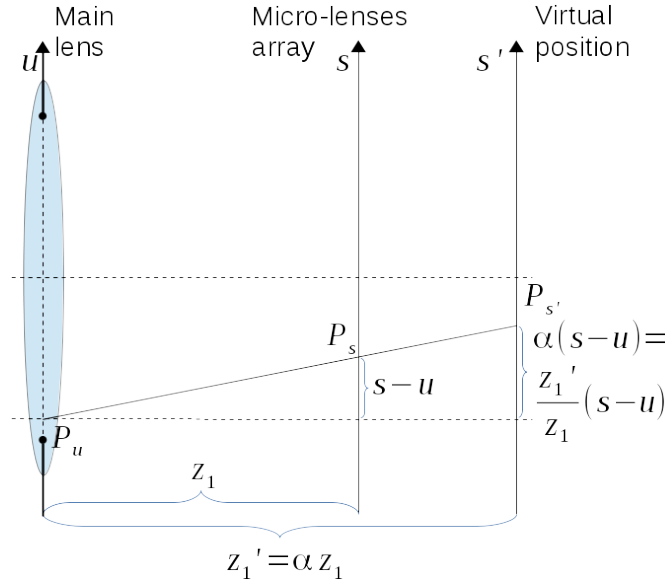$$t' = v + \alpha(t - v) = \alpha t + (1 - \alpha)v, \tag{10}$$

Fig. 5. Ray parametrization proposed in [1].

where s' and t' are the new coordinates where the rays cross the new virtual refocusing plane, and $\alpha = z_1'/z_1$.

We now define the set of rays R_u, as the rays sharing the same u coordinate over the main lens, or, in other words, the rays belonging to the same sub-aperture image. From Eq. (9), we deduce that for the rays in R_u, the shift over the s'-axis is determined by two components: a fixed shift $(1 - \alpha)u$, identical for every ray in the set, and a s-dependent shift which is ray-specific and equal to $(\alpha - 1)s$. This is consistent with the cone-beam geometry identified in section 2.3 for such ray sets.

The cone-beam geometry can be considered to cause the shrinking and dilation of the scene, when refocusing at distances different from the acquisition focal plane. This is shown and discussed in [1], where the author suggests to ignore the dilation factor α for the (s, t) coordinates in order to avoid such effect. On the other hand, by doing so, the s-dependent displacement equal to $(\alpha - 1)s$ is removed, effectively resulting in a parallel-beam geometry for the ray-set R_u. In this case, the rays form a set of parallel rays, with spacing equal to the lenslet size and slope equal to u.

The parallel-beam approximation has two different effects on the refocusing performance, namely the size of the refocused objects outside the acquisition focal plane is dilated by α, and the refocusing distances are not preserved. The relationship between the α parameters in both parametrizations are reported here:

$$\alpha_{\mathrm{p}} = 2 - \frac{1}{\alpha_{\mathrm{c}}}, \tag{11}$$

$$\alpha_{\mathrm{c}} = \frac{1}{2 - \alpha_{\mathrm{p}}}, \tag{12}$$

where α_{p} and α_{c} are respectively the parallel-beam and cone-beam refocusing parameters α. The derivation of Eqs. 11-(12) is detailed in appendix C.

2.5. Refocusing in the object space vs image space

The parametrization seen in section 2.4 is defined in the image space. Its aim is to compute refocused images of the scene as they would be seen on the camera sensor at different acquisition distances. This is equivalent to reconstructing the scene in the object space, in the case of magnification modulus $|M| = z_0/z_1 = 1$. For different magnifications, the ray geometry of the two spaces is different. In fact, for a certain magnification M, the object space coordinates (s_o, t_o) are related to the image space coordinates (s_i, t_i) through the relationships:

$$M s_o = s_i,\tag{13}$$

$$M t_o = t_i,\tag{14}$$

where the magnification is defined as $M = -z_0/z_1$ [12]. While the ray propagation equations in the image space are given in Eqs. 9-(10), for the object space we have the following set of equations:

$$M s_o' = \alpha M s_o + (1 - \alpha)u,\tag{15}$$

$$M t_o' = \alpha M t_o + (1 - \alpha)v,\tag{16}$$

where we have substituted Eqs. 13-(14) into Eqs. 9-(10). Equations 15-(16) mean that, with different values of the magnification M, the coordinates (u, v) remain unchanged, while the coordinates (s, t) are scaled differently in the image space and the object space. This results in a different angular range and resolution of the ray-sets R_u, and, especially for the cone-beam case, in a different matching of the rays outside the focal plane. We address this topic in more details in appendix D.

For these reasons, we propose to reconstruct the scene directly in the object space, like in the tomography setting. The proposed parametrization is very similar to the one from Eqs. 9-(10): aside from being reversed along the z-axis, it is based on the change of distance z_0 to z_0' defined as $z_0' = \alpha z_0$. The resulting coordinate change reduces to Eqs. 9-(10), with only a slight change to the α parameter definition. The relationship between the object space $\alpha_{c,o}$ and image space $\alpha_{c,i}$ is then given by:

$$\alpha_{c,o} = \frac{\alpha_{c,i}}{(1 - |M|)\,\alpha_{c,i} + |M|},\tag{17}$$

$$\alpha_{c,i} = \frac{|M|\alpha_{c,o}}{1 - \alpha_{c,o}\,(1 - |M|)}.\tag{18}$$

For a derivation of Eqs. 17-(18), we refer again to appendix D.

The refocusing in the object space is in essence similar to Eq. (2.1) from [1], but the image space (s_i, t_i) coordinates are substituted with the (s_o, t_o) coordinates, and the distance z_1 with the distance z_0:

$$E_{z_0}(s', t') = \frac{1}{z_0^2} \iint \mathcal{L}_{z_0}(s, t, u, v)\,\mathrm{d}u\,\mathrm{d}v,\tag{19}$$

where $E_{z_0}(s', t')$ is the computed intensity for the refocusing distance z_0', at the pixel coordinates (s', t') and $\mathcal{L}(s, t, u, v)$ is the recorded light field.

In this setting, the refocusing equation is the same to equation 4.1 in [1], but obtained by substituting Eqs. 15-(16) into Eq. (19):

$$E_{z_0'}(s', t') = \frac{1}{z_0^2} \iint \mathcal{L}_{z_0}\left(\frac{s}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\frac{u}{M}, \frac{t}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\frac{v}{M}, u, v\right)\mathrm{d}u\,\mathrm{d}v.\tag{20}$$

The refocusing is performed by integrating all the ray contributions for matching rays in each pixel of the reconstruction at the new distance z_0'.

Equation (20) also describes the back-projection in tomography, when the coordinates (u, v) are intended as angular coordinates of the individual sub-aperture images. This can be verified by expressing the sub-aperture images as the functions $P_{u,v}$, and consequently the light field as $\mathcal{L}(s, t, u, v) = \{P_{u,v}(s, t)\}$. Equation 5.43 from chapter 5 of [13] can then be written in the light field formalism:

$$E_{z_0'}(s', t') = \iint_{-U,-V}^{+U,+V} P_{u,v}(s, t) \, \delta\left(\frac{s}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\frac{u}{M} - s', \frac{t}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\frac{v}{M} - t'\right) du \, dv, \quad (21)$$

where we have omitted the coefficient $1/z_0^2$ and we have assumed that $U = |u_{\max}| = |u_{\min}|$ and $V = |v_{\max}| = |v_{\min}|$.

This results in the possibility to perform the scene reconstruction using existing tomographic tools and routines, which have been extensively studied and optimized over the years and for which efficient implementations exist, for example the ASTRA toolbox [14].

2.6. Volumetric reconstructions and resolution considerations

So far, we only considered traditional light field refocusing, as a mean to reconstruct the photographed scene. As mentioned in section 2.1, light field refocusing addresses the problem slice-by-slice, and the outcome of the scene reconstruction is a stack of refocused images, generally called *focal stack* [1].

In tomography, it is customary to consider all the slices as part of a volume. In such context, the back-projection in Eq. (21) can be rewritten as a volumetric back-projection:

$$E(s', t', z') = \iint_{-U,-V}^{+U,+V} P_{u,v}(s, t) \, \delta\left(\frac{z_0}{z'}s + \left(1 - \frac{z_0}{z'}\right)\frac{u}{M} - s', \frac{z_0}{z'}t + \left(1 - \frac{z_0}{z'}\right)\frac{v}{M} - t'\right) du \, dv,$$

$$(22)$$

$$L(s', t', u', v') = \int_{-Z}^{+Z} E(s, t, z) \, \delta\left(\frac{z}{z_0}s + \left(1 - \frac{z}{z_0}\right)\frac{u'}{M} - s', \frac{z}{z_0}t + \left(1 - \frac{z}{z_0}\right)\frac{v'}{M} - t'\right) dz,$$

$$(23)$$

where we have substituted α with its definition $\alpha = z/z_0$ and we have reported the forward-projection model that describes the ray tracing based image formation process. The forward projection is composed by the line integrals along the rays defined by Eqs. 52-(53), over the range $[-Z, +Z]$.

Equations 22-(23) give the relationship between the focal stack and the light field $\mathcal{L}(s, t, u, v)$, which can be used to reconstruct the slices all together. This corresponds to performing a tomographic reconstruction of the focal stack, as opposed to a refocusing reconstruction of the single slices. When referring to a volumetric reconstruction of the scene, the reconstruction pixels are called voxels. The advantage of volumetric reconstructions is the ability to distinguish the features from the different slices.

A tomographic reconstruction in this setting is characterized by a strongly anisotropic resolution of the reconstruction volume, due to the intrinsic limited angle nature of the problem. Here we use a Fourier-space representation of the acquired data, that is well understood in the tomography community, and which allows to deduce the depth and lateral resolutions achievable in a volumetric reconstruction. This three-dimensional Fourier representation was first introduced for light field images as a way to reduce their memory footprint, and it was called a "Radon image" [11].

As an example, we consider the acquisition geometry of the *Tarots Cards* dataset from the Stanford light field archive (http://lightfield.stanford.edu/lfs.html). The

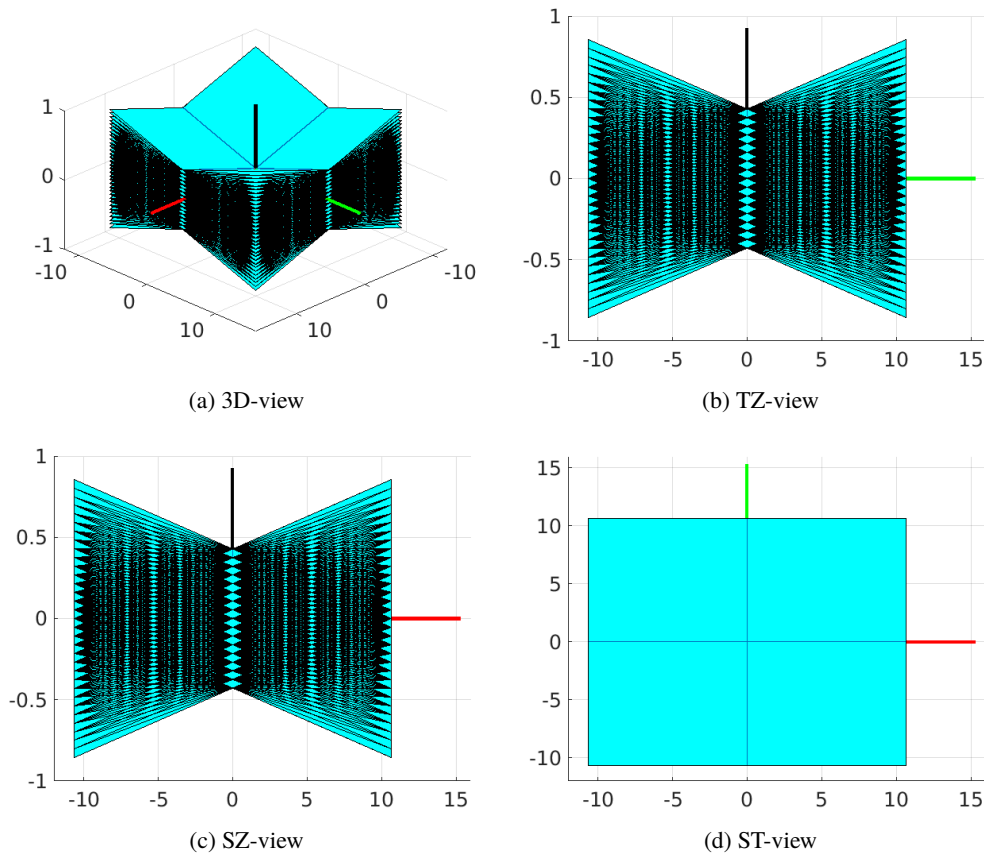(a) 3D-view

(b) TZ-view

(c) SZ-view

(d) ST-view

Fig. 6. Plot of the three-dimensional Fourier sampling associated with the Tarots Cards light field image from the Stanford light field archive.

sampling of the three-dimensional Fourier-space corresponding to this light field acquisition [10] is depicted in Fig. 6. This sampling pattern is very similar to the one of a very limited angle laminography or a tomosynthesis setup. We can also observe that, for the specific case in Fig. 6, the highest achievable resolution in the z (depth) direction is more than ten times lower than the lateral resolution.

These resolution considerations can raise some concerns related to the correct intensity determination of the reconstructed pixels. In fact, with such imbalanced anisotropic resolutions, reconstructing sharp features, that are not aligned with the sampling grid in the depth direction, results in smeared reconstructions. This effect is similar to the blur observed for the refocusing of features outside of their focus plane.

If we assume that an object resides at a certain distance z_0' from the camera, and that we refocus at a distance z_0'', the observed blur can be modeled as a box filter smoothing [15]. As discussed in [15], the width of the smoothing kernel scales linearly with both the distance $\Delta z = |z_0'' - z_0'|$ and the aperture sizes $2U$ and $2V$, in the s and t coordinates respectively:

$$w_s = \frac{2U}{|M|\delta s} \frac{\Delta z}{z_0'}, \tag{24}$$

$$w_t = \frac{2V}{|M|\delta t} \frac{\Delta z}{z_0'}, \tag{25}$$

where $|M|$ is the modulus of the magnification, w_s and w_t are the widths of the box filter in the direction of the s and t coordinates respectively, and δs and δt the corresponding voxel sizes. As a consequence, in a traditional refocusing reconstruction, an *extended depth-of-field* reconstruction can be obtained by selecting the sharpest features along the z-axis, out of a densely sampled focal stack [1].

A refocusing reconstruction is able to produce sharp images at the correct position along z, but it does not benefit from the interplay of the different slices. On the other hand, a volumetric reconstruction, by reconstructing the scene all together, can separate features from the different slices. However, if we were to oversample a volumetric reconstruction along the z-axis, another type of smearing artifact would appear along this direction. This blur is the typical effect of a so-called missing wedge in tomographic reconstructions, and the shape of the smoothing kernel along the z-axis can be approximated by a triangle. In terms of voxels, the width of this kernel is inversely proportional to the voxel size Δz (along the z-axis), and to half of the aperture sizes $2U$ and $2V$, in the s and t coordinates respectively:

$$w_{z,s} \propto \frac{\delta \text{obj}_s}{U} \frac{\delta z}{\Delta z}, \tag{26}$$

$$w_{z,t} \propto \frac{\delta \text{obj}_t}{V} \frac{\delta z}{\Delta z}. \tag{27}$$

Here $w_{z,s}$ and $w_{z,t}$ are s and t components of the triangular smoothing kernel width along z, δz is the voxel size along z, and δobj_s and δobj_t are the lateral sizes of the considered object along the s and t coordinates respectively.

For the aforementioned reasons, both the refocusing and the volumetric reconstructions present advantages and drawbacks, so a mixture of the two approaches could be used to sharply reconstruct multiple objects in the photographed scene.

As a final remark, the blur generated by the anisotropic voxel sizes in the volumetric reconstructions can be alleviated by increasing the aperture sizes $2U$ and $2V$. In compound plenoptic cameras this translates into an increased main lens diameter, while in gantry setups, this is implemented by taking the sub-aperture images on a wider range.

2.7. Depth estimation and impact of the aperture size

As mentioned in section 2.6, the refocusing reconstructions are based on the assumption that the full object resides at a certain fixed distance from the camera. This can deliver sharp and quantitatively correct image reconstructions of the photographed scene, when refocused at the correct distance z'_0. Depth estimation techniques take advantage of this piece of information, through the use of defocus and correspondence methods [3]. From Eqs. 24-(27), we derived that the aperture size of the main lens plays an important role in this context.

The depth-of-field of a refocused image is the distance along the optical axis that is still in focus in such image [15]. The smaller the depth-of-field, the higher the depth estimation resolution. Similarly to the volumetric reconstruction, the depth-of-field of a single refocused image is determined by the highest observable frequency in the z direction.

By using the reconstruction model in the object space, developed in section 2.5, it is possible to determine by geometric arguments the size of the depth-of-field of a refocused image:

$$\text{DoF}_0, s = \frac{2|M|\delta s z'_0}{U}, \tag{28}$$

$$\text{DoF}_0, t = \frac{2|M|\delta s z'_0}{V}, \tag{29}$$

where M is the magnification, $\text{DoF}_{0,s}$ and $\text{DoF}_{0,t}$ are the contributions to the depth-of-field from the s and t coordinates respectively, and z'_0 the position of the analyzed object along the z-axis.

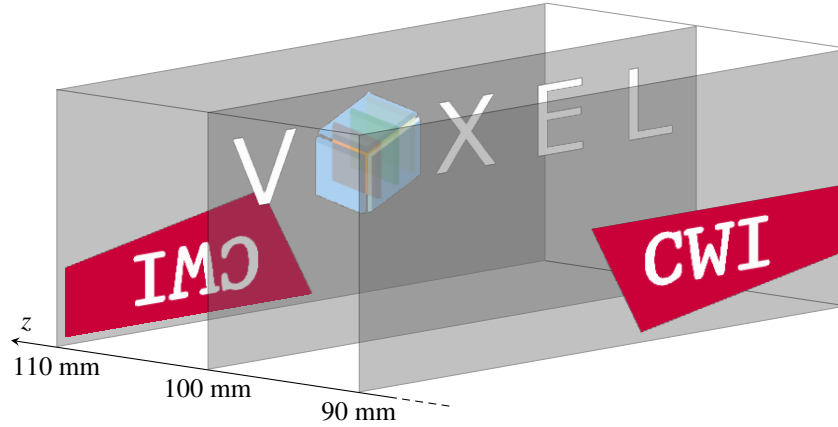For details on the derivation of Eqs. 28-(29) we refer to appendix E.

Fig. 7. Phantom used to simulate light field data for the logos scene: a VOXEL logo at the acquisition focal distance $z = 100$ mm, and two CWI logos at the distances of 110 mm, and 90 mm.
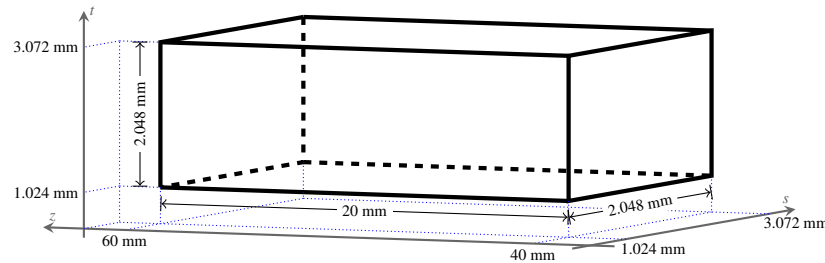

Fig. 8. Phantom of the wireframe scene: a cube centered on the acquisition focal distance 100 mm, which spans multiple refocusing planes.

3. Experiments

In this section, we present a series of experiments to illustrate the effect of the proposed methods on the scene reconstruction from light field data. We mostly present experiments based on simulated data, because they allow us to completely control and manipulate both the scene and the acquisition parameters. Finally, we add a real world example, which recreates similar conditions to the simulated examples, to show the behavior of the different models on real data.

3.1. Synthetic data description

We use two test cases, which both simulate the acquisition of a light field through the use of a compound plenoptic camera, but with different geometries and different scenes. The two scenes are referenced through the text as *logos* and *wireframe* test case, respectively. The logos scene is composed of three different objects at the three distances of 90, 100 and 110 mm each from the main lens, as displayed in Fig. 7. The wireframe scene is composed by the edges of a parallelepiped, and it serves two purpouses: it is used to show the combined effect of the aspects analyzed using the logos scene, on a three-dimensional structure that spans multiple planes, and it takes into account both the effects of noise corruption of the recorded images and the misalignment of the main lens. The wireframe case spans the space between 30 mm and 70 mm from the main lens (Fig. 8), but the wireframe structure only spans from 40 mm to 60m and
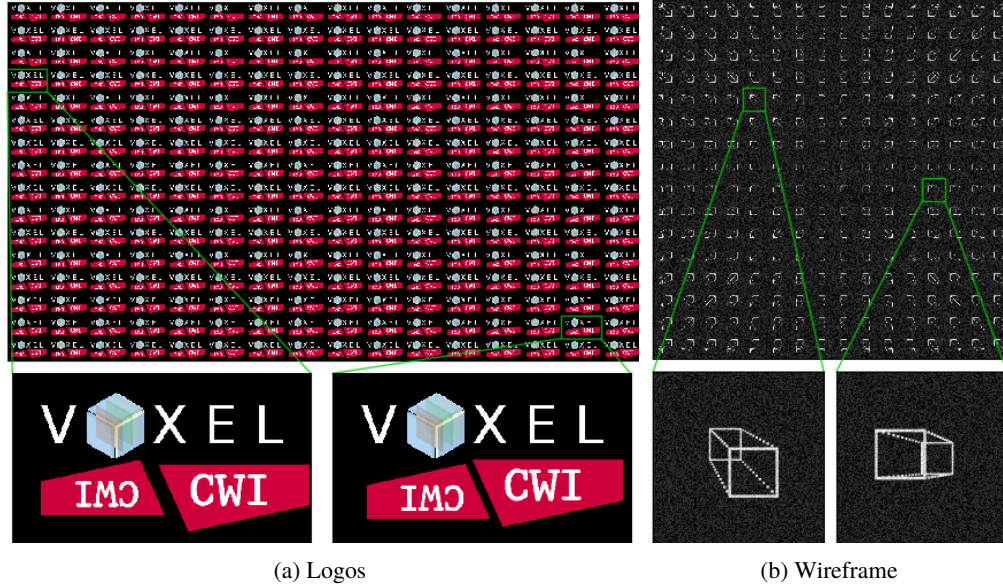
(a) Logos   (b) Wireframe

Fig. 9. Simulated light fields for both the synthetic logos and wireframe cases. Two sub-aperture images in each each case are extracted for clarity.

it has lateral sizes of 2.048×2.048 mm. The choice of such simple structure is based on the fact that its constituting elements can be easily segmented, identified, and measured. The tilt of the main lens results in a diagonal shift to the pin-hole positions of up to 10% of the (u, v) resolution, and the white noise amplitude goes to a maximum of 30% of the recorded wireframe signal amplitude.

The datasets are composed of 16×16 sub-aperture images, each having 512×256 pixels for the logos and 64×64 pixels for the wireframe scenes. The sub-aperture images for both cases are presented in Fig. 9.

Each lenslet has a diameter of 16μm, and in the examples of sections 3.3 and 3.4 the diameter of the main lens is 16 mm, with f_1 equal to 20 mm for the logogs case and 16.667 mm for the wireframe case. In section 3.6, the size of the aperture is one of the investigation parameters. The distance z_2 is 50μm, and the detector pixel size 1μm.

The acquisition distance of the MLA from the main lens z_1 is 25 mm for both the scenes, so that the focal plane in the object space is at 100 mm ($|M| = 4$) for the logos case, and 50 mm ($|M| = 2$) for the wireframe case.

The light fields produced for the logos and wireframe cases are shown in Figs. 9(a) and 9(b), respectively.

3.2. Real data description

This test case is based on an open setup, visible in Fig. 10(a). As MLA we used a HASO3 128GE2 from Imagine Optic, with 128×128 lenslets, each having a diameter $\sim 110 \, \mu$m. For the main lens we used an AC254-200-A-ML from Thorlabs, with f_1 of $200 \, mm$. The distance z_1 was set to $1000 \, mm$, resulting in a z_0 of $250 \, mm$ and $|M| = 0.25$, while z_2 was set to the value of f_2.

We used the 1951 USAF Resolution Test Target from Thorlabs to perform a series of image captures, at 11 evenly spaced distances z'_0, ranging from 240 mm till 260 mm, with an uncertainty of less than 1 mm. The image capture at distance z_0 equal to 250 mm is shown in Fig. 10(b).
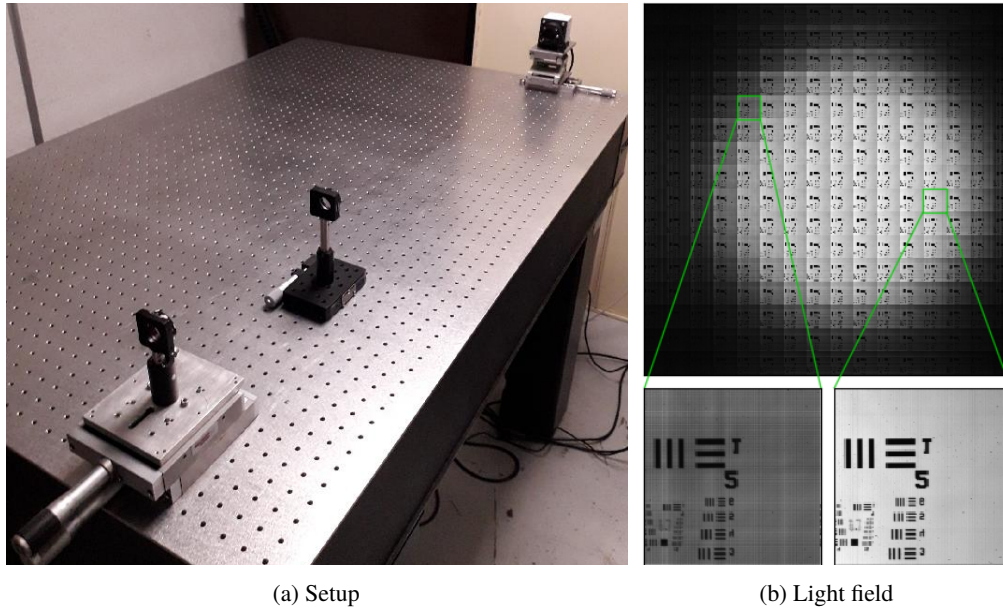
(a) Setup

(b) Light field

Fig. 10. Experimental setup and acquired light field for the real data experiment: (a) experimental setup, with the Thorlabs 1951 USAF Resolution Test target in the bottom left corner, the Thorlabs AC254-200-A-ML main lens in the middle, and the Imagine Optic HASO3 128GE2 in the top right corner, (b) the acquired light field for the distance $z_0' = z_0 = 250\,mm$.

3.3. Object space vs image space refocusing

In section 2.5 we addressed the difference between reconstructing the scene in the object space and the image space. A visual example of the considerations done in section 2.5 is given by the comparison of Figs. 11(b) and 11(d) against Fig. 11(a).

We expect the CWI logo in the bottom right corner to be at 90 mm from the main lens along the optical axis. This corresponds to $z_0 = 90$ mm and $\alpha_{o,c} = 0.9$. We see that when using the correct cone-beam geometry for the refocusing, if it is performed in the object space we obtain a refocused image at the expected $z_0 = 90$ mm ($\alpha_{o,c} = 0.9$). When it is performed in the image space, we obtain a refocused image at $z_1 = 24.37$ mm, corresponding to $\alpha_{i,c} = 0.9749$, and through the use of the thin lens equation, to $z_0 = 111.5$ mm.

Moreover, from Figs. 11(b) and 11(d), we can compare the position of the two points from the insets: the one on the upper left corner of the logo, and the one at the bottom of the letter "I" of the word CWI. Besides some blur, due to the refocusing based on the integration algorithm, the points in Fig. 11(b) correspond closely to the ones in Fig. 11(a). On the other hand, Fig. 11(d) presents a dilated CWI logo, and the positions of the identified points diverge from their real positions, as a function of their distance from the optical axis.

3.4. Cone-beam vs parallel-beam refocusing

Here we address the problem of the parallel-beam approximation, which was treated in section 2.4. We do it in a similar fashion to the previous section, by comparing the results from Figs. 11(b) and 11(c) against Fig. 11(a).

We focus again on the CWI logo in the bottom right corner. For the parallel-beam reconstruction we obtain a refocused image at $z_0 = 88.89$ mm and corresponding $\alpha_{o,p} = 0.8889$. This distance

(a) All-in-focus phantom

(b) Object space, cone-beam, $\alpha_{o,c} = 0.9$

(c) Object space, parallel-beam, $\alpha_{o,p} = 0.8889$

(d) Image space, cone-beam, $\alpha_{i,c} = 0.9749$

Fig. 11. Refocusing sizes and distances of the different geometric models: (a) Flattened all-in-focus phantom from Fig. 7, where the distances different from 90 mm have been made semi-transparent, (b) Refocusing in the object space with a cone-beam geometry, at a distance of $z_0 = 90$ mm, (c) Refocusing in the object space with a parallel-beam geometry, at a distance of $z_0 = 88.89$ mm, (d) Refocusing in the image space with a cone-beam geometry, at a distance $z_1 = 24.37$ mm, corresponding to a $z_0 = 111.5$ mm.

corresponds to the one computed theoretically using Eqs. 11-(12), from appendix C, with $\alpha_{o,c} = 0.9$.

When comparing the size of the investigated CWI logo, the parallel-beam geometry fails to retrieve the correct size when it is in focus. This observation is confirmed by the coordinate points underpinned in the insets of Figs. 11(b) and 11(c). Similarly to our observations from the previous example, the points in Fig. 11(c) diverge more from their real positions as a function of their distance from the optical axis.

3.5. Interpretation of distances and sizes

We now analyze the behavior of the three different models with respect to a varying distance along the optical axis of the "CWI" logo in the logos scene. As we observed in sections 3.3 and 3.4, under each parametrization returned different refocusing distances along the optical axis for the same object. These distances obey to the relationships of Eqs. 11-(12) and Eqs. 17-(18), respectively derived in appendices C and D.

The differences in scaling and distances for the different parametrizations are displayed in
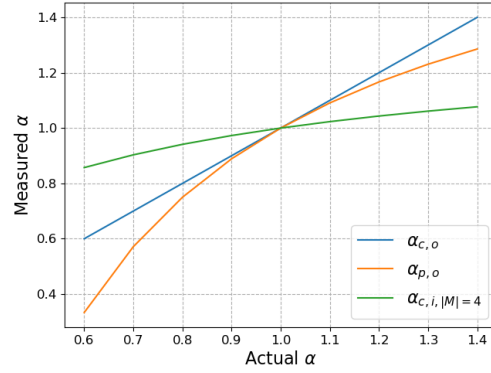
(a) Deviation in α
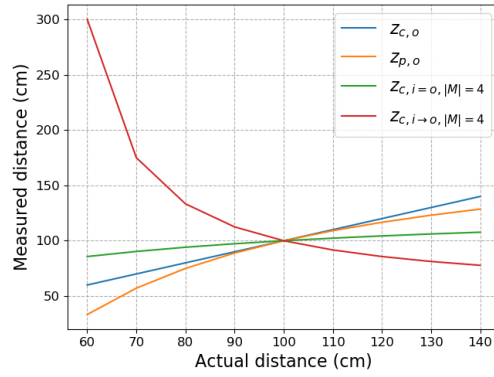
(b) Deviation in distance
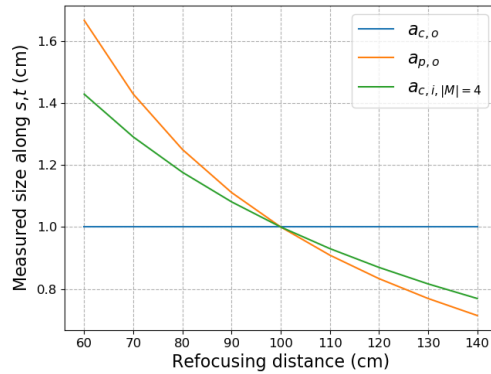
(c) Deviation in size

Fig. 12. Effect of the different parametrizations on the scaling of estimated size and depth, as a function of the object position along the optical axis, for the logos case. More precisely: (a) effect on the α parameters, (b) effect on the estimated distance along the z-axis, (c) effect on the estimated size along the s, t-axes. The chosen parametrizations were the three outlined in section 2.

Fig. 12. Figure 12(b) presents two different curves for the interpretation of the refocusing parameter $\alpha_{c,i,|M|=4}$, namely $z_{c,i=o,|M|=4}$ (green) and $z_{c,i\to o,|M|=4}$ (red). The first one corresponds to the case where we ignore the fact that the refocusing was performed in the image space, and multiply the obtained refocusing $\alpha_{c,i,|M|=4}$ with z_0 to obtain the refocusing distances $z_{c,i=o,|M|=4}$. The second one performs the same operation in the image space, obtaining the set of distances $\alpha_{c,i,|M|=4}z_1 = z_{c,i,|M|=4}$ and then converts them to the corresponding distances in the object space using the thin lens equation, defined as $1/f_1 = 1/z_o + 1/z_i$, where f_1 is the focal length of the main lens, z_i the distance in the image space, and z_o the corresponding distance in the object space [12]. By applying the thin lens equation to the distances $z_{c,i,|M|=4}$, we obtain the distances $z_{c,i\to o,|M|=4}$. The results for the two procedures are very different, but also both incorrect. The cone-beam object space parametrization always correctly retrieves the object distance (blue curve), while the parallel-beam object space parametrization always underestimates it.

In Fig. 12(c), we present the estimated normalized sizes, using the different parametrizations, when the same object is placed at different depths z. These results are consistent with the observations made from Fig. 11 and the discussions of sections 3.3 and 3.4. The estimated size of the objects using the cone-beam object space parametrization does not depend on the distance of such objects along the z-axis, while they follow different scaling laws for the other parametrizations.

3.6. Depth estimation and aperture size

We address in this section the aperture size problem mentioned in section 2.7, by following the state of the art methodology [3].

We perform depth estimation of the scene from Fig. 7 for three different aperture sizes, namely $d_1 = 2$ mm, $d_1 = 8$ mm, and $d_1 = 32$ mm. The corresponding depth maps are respectively displayed in Figs. 13(a) to 13(c). The defocus and correspondence response functions are used to identify the depth of the objects in the scene on a pixel per pixel basis [3].

For each position on the (s,t)-axes, a response value for both defocus and correspondence functions is associated to each sampled distance in the scene. Larger amplitude of these functions correspond to higher likelihood for the trial distances to correctly represent depth at the chosen position. Ideally, perfectly identified distances should be characterized by the defocus and correspondence functions presenting Dirac deltas at the right distances. Broader peaks, and the appearance of secondary peaks, can lead to incorrect identification of object distances in the scene.

In Fig. 13, we see that increasing the aperture size d_1 leads to an increase in quality of the defocus and correspondence response functions at the right positions in the depth map. As a consequence, we also observe an increased quality of the estimated depth maps.

These results are well correlated to the considerations exposed in section 2.7: increasing aperture sizes correspond to improved depth estimation resolutions. In fact, in the example of Fig. 13, all the three cases are sampled at the same rate along the z-axis, but we observe an increased ability to discriminate neighbouring distances with the defocus and correspondence response functions when using larger aperture sizes.

3.7. Wireframe reconstructions

This final example combines the quantitative aspects studied above, by analyzing the three-dimensional structure of a known object reconstructed with the different parametrizations. The refocusing reconstructions obtained for 21 distances between 40 mm and 70 mm from the main lens have been segmented and assembled to produce the final reconstructions presented in Fig. 14. The resulting volumes are displayed in the form of projections on the planes ST, SZ, and TZ.

For the depth estimation in the image space we used the $z_{c,i=o,|M|=4}$ type of interpretation from section 3.5 (green curve from Fig. 12(b)).
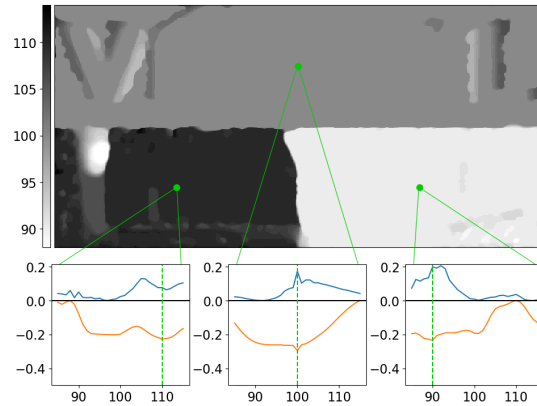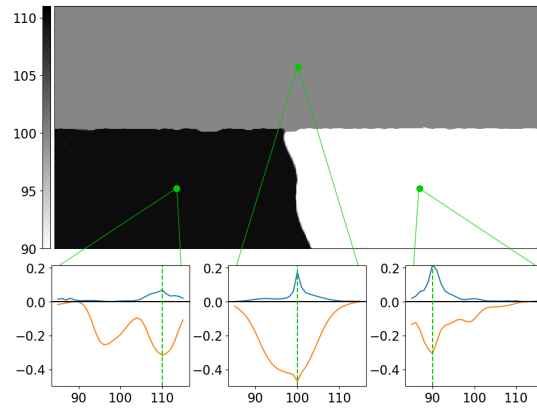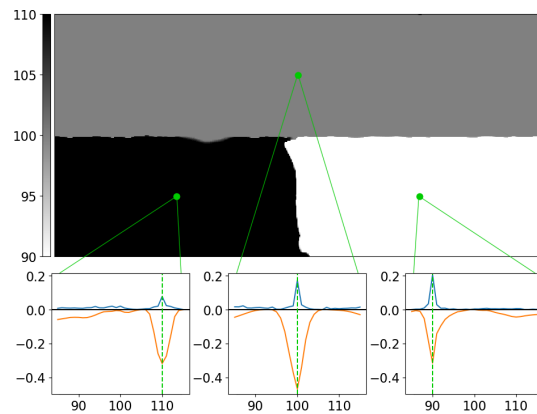
(a) $\mathrm{d}u, \mathrm{d}v = 0.125$ mm, $d_1 = 2$ mm

(b) $\mathrm{d}u, \mathrm{d}v = 0.5$ mm, $d_1 = 8$ mm

(c) $\mathrm{d}u, \mathrm{d}v = 2$ mm, $d_1 = 32$ mm

Fig. 13. Depth maps for three different aperture sizes and depth estimation cues response functions for three points in the depth map: (a) aperture size $d_1 = 2$ mm, (b) aperture size $d_1 = 8$ mm, (c) aperture size $d_1 = 32$ mm. The blue and orange curves in the function response plots correspond to the defocus and correspondence functions values for each of the sampled distances along the optical axis. The vertical green dotted line in the function response plots indicates the estimated depth for the given point.

(a) Phantom　　(b) Image space cone-beam　　(c) Object space cone-beam　　(d) Object space parallel-beam
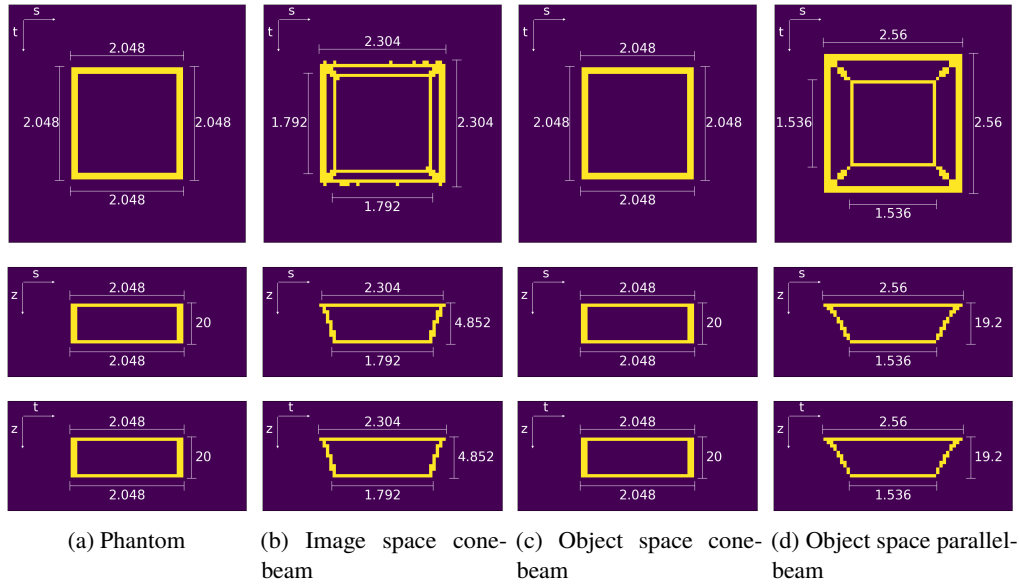
Fig. 14. Comparison of the reconstruction sizes of the wire-frame test case. The figures are projections on the ST, SZ, and TZ planes from top to bottom, and the columns represent: (a) the phantom, (b) reconstruction in the image space with cone-beam geometry, (c) reconstruction in the object space with cone-beam geometry, (d) reconstruction in the object space with parallel-beam geometry.

By comparing the rows of Fig. 14, we notice that the object space refocusing, using the cone-beam geometry, correctly reconstructs the shape and the size of the phantom in every direction. The other two parametrizations exhibit shape and size estimation errors. In these cases, the reconstruction of the parallelepiped returns a truncated pyramid, because both parameterizations tend to overestimate the size of objects when they are closer to the camera than the acquisition focal plane, and to underestimate their size when they are further away.

The cone-beam object space reconstruction reports a correct estimation of the depth of the sample (20 mm), and correct front and rear lateral sizes of 2.048 mm. The cone-beam image space reconstruction gives a much lower estimation of the depth of the sample (4.852 mm), and a slightly incorrect lateral size: front 2.304 mm, rear 1.792 mm. Finally, the parallel-beam object space reconstruction presents an underestimated depth of the sample of 19.2 mm, and also an incorrect lateral size: front 2.56 mm, rear 1.536 mm.

The reconstruction that suffered the most from noise is the one based on the image space cone-beam model, together with the worst estimation of the object length along the z-axis. The second problem should be attributed to a joint effect of the two following causes: the fact that in the image space the distances scale along the z-axis with respect to z_1 instead of z_0, and the fact that the ray geometry is from the main lens is increasingly different from the object space, due to the magnification.

3.8. Real data analysis

We estimated the distance of the target in each acquired image with the setup from Fig. 10(a), for the three models considered in section 3.5. We then segmented the refocused images and estimated the size of the target in number of pixels, from the left edge of the pattern in the top of Fig. 10(b), horizontally until the reversed "1" on the top right. The estimated size of such length is $\sim 1800 \, \mu m$, which corresponds to 62.6 pixels.
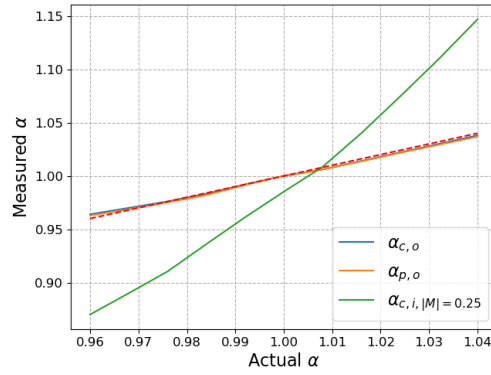
(a) Deviation in α
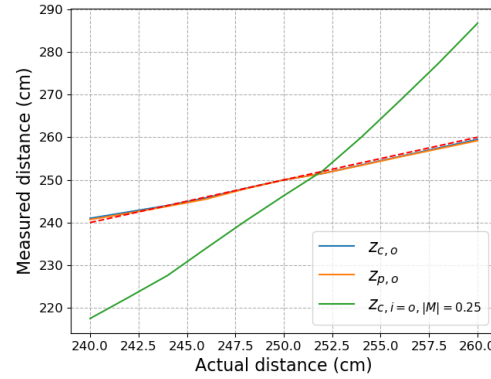
(b) Deviation in distance
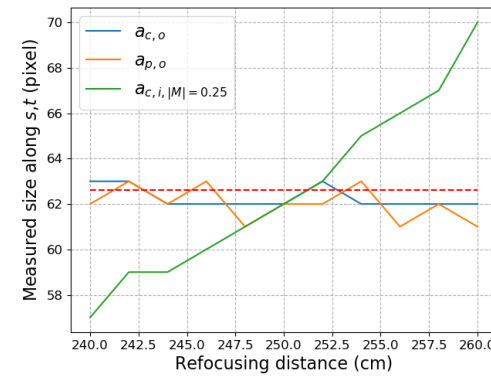
(c) Deviation in size

Fig. 15. Effect of the different parametrizations on the scaling of estimated size and depth, as a function of the object position along the optical axis, for the real-world example discussed in section 3.2. More precisely: (a) effect on the α parameters, (b) effect on the estimated distance along the z-axis, (c) effect on the estimated size along the s, t-axes. The chosen parametrizations were the three outlined in section 2. The red dotted line represents the expected value for the plotted quantity.

In Fig. 15, we plotted the estimated parameter α for the three models, together with the fitted target distance, and lateral size. For the chosen magnification of $|M| = 0.25$, and deviations in α, we notice only a small difference between the fitted distance between the cone-beam and parallel-beam approaches in the object space. This is in accordance with the fact that the expected difference between α_c and α_p is only a few percents at the investigated distance range. The difference is, instead, remarkably bigger between the cone-beam approach in object space, compared to the image space, as it can be seen clearly in Figs. 15(a) and 15(b).

For what concerns Fig. 15(c), we see the same trend, with differences between the cone-beam and parallel-beam approaches in object space giving size differences within the uncertainty of one pixel, while the cone-beam image space approach follows the same trend as for Figs. 15(a) and 15(b).

4. Discussion

From section 3 we can observe that all the different parametrizations are able to produce refocused images, but only the cone-beam object space parametrization can directly retrieve the correct depth and size of the photographed objects, based on the refocused images. This is not in contradiction with all the previous work in this field, which assumed a different parametrization of the reconstruction space. In fact, an important aspect of Eqs. 11-(12) and Eqs. 17-(18) is their ability to convert the distances computed with other parametrizations into the cone-beam object space parametrization proposed in this article. Given these equations, the conversion of estimated depths is straightforward:

$$z_0' = z_0 \alpha_{c,o} = z_0 \frac{\alpha_{c,i}}{(1 - |M|)\, \alpha_{c,i} + |M|}, \tag{30}$$

$$z_0' = z_0 \alpha_{c,o} = z_0 \frac{1}{2 - \alpha_{p,o}}, \tag{31}$$

where z_0' is the correct object distance in the scene, z_0 is the acquisition focal distance in the object space, and $\alpha_{c,i}$ and $\alpha_{p,o}$ are the refocusing αs for the cone-beam image space and parallel-beam object space parametrizations respectively. Concerning the size estimation, it is easily seen from Eqs. 9-(10) and Eqs. 15-(16) that the coordinates have to be scaled accordingly to the cone-beam object space parametrization. This results in:

$$s_0' = s_{c,i}' \frac{\alpha_{c,o}}{\alpha_{c,i}}, \tag{32}$$

$$s_0' = s_{p,o}' \alpha_{c,o}, \tag{33}$$

where s_0' is the correct object size in the scene, $s_{c,i}'$ and $s_{p,o}'$ are the estimated object sizes by the cone-beam image space and parallel-beam object space parametrizations respectively.

5. Conclusions and outlook

In this article we have used the transformation between the ray tracing geometry of a plenoptic camera and a cone-beam tomographic setup to describe plenoptic photography problems as tomographic problems. We then used it to establish an accurate and precise reconstruction framework, enabling quantitative analysis based on light field images.

We showed that many aspects of the scene reconstruction from a light field image can be understood in simple terms with the use of tomographic tools. This allowed us to identify the most suitable representation of the data to perform the correct estimation of the photographed object sizes and positions. We deduced that the scene reconstruction should be performed in the object space, assuming the cone-beam projection geometry of a tomographic reconstruction. However, existing tools can still be used without modification, provided that their output is

interpreted with the relationships identified in this article. In fact, we also obtained simple formulas to convert the results from alternative parametrizations into the results obtainable with the cone-beam object space parametrization proposed here.

A final important consequence of the explicit transformation from light field images to limited angle cone-beam tomography geometries is the ability to use the advanced algorithms from the tomography world for the reconstruction of higher quality refocused images. This means that recent developments in the tomographic reconstruction algorithms and theory [16–19] can be directly applied to light field photography, as it was done in [9] for light field microscopy applications.

Appendices

A. Radiance

We derive here the functions describing the intensity deposited on the detector pixels for three setups: the two plenoptic imaging setups, namely the compound camera and camera arrays, and the cone-beam tomography setup. A similar approach was taken in [20], to derive the spatial and directional resolutions of the plenoptic camera. In that case, the authors focus on the intensity deposited by the footprint of the lenslets on the detector. Here instead, we take the perspective of the pin-hole images, and look at the corresponding pixels for only one sub-aperture image. In the case of a compound plenoptic camera we look at the intensity coming from the corresponding region on the main lens. For the camera array setup we look at one specific camera of the array. These two subsets of rays compare to the rays of one specific projection in the tomography setup.

In Fig. 16(b), we present again the traditional compound plenoptic camera setup from Fig. 4, while in Fig. 16(a) we present the camera arrays. In both figures, at the crossing point P of the rays over the axis u, the radiance can be described by a function of the coordinates $\mathbf{x} = (q, p)$ where p represents the rays direction and q represents the ray position. This function is completely separable in the two coordinates, and it is expressed as

$$r_P(\mathbf{x}) = r_P(q, p) = \delta_0(q) F(p), \tag{34}$$

where δ_0 is a Dirac-delta centered on the point P, which for convenience is the 0 of the coordinates, and $F(p)$ is the directional component of the radiance captured by the sub-aperture image under analysis.

The transformations that describe the propagation of light from such point to the detector in Fig. 16(a) and to the lenslets centers in Fig. 16(b) are represented by 2×2 matrices:

$$\mathbf{A}_{\text{lensless}} = \mathbf{T}_{z_1} = \begin{pmatrix} 1 & z_1 \\ 0 & 1 \end{pmatrix}, \tag{35}$$

$$\mathbf{A}_{\text{lensless}}^{-1} = \mathbf{T}_{z_1}^{-1} = \mathbf{T}_{-z_1} = \begin{pmatrix} 1 & -z_1 \\ 0 & 1 \end{pmatrix}, \tag{36}$$

$$\mathbf{A}_{\text{compound}} = \mathbf{L}_{f_1} \mathbf{T}_{z_1} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f_1} & 1 \end{pmatrix} \begin{pmatrix} 1 & z_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & z_1 \\ -\frac{1}{f_1} & 1 - \frac{z_1}{f_1} \end{pmatrix}, \tag{37}$$

$$\mathbf{A}_{\text{compound}}^{-1} = \mathbf{T}_{z_1}^{-1} \mathbf{L}_{f_1}^{-1} = \begin{pmatrix} 1 & -z_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{1}{f_1} & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{z_1}{f_1} & -z_1 \\ \frac{1}{f_1} & 1 \end{pmatrix}, \tag{38}$$

where the matrices $\mathbf{T}_\bullet$ represent the propagation of the rays along straight lines, and the matrix $\mathbf{L}_{f_1}$ represents the effect of the main lens. As also discussed in [20] and [21], we transform the
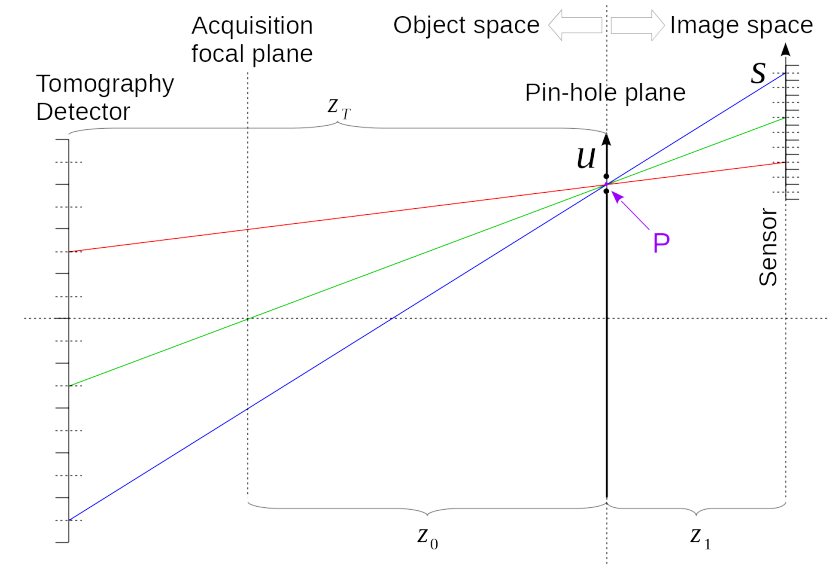
(a) Lensless
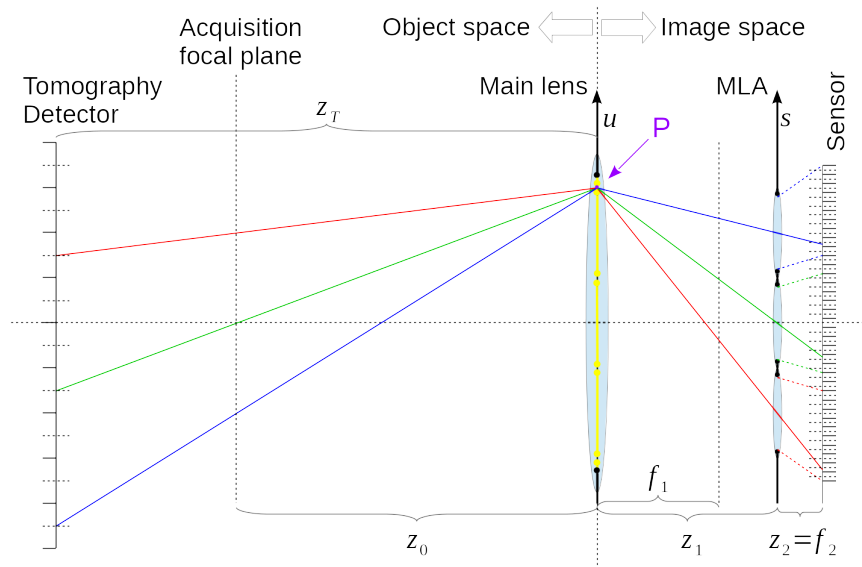
(b) With lenses

Fig. 16. Ray tracing geometry for one sub-aperture image of both the plenoptic acquisition setups.

radiance from the point P to the points observed by the detector in the lensless case and by the lenslets with the following equivalences:

$$r_{\text{MLA}}(\mathbf{x}) = r_P\left(\mathbf{A}_{\text{compound}}^{-1}\mathbf{x}\right) = \delta_0\left(q - z_1(q\frac{1}{f_1} + p)\right)F\left(q\frac{1}{f_1} + p\right), \tag{39}$$

$$r_{\text{D}}(\mathbf{x}) = r_P\left(\mathbf{A}_{\text{lensless}}^{-1}\mathbf{x}\right) = \delta_0\left(q - z_1 p\right)F\left(p\right), \tag{40}$$

For the lensless case in Fig. 16(a), the intensity collected by the detector pixels is:

$$I_{\text{D}}(q) = \int_{-\infty}^{+\infty} r_{\text{D}}(q, p)\,\mathrm{d}p = \int_{-\infty}^{+\infty} \delta_0(q - z_1 p)\,F(p)\,\mathrm{d}p = \frac{1}{z_1}F\left(\frac{q}{z_1}\right), \tag{41}$$

which means that on the detector we record a dilated version of the directional component $F(p)$ of the radiance $r_P(\mathbf{x})$. For what concerns the case in Fig. 16(b), the intensity deposited at each lenslet position is computed by the same integral:

$$\begin{aligned} I_{\text{MLA}}(q) &= \int_{-\infty}^{+\infty} r_{MLA}(q, p)\,\mathrm{d}p \\ &= \int_{-\infty}^{+\infty} \delta_0\left(q - z_1(q\frac{1}{f_1} + p)\right)F\left(q\frac{1}{f_1} + p\right)\mathrm{d}p, \end{aligned} \tag{42}$$

where this time we have to operate a change of variables to reduce ourselves to the same condition of Eq. (41):

$$p' = q\frac{1}{f_1} + p, \ \mathrm{d}p' = \mathrm{d}p \tag{43}$$

$$I_{\text{MLA}}(q) = \int_{-\infty}^{+\infty} r_{\text{D}}(q, p)\,\mathrm{d}p = \int_{-\infty}^{+\infty} \delta_0(q - z_1 p')\,F(p')\mathrm{d}p' = \frac{1}{z_1}F\left(\frac{q}{z_1}\right), \tag{44}$$

which proves that the two different light field setups record the same type of sub-aperture images.

We can obtain the same expression for the tomography case, where the point P is now the source, and the tomography detector lives in the object space. The radiance at the point P is the same from Eq. (34), and the transformation matrices for the coordinate change from the point P to the tomography detector are:

$$\mathbf{A}_{\text{tomography}} = \mathbf{T}_{z_{\text{T}}} = \begin{pmatrix} 1 & -z_{\text{T}} \\ 0 & 1 \end{pmatrix}, \tag{45}$$

$$\mathbf{A}_{\text{tomography}}^{-1} = \mathbf{T}_{z_{\text{T}}}^{-1} = \mathbf{T}_{-z_{\text{T}}} = \begin{pmatrix} 1 & z_{\text{T}} \\ 0 & 1 \end{pmatrix}, \tag{46}$$

where the propagation happens in the opposite direction with respect to the other two cases, and the sign of the propagation is reversed. This translates to:

$$r_{\text{T}}(\mathbf{x}) = r_P\left(\mathbf{A}_{\text{tomography}}^{-1}\mathbf{x}\right) = \delta_0\left(q + z_{\text{T}} p\right)F\left(p\right), \tag{47}$$

$$I_{\text{T}}(q) = \int_{-\infty}^{+\infty} r_{\text{T}}(q, p)\,\mathrm{d}p = \int_{-\infty}^{+\infty} \delta_0(q + z_{\text{T}} p)\,F(p)\,\mathrm{d}p = \frac{1}{z_{\text{T}}}F\left(-\frac{q}{z_{\text{T}}}\right), \tag{48}$$

where the directional component $F(p)$ of the radiance $r_P(\mathbf{x})$ is recorded upside-down compared to the light field case, with a different magnification.

Equations (41), (44) and (48) show the close relationship between the light field and the tomography geometry. Besides a flip of the image and different magnification factors, a sub-aperture image in a light field acquisition can be exactly transformed into one image of a cone-beam tomography acquisition.
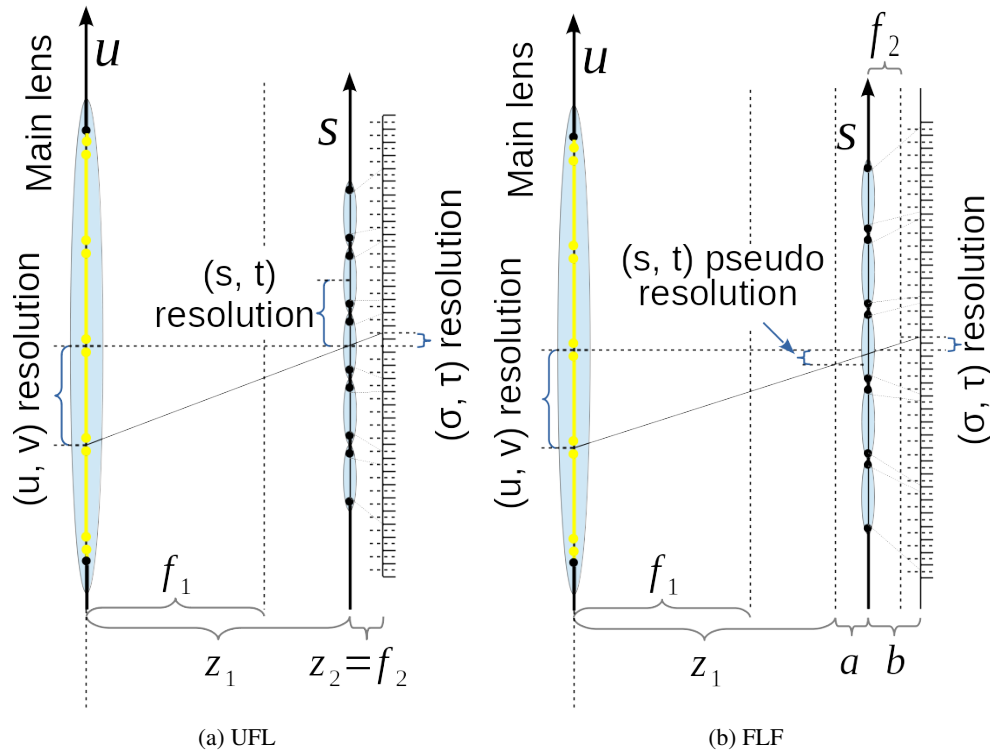
Fig. 17. Comparison of the (s, u, σ) resolutions in the: (a) unfocused light field camera, (b) focused light field camera. In the FLF case, we speak about pseudo-resolution, due to the sheared sampling of Fig. 3(b).

B.   ULF and FLF phase space sampling

The ULF and FLF phase space sampling can be derived by geometric construction, following the ray-tracing carried out in Fig. 17.

In the ULF case, the (s, t) axes reside on the MLA plane and the (s, t) sampling is therefore imposed by the lenslet positions. The (u, v) coordinates corresponding to each pixel at position (σ, τ) are computed by simply prolonging the ray originating from the said pixel and traversing the corresponding lenslet center (chief ray) to the main lens plane. This results in Eqs. 1-(4).

In the FLF case, we follow the same procedure as for the ULF. For the (u, v) coordinates this results in just a slightly different version, compared to the ULF case, which accounts for the distances $z_1 + a$ or $z_1 - a$, depending on the type of focused configuration. In Eqs. 5-(8) we chose to represent the first of the two possibilities. For what concerns the (s, t) coordinates, for a given lenslet each pixel (σ, τ) has an additional shift of $\sigma(a/b)$ and $\tau(a/b)$ to the s and t coordinates of the lenslet, respectively. This results in Eqs. 5-(8).

The additional (s, t) shift in Eqs. 5-(6) is responsible for the higher spatial quality available in the FLF camera configurations. In Fig. 17, we denoted such shift as the FLF (s, t) pseudo resolution. In fact, the number of spatial samples in the (s, t, u, v) phase space is identical for both the ULF and FLF configurations, but the sheared acquisition geometry, embodied by such shift, contributes to higher spatial resolution in image refocusing.

C. Refocusing distances for cone-beam and parallel-beam geometries

We now address the difference in the α parameter between the cone-beam and parallel-beam geometry. We define the set of functions describing the intensity patter for each sub-aperture image with positions (u, v) on the main lens, as the set of functions $P_{u,v}$. Such functions are hyperplanes in the light field $\mathcal{L}(s, t, u, v)$ at the positions (u, v). We can then rewrite Eq. (20):

$$E_{(z')}(s', t') = \frac{1}{z^2} \sum_u \sum_v P_{u,v}\left(\frac{s}{\alpha} + \left(1 - \frac{1}{\alpha}\right)u, \frac{t}{\alpha} + \left(1 - \frac{1}{\alpha}\right)v\right), \tag{49}$$

where we have substituted the integral with the sum of the contributions from each sub-aperture image. To have correct refocusing, the functions $P_{u,v}$ need to have matching intensities for a certain α. This means that given any set $C = \{(s_{u_1}, t_{v_1}), (s_{u_2}, t_{v_2}), \dots, (s_{u_U}, t_{v_V})\}$ of initial coordinates on the focal plane of the image space, they have to match at the refocusing positions $(s'_{u_1}, t'_{v_1}) = (s'_{u_2}, t'_{v_2}) = \dots = (s'_{u_U}, t'_{v_V})$.

Let us now focus on the first two pairs of coordinates, for the cone-beam geometry. From Eq. (9) for the s coordinate, we have $s'_{u_1} = s'_{u_2} = \alpha_c s_{u_1} + (1 - \alpha_c)u_1 = \alpha_c s_{u_2} + (1 - \alpha_c)u_2$. This translates into:

$$\alpha_c = \frac{u_1 - u_2}{(u_1 - u_2) - (s_{u_1} - s_{u_2})}, \tag{50}$$

while for the coordinate transformation of the parallel-beam geometry, we obtain:

$$\alpha_p = \frac{(u_1 - u_2) + (s_{u_1} - s_{u_2})}{u_1 - u_2} = 1 + \frac{s_{u_1} - s_{u_2}}{u_1 - u_2}. \tag{51}$$

By substituting Eq. (51) in Eq. (50), we obtain the two relationships in Eqs. 11-(12).

From Eqs. 11-(12) we deduce that the two parameters α_c and α_p can only be equal in the case where $\alpha_c = \alpha_p = 1$.

D. Ray geometry in object space and image space

In this appendix, we show that the images reconstructed in the object space differ from the corresponding refocused images in the image space. We follow a similar rationale as the one outlined in appendix C.

We rewrite Eqs. 15-(16) as the following:

$$s'_o = \alpha s_o + (1 - \alpha)\frac{u}{M}, \tag{52}$$

$$t'_o = \alpha t_o + (1 - \alpha)\frac{v}{M}, \tag{53}$$

which can be substituted in Eq. (49), to obtain the corresponding intensity function in the object space for the same recorded sub-aperture images:

$$E_{(z'_o)}(s'_o, t'_o) = \frac{1}{z_o^2} \sum_u \sum_v P_{u,v}\left(\frac{s_o}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\frac{u_o}{M}, \frac{t_o}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\frac{v_o}{M}\right), \tag{54}$$

in which case again, to have correct refocusing, the functions $P_{u,v}$ need to match for a certain α. In the image space, we have exactly the same procedure as outline in appendix C, for the cone-beam case. For clarity, we now indicate the set of initial coordinates on the focal plane of the image space as $C_i = \{(s_{i,u_1}, t_{i,v_1}), (s_{i,u_2}, t_{i,v_2}), \dots, (s_{i,u_U}, t_{i,v_V})\}$. The matching coordinates at the refocusing positions are instead: $(s'_{i,u_1}, t'_{i,v_1}) = (s'_{i,u_2}, t'_{i,v_2}) = \dots = (s'_{i,u_U}, t'_{i,v_V})$. Let us now

focus again on the first two pairs of coordinates. From Eq. (9), for the coordinate s we have $s'_{i,u_1} = s'_{i,u_2} = \alpha_i s_{i,u_1} + (1 - \alpha_i)u_1 = \alpha_i s_{i,u_2} + (1 - \alpha_i)u_2$. This translates into

$$\alpha_{c,i} = \frac{-(u_1 - u_2)}{\left(s_{i,u_1} - s_{i,u_2}\right) - (u_1 - u_2)} = \frac{(u_1 - u_2)}{(u_1 - u_2) - \left(s_{i,u_1} - s_{i,u_2}\right)}, \tag{55}$$

where we only take the relationship along the s coordinate and where we assume a cone-beam geometry.

On the focal plane in the object space we have the same coordinates $(s_{o,u_1}, t_{o,v_1}) = (s_{i,u_1}, t_{i,v_1})$, $(s_{o,u_2}, t_{o,v_2}) = (s_{i,u_2}, t_{i,v_2})$, etc. For the focus condition to be verified, they also have to match at the focusing position, meaning that $(s'_{o,u_1}, t'_{o,v_1}) = (s'_{o,u_2}, t'_{o,v_2}) = \ldots = (s'_{o,u_U}, t'_{o,v_V})$. However, the coordinate change in the object space follows Eq. (52), resulting in $\alpha_o = (-u_1 + u_2)/\left(M\left(s_{i,u_1} - s_{i,u_2}\right) - u_1 + u_2\right)$. If we now substitute $(s_{o,u_1}, t_{o,v_1}) = (s_{i,u_1}, t_{i,v_1})$, …, $(s_{o,u_U}, t_{o,v_V}) = (s_{i,u_U}, t_{i,v_V})$, we obtain

$$\alpha_{c,o} = \frac{-(u_1 - u_2)}{M\left(s_{i,u_1} - s_{i,u_2}\right) - (u_1 - u_2)} = \frac{(u_1 - u_2)}{(u_1 - u_2) - M\left(s_{i,u_1} - s_{i,u_2}\right)}, \tag{56}$$

which shows that in general $\alpha_{c,o} \neq \alpha_{c,i}$ for $M \neq 1$. As a consequence, the refocused image dilation is different for the two images, and it means that images reconstructed in the image space are not correct reconstructions of the scene in the object space.

We can now proceed to the computation of the relationship between $\alpha_{c,o}$ and $\alpha_{c,i}$, by taking their inverse:

$$\frac{1}{\alpha_{c,o}} = 1 - M \frac{\left(s_{i,u_1} - s_{i,u_2}\right)}{(u_1 - u_2)}, \tag{57}$$

$$\frac{1}{\alpha_{c,i}} = 1 - \frac{\left(s_{i,u_1} - s_{i,u_2}\right)}{(u_1 - u_2)}, \tag{58}$$

which then translates to:

$$\alpha_{c,o} = \frac{\alpha_{c,i}}{(1 - M)\,\alpha_{c,i} + M}, \tag{59}$$

$$\alpha_{c,i} = \frac{M\alpha_{c,o}}{1 - \alpha_{c,o}\,(1 - M)}, \tag{60}$$

where we see that it is only possible to have $\alpha_{c,o} = \alpha_{c,i}$ for $M = 1$. A minor complication is raised by the sign of the magnification M. Positive magnifications correspond to transformations in the same space (image to image, and object to object), while negative magnifications correspond to transformations across spaces (image to object, and object to image). Ideally, we would want to use the negative magnification, but the parametrization in the image space proposed in [1] assumes a mirrored representation of what happens in the object space. To preserve this mirroring, we have to use the modulus of the magnification M, transforming the relationships from Eqs. 59-(60) into Eqs. 17-(18), respectively.

As a final remark, this does not happen in the case of the parallel-beam approximation, because the rescaling of the coordinates (s, t) is removed, this means that we have the same α_p for refocusing in the image space and the object space.

E. Depth-of-field of a refocused image

Estimating the depth-of-field of an image can be done by identifying the z regions around it that have light-rays being focused by the main lens to the same detector pixels as the image. In traditional compound plenoptic cameras, this is equivalent to having the rays collimated on the

same microlens. While these considerations apply in the image space, for determining bounds on the depth estimation resolution, we need to translate the same concept into the object space. For this purpose, we use the model given by the parametrization of Eqs. 52-(53), and focus on the reconstruction pixels on the optical axis. For simplicity, we also only consider the problem along the s coordinate. Let us consider a refocusing distance z_o in the object space, and a reconstruction pixel size of $|M| \times \delta s$, where δs is the lenslet size along the s coordinate. Let us also consider the physical sampling bound on the u-axis, the quantity $u_{max} = U$, we have:

$$\tan \theta = \frac{U}{z_o}, \tag{61}$$

where θ is the angle between the optical axis and the line connecting the center of the reconstruction pixel with the largest sampled component on the u axis. A first order approximation for the depth-of-field is:

$$\text{DoF} = \frac{2|M|\Delta s}{\tan \theta}. \tag{62}$$

If we substitute Eq. (61) in Eq. (62), we obtain Eqs. 28-(29).

As expected, the depth-of-field increases as a function of distance from the camera, and it decreases with larger main lens aperture sizes.

Funding

Acknowledgments

References

1. R. Ng, "Digital light field photography," Ph.D. thesis, Stanford University (2006).
2. E. Y. Lam, "Computational photography with plenoptic camera and light field capture: tutorial," J. Opt. Soc. Am. A **32**(11), 2021–2032 (2015).
3. M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proceedings of IEEE International Conference on Computer Vision,* (IEEE, 2013), pp. 673–680.
4. C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Refocusing distance of a standard plenoptic camera," Opt. Express **24**(19), 21521–21540 (2016).
5. R. Ng, M. Levoy, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanf. Univ. Tech. Rep. CSTR 2005-02 (Stanford University, 2005).
6. A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *IEEE International Conference on Computational Photography (ICCP),* (IEEE, 2009), pp. 1–8.
7. S. Zhu, A. Lai, K. Eaton, P. Jin, and L. Gao, "On the fundamental comparison between unfocused and focused light field cameras," Appl. Opt. **57**(1), A1 (2018).
8. M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," ACM Trans. Graph. **25**(3), 924–934 (2006).
9. M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-D deconvolution for the light field microscope," Opt. Express **21**(21), 25418–25439 (2013).
10. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, 1988).
11. T. G. Georgiev, A. Lumsdaine, J. Gille, and A. Veeraraghavan, "The Radon image as plenoptic function," in *Proceedings of IEEE International Conference on Image Processing,* (IEEE, 2014), pp. 1922–1926.
12. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3 (McGraw-Hill, 1957), 3rd ed.
13. T. M. Buzug, *Computed Tomography* (Springer-Verlag, 2008).
14. W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)," J. Struct. Biol. **176**(2), 250–253 (2011).
15. L. G. Shapiro and G. C. Stockman, *Computer Vision* (Pearson, 2001).

16. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. on Imaging Sci. **2**, 183–202 (2009).

17. A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," J. Math. Imaging Vis. **40**(1), 120–145 (2010).

18. E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm," Phys. Medicine Biol. **57**(10), 3065–3091 (2012).

19. E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Commun. on Pure Appl. Math. **59**(8), 1207–1223 (2006).

20. A. Lumsdaine, T. G. Georgiev, and G. Chunev, "Spatial analysis of discrete plenoptic sampling," Proc. SPIE **8299**, 829909 (2012).

21. T. G. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," J. Electron. Imaging **19**(2), 021106 (2010).