



Advanced light-field refocusing through tomographic modeling of the photographed scene

NICOLA VIGANÒ,^{1,*} PABLO MARTÍNEZ GIL,² CHARLOTTE HERZOG,^{3,4,5} OMBELINE DE LA ROCHEFOUCAULD,³ ROBERT VAN LIERE,^{1,6} AND KEES JOOST BATENBURG^{1,7}

¹Centrum Wiskunde & Informatica (CWI), NWO, Amsterdam, The Netherlands

²Instituto Fusión Nuclear "Guillermo Velarde" (IFN-GV), ETSII, Universidad Politécnica de Madrid, 28006 Madrid, Spain

³Imagine Optic, Bordeaux, France

⁴Laboratoire Photonique, Numérique et Nanosciences UMR 5298: Institut d'Optique Graduate School, CNRS, Univ. Bordeaux, France

⁵Inria Bordeaux Sud-Ouest, Bordeaux, France

⁶Technical University Eindhoven, Eindhoven, The Netherlands

⁷Leiden Universiteit, Leiden, The Netherlands

*N.R.Vigano@cw.nl

Abstract: Recently we have shown that light-field photography images can be interpreted as limited-angle cone-beam tomography acquisitions. Here, we use this property to develop a direct-space tomographic refocusing formulation that allows one to refocus both unfocused and focused light-field images. We express the reconstruction as a convex optimization problem, thus enabling the use of various regularization terms to help suppress artifacts, and a wide class of existing advanced tomographic algorithms. This formulation also supports super-resolved reconstructions and the correction of the optical system's limited frequency response (point spread function). We validate this method with numerical and real-world examples.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Portable plenoptic cameras offer unique features which do not exist with traditional cameras, such as the possibility to refocus on different planes of the scene after acquisition [1, 2], and the ability to extract depth information regarding the photographed objects [3]. These features come at the cost of a greatly reduced spatial resolution. As a result, researchers and developers have invested a considerable effort on developing custom and sometimes very specialized algorithms for computational photography techniques, in particular for super-resolution applications.

Higher light-field refocusing quality and resolution are prerequisites for important applications that can benefit from the ability to refocus on multiple distances after acquisition, depth estimation and volume rendering, all from a single light-field. Examples of these applications can be found in fields ranging from industrial quality control and automation processes to autonomous driving. As light-field imaging will be increasingly adopted by these applications, greater importance will be given to quality and richness of details of the refocused images.

Many approaches attempt to obtain spatially super-resolved light-fields from lower resolution acquisitions [4–9]. The resulting super-resolved light-fields contain finer detailed structures that can deliver higher quality depth-estimation (distance from the camera) of the objects in the scene, and deliver higher quality image reconstruction with traditional refocusing algorithms. These approaches need strong *a priori* information to work properly, like a map of the distances from the camera of all the objects in the scene, or the knowledge of their composing structures, which limits their applicability.

A different approach is taken in [10] for a compound-eye imaging system, and in [11] for the light-field microscope, where the authors use the redundancy of the acquired data to super-resolve the refocused images, instead of creating new information in the data. In both [10] and [11], the image formation process is modelled as a linear process, and the high-resolution image reconstruction is formulated as the solution to a minimization problem. The limited frequency response of the optical system, modelled as a point spread function (PSF), is taken into account in the image formation model.

As specified in [11], the traditional light-field acquisition scheme, called unfocused light-field (ULF), does not allow to super-resolve images at the acquisition focal distance. In [12] the authors propose a deconvolution-like solution to the problem, while in [13, 14], the authors propose a new popular acquisition scheme, which is known as focused light-field (FLF). Traditionally, the ULF and FLF have been treated as two different acquisition setups that needed different algorithms for similar tasks like refocusing. In the ULF case, notable examples of refocusing approaches are: the integration method [15], the Fourier-space method based on the Fourier slice theorem [16], and a volumetric filtered version of the latter [17]. In the FLF case, we have mainly the patch method [14], and a more sophisticated super-resolution method that explicitly takes into account the positioning of the pixels on the sensor [18]. Recently, in [19], the authors have demonstrated that the algorithms developed for one camera setup can be applied to the other setup. However, the resulting image resolutions and quality depend on the chosen algorithm.

In a recent paper, we described a framework in which a light-field photography image can be interpreted as limited angle cone-beam tomography acquisition, [20]. The framework presents an in-depth analysis of how different parametrizations of the ray-tracing model can be formulated so that the well-known back-projection methods for cone-beam tomography can be used to solve the refocusing problem.

In this paper, we propose to replace the back-projection approach introduced earlier by an *algebraic model* of the cone-beam tomography problem. This new formulation exposes a broad range of more sophisticated tomographic reconstruction techniques, introducing a number of powerful modelling and computational formulations for light field reconstruction. Specifically:

- We extend the formulation by unifying the treatment of unfocused light-field (ULF) and focused light-field (FLF) images.
- We formulate the refocusing reconstruction as a convex optimization problem, introduce additional regularization terms to cope with artifacts, and show how advanced iterative algorithms can be used to solve the problem.
- We show how super-resolved refocused images can be computed.
- We correct for the limited frequency response of the optical system (point spread function).

Each of these extensions contributes to a measurable improvement of the refocusing reconstruction quality.

1.1. Detailed content description

The introduction of tomographic projection operations naturally allows to express the image formation process as a linear system, similarly to [10, 11], where the forward operator is composed for both geometries of: a tomographic forward-projection operator; a down-sampling operator in the case of super-resolved images; and a blurring operator. The blurring operator takes into account the effects of Fraunhofer diffraction on the lenslets [10, 12] and the Nyquist limitation on the refocusing resolution [16].

We propose the use of regularizations like the Total Variation minimization (TV-min) or the stationary wavelet transform (SWT) [21, 22], instead of the traditional smoothness constraints [23].

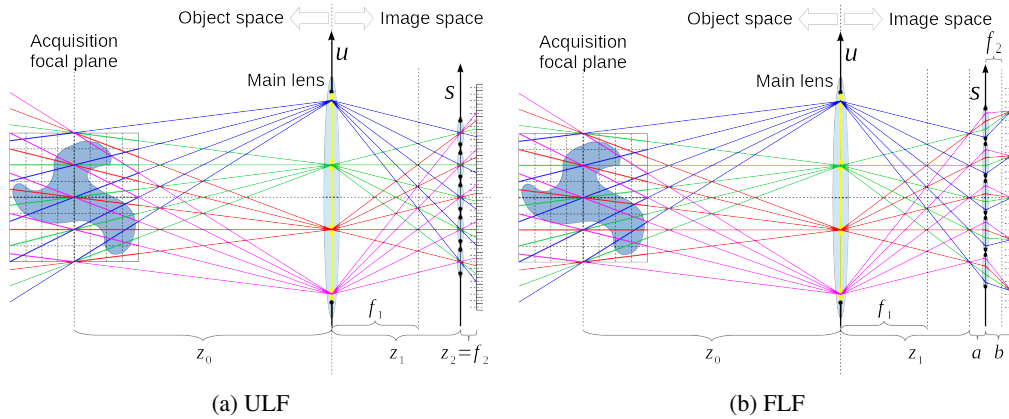


Fig. 1. Comparison of the two light-field setups, and their naming conventions: (a) unfocused light-field (ULF); (b) focused light-field (FLF).

These advanced constraints allow to effectively suppress artifacts related to the super-resolution reconstruction, while enforcing prior knowledge in the reconstruction, like sharpness of the in-focus objects, or its compressibility in the wavelet domain. We also demonstrate the use of both established and recent customizable algorithms, which are widely used in the tomography community and include the Simultaneous Iterative Reconstruction Technique (SIRT) [24, 25], and the Chambolle-Pock (CP) algorithm instances [26, 27]. The use of consolidated algorithms allows to re-use existing high-performance solutions, while the use of CP algorithm instances allows to have tailored algorithms for the most popular regularization functionals in a straight-forward manner.

The rest of the article is structured as follows. In section 2 we introduce the used formalism, explain the geometric adjustments needed for the ULF and FLF cases, construct the minimization problem, and present the used reconstruction techniques. In section 3 we present the results of the reconstruction methods detailed in section 2 on simulated and real-world examples, and compare them against a representative selection of existing methods. In section 4 we conclude the paper.

2. Methods

In this section we introduce and discuss the mathematical tools used in the paper. In section 2.2 we establish the relationship between the projection geometry, the optical system's point spread function (PSF) and the image formation. Then in section 2.3 we express the refocusing as a convex optimization problem. Finally in section 2.4 we introduce the regularization terms used in this article.

2.1. Conventions

We schematically represent the two typical plenoptic acquisition setups in Fig. 1. They are composed of a main lens (ML), a micro-lenses array (MLA), and a sensor. The camera array setup is not considered explicitly here, but it is equivalent to the setup in Fig. 1(a), with the only difference being that it lacks the ML [20].

We adopt the conventions and representations from [20]. In both Figs. 1(a) and 1(b) the ML divides the space into the object space which is the photographed scene (left side) and the image space which is the inside of the camera (right side).

The focal lengths of the ML and the lenslets in the MLA are called f_1 and f_2 respectively, and the distance between the natural focal plane in the object space and the ML is called z_0 . In the

ULF case (Fig. 1(a)), the MLA-to-sensor distance is called z_2 and is usually set to f_2 , while the ML-to-MLA distance is z_1 . Given the fact that $f_2 \ll z_1$, the ML is considered in this case to be at the optical infinity of the lenslets in the MLA. In the FLF case, we can have two configurations: “Keplerian” and “Galilean” [28]. For simplicity, throughout this paper, we only work with a Keplerian setup (represented in Fig. 1(b)), but our approach equally works for the Galilean setup. In Keplerian mode, the image from the ML is formed in front of the MLA, at the distance z_1 , and that has a ML-to-MLA distance equal to $z_1 + a$, with $a > 0$. In Galilean mode, the image from the ML is virtual and it is formed behind the MLA, which is placed again at a distance of $z_1 + a$ from the ML, but with $a < 0$. For an FLF in Keplerian configuration, the MLA-to-sensor distance is called b and satisfies the thin lens equation: $1/f_2 = 1/a + 1/b$.

In the ULF configuration, 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) , whose origin is on the center of the lenslet footprint. In this setup, there is a one-to-one correspondence with coordinates (u, v) on the main lens, given by:

$$u_{\text{ULF}} = \frac{z_1}{f_2} \sigma, \quad (1)$$

$$v_{\text{ULF}} = \frac{z_1}{f_2} \tau. \quad (2)$$

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 apertures in Fig. 1(a).

Compared to the ULF case, we showed in [20] that in the FLF case, we observe a shift in the (s, t) coordinates of each sub-aperture image, which depends on its associated (σ, τ) coordinates:

$$s_{\text{FLF}} = s_{\text{MLA}} + \frac{a}{b} \sigma, \quad (3)$$

$$t_{\text{FLF}} = t_{\text{MLA}} + \frac{a}{b} \tau, \quad (4)$$

where s_{MLA} and t_{MLA} are the coordinates of the center of the lenslets on the MLA, while the (u, v) coordinates become:

$$u_{\text{FLF}} = \frac{z_1 + a}{b} \sigma, \quad (5)$$

$$v_{\text{FLF}} = \frac{z_1 + a}{b} \tau. \quad (6)$$

2.2. Forward model

In [20] we derived the tomographic volumetric forward-projection and back-projection formulas for the ULF, which relate lines in the considered volume of the photographed scene to points in the light-field and vice versa. The complete forward-projection formula is:

$$L'(s_i, t_i, u, v) = \int_{\Omega_o} \delta \left(\frac{z}{z_0} s_o + \left(1 - \frac{z}{z_0} \right) u - M s_i, \frac{z}{z_0} t_o + \left(1 - \frac{z}{z_0} \right) v - M t_i \right) \times E(s_o, t_o, z) ds_o dt_o dz_o, \quad (7)$$

where $M = z_0/z_1$ is the magnification of the recorded pixels in the photographed scene, (s_o, t_o) are the transverse coordinates in the object space, (s_i, t_i) are the transverse coordinates in the image space, $L'(s_i, t_i, u, v)$ is the computed light-field, $E(s_o, t_o, z)$ is the function describing the

photographed scene, with support $\Omega_o = \Omega_{s_o} \times \Omega_{t_o} \times \Omega_z$, where $\Omega_{s_o} = [-S_o, S_o]$, $\Omega_{t_o} = [-T_o, T_o]$, and $\Omega_z = [-Z, Z]$. The back-projection (integration refocusing) formula is instead:

$$E'(s_o, t_o, z) = \int_{\Omega_i} \delta \left(\frac{z}{z_0} s_o + \left(1 - \frac{z}{z_0} \right) u - M s_i, \frac{z}{z_0} t_o + \left(1 - \frac{z}{z_0} \right) v - M t_i \right) \times L(s_i, t_i, u, v) ds_i dt_i du dv, \quad (8)$$

where $E'(s_o, t_o, z)$ is the produced focal stack, $L(s_i, t_i, u, v)$ is the recorded light-field, with support $\Omega_i = \Omega_{s_i} \times \Omega_{t_i} \times \Omega_u \times \Omega_v$, where $\Omega_{s_i} = [-S_i, S_i]$, $\Omega_{t_i} = [-T_i, T_i]$, $\Omega_u = [-U, U]$, and $\Omega_v = [-V, V]$. $|U|$ and $|V|$ are the largest sampled coordinates on the ML.

Equations (7) and (8) are also valid for the FLF case when considering the coordinates (s_i, t_i) on the plane at distance z_1 inside the camera. From Eqs. (3) and (4), we deduce that each recorded sub-aperture image presents a displacement equal to $\Delta s_i = (a/b)\Delta\sigma$ and $\Delta t_i = (a/b)\Delta\tau$, respectively. Using Eqs. (5) and (6), we can compute the displacements of the recorded sub-aperture images in the image space, as functions of the (u, v) coordinates:

$$\Delta s_{i,\Delta u} = a/(z_1 + a) \Delta u, \quad (9)$$

$$\Delta t_{i,\Delta v} = a/(z_1 + a) \Delta v. \quad (10)$$

In the object space, the shifts become:

$$\Delta s_{o,\Delta u} = M \Delta s_{i,\Delta u} = Ma/(z_1 + a) \Delta u, \quad (11)$$

$$\Delta t_{o,\Delta v} = M \Delta t_{i,\Delta v} = Ma/(z_1 + a) \Delta v. \quad (12)$$

The resulting forward-projection equation for the FLF case is simply:

$$L'(s_i, t_i, u, v) = \int_{\Omega_o} \delta \left(\frac{z}{z_0} s_o + \left(1 - \frac{z}{z_0} + \frac{Ma}{z_1 + a} \right) u - M s_i, \frac{z}{z_0} t_o + \left(1 - \frac{z}{z_0} + \frac{Ma}{z_1 + a} \right) v - M t_i \right) \times E(s_o, t_o, z) ds_o dt_o dz_o, \quad (13)$$

and the back-projection equation is:

$$E'(s_o, t_o, z) = \int_{\Omega_i} \delta \left(\frac{z}{z_0} s_o + \left(1 - \frac{z}{z_0} + \frac{Ma}{z_1 + a} \right) u - M s_i, \frac{z}{z_0} t_o + \left(1 - \frac{z}{z_0} + \frac{Ma}{z_1 + a} \right) v - M t_i \right) \times L(s_i, t_i, u, v) ds_i dt_i du dv. \quad (14)$$

From a tomographic perspective, in the object space, both Eq. (7) for the ULF and Eq. (13) for the FLF compute the line integral for the pixel (s_o, t_o, z_o) of a projection having source in $(u, v, 0)$, in a cone-beam geometry acquisition. The only difference between the two resides in the definition of the pixel (s_o, t_o, z_o) , which is shifted by the amounts $(\Delta s_{o,u}, \Delta t_{o,v}, 0)$ in the FLF configuration. Due to this, they can be both re-interpreted as the application of the integral forward-projection operator $\mathbb{A}[\cdot](s_o, t_o, z) : \mathbb{R}^3 \mapsto \mathbb{R}^4$ (which we call $\mathbb{A}(s_i, t_i, u, v; s_o, t_o, z)$ with a small abuse of notation) to the function $E(s_o, t_o, z)$, transforming Eqs. (7) and (13) into:

$$L'(s_i, t_i, u, v) = \mathbb{A}[E(s_o, t_o, z)](s_i, t_i, u, v), \quad (15)$$

where the integration is over the coordinates (s_o, t_o, z) , and the details of the geometry are absorbed into the operator $\mathbb{A}(s_i, t_i, u, v; s_o, t_o, z)$. Equations (8) and (14) are the adjoint operations of Eqs. (7) and (13), respectively. So, Eqs. (8) and (14) can also be re-interpreted both as the application of the integral back-projection operator $\mathbb{A}^\dagger[\cdot](s_i, t_i, u, v) : \mathbb{R}^4 \mapsto \mathbb{R}^3$ (or $\mathbb{A}^\dagger(s_o, t_o, z; s_i, t_i, u, v)$, as before), to the function $L(s_i, t_i, u, v)$:

$$E'(s_o, t_o, z) = \mathbb{A}^\dagger[L(s_i, t_i, u, v)](s_o, t_o, z), \quad (16)$$

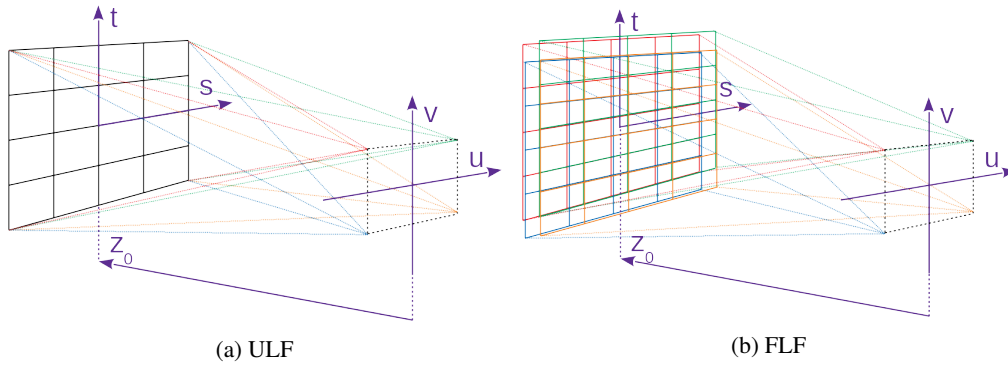


Fig. 2. Projection geometry of the two light-field acquisition geometries: (a) unfocused light-field (ULF); (b) focused light-field (FLF). At the distance z_0 all the sub-aperture images coincide on the (s, t) axes for the ULF case, while they experience shifts $\Delta s = (a/b)\Delta\sigma$ and $\Delta t = (a/b)\Delta\tau$ for the FLF case.

where the integration is over the coordinates (s_i, t_i, u, v) , and \mathbb{A}^\dagger is the adjoint of \mathbb{A} . The forward-projection matrix \mathbf{A} is the matrix representation of the operator $\mathbb{A}(s_i, t_i, u, v; s_o, t_o, z)$, where the rows are selected by the set of indexes $\{s_i, t_i, u, v\}$, and the columns are selected by the set of indexes $\{s_o, t_o, z\}$. The back-projection matrix \mathbf{A}^T is then the matrix representation of the operator $\mathbb{A}^\dagger(s_o, t_o, z; s_i, t_i, u, v)$ where the rows are selected by the set of indexes $\{s_o, t_o, z\}$, and the columns are selected by the set of indexes $\{s_i, t_i, u, v\}$. The matrices \mathbf{A} and \mathbf{A}^T are sparse, and have real non-negative values [29].

In this article, we restrict ourselves to refocusing at a single distance, as opposite to reconstructing an entire volume from one light-field image. This turns Eq. (7) into the projection from a slice at position z in the volume to the light-field:

$$L'(s_i, t_i, u, v) = \int_{\Omega_{o,z'}} \delta(\alpha s_o + (1 - \alpha)u - Ms_i, \alpha t_o + (1 - \alpha)v - Mt_i) \times E(s_o, t_o, z') ds_o dt_o, \quad (17)$$

and Eq. (8) into the projection from the light-field to such slice at position z :

$$E'(s_o, t_o) = \int_{\Omega_i} \delta(\alpha s_o + (1 - \alpha)u - Ms_i, \alpha t_o + (1 - \alpha)v - Mt_i) \times L(s_i, t_i, u, v) ds_i dt_i du dv, \quad (18)$$

where $\alpha = z'/z_0$, and the support $\Omega_{o,z}$ is the slice at position $z = z'$. In Fig. 2(a) we show the geometric representation for Eqs. (17) and (18) and $\alpha = 1$. For $\alpha \neq 1$, the dotted lines remain the same, but the image grid is shifted to the corresponding $z = \alpha z_0$ position.

For the FLF case, Eq. (13) becomes:

$$L'(s_i, t_i, u, v) = \int_{\Omega_{o,z'}} \delta\left(\alpha s_o + \left(1 - \alpha + \frac{Ma}{z_1 + a}\right)u - Ms_i, \alpha t_o + \left(1 - \alpha + \frac{Ma}{z_1 + a}\right)v - Mt_i\right) \times E(s_o, t_o, z') ds_o dt_o, \quad (19)$$

and Eq. (14) becomes:

$$E(s_o, t_o) = \int_{\Omega_i} \delta\left(\alpha s_o + \left(1 - \alpha + \frac{Ma}{z_1 + a}\right)u - Ms_i, \alpha t_o + \left(1 - \alpha + \frac{Ma}{z_1 + a}\right)v - Mt_i\right) \times L(s_i, t_i, u, v) ds_i dt_i du dv. \quad (20)$$

Figure 2(b) gives a representation of the projection geometry of different sub-aperture images, for the FLF case.

Without loss of generality we can identify again A_z as the matrix representation of both the forward-projection operators in Eqs. (17) and (19), and A_z^T as the matrix representation of the back-projection operators in Eqs. (18) and (20).

If we now consider the discretization of the functions $L(s_i, t_i, u, v)$ and $E_z'(s_o, t_o)$ as the vectors \mathbf{b} and \mathbf{x}_z respectively, the image formation process for one slice is represented by the following linear system:

$$A_z \bar{\mathbf{x}}_z = \mathbf{b}, \quad (21)$$

where $\bar{\mathbf{x}}_z$ is the vector representation of the considered slice in the photographed scene, and \mathbf{b} is the vector representation of the recorded light-field. In Eq. (21) we assume that all the recorded light comes from the chosen slice at distance z .

Any object outside the slab defined by the acquisition depth-of-field (DoF) around z , is not reconstructed correctly and appears out of focus in the refocused images. The acquisition DoF is the sum of two factors: a geometric and a diffractive component. The former was also considered in [20] from the point of view of a tomographic acquisition. It is linearly proportional to the lenslet diameter and inversely proportional to the ML numerical aperture (NA). The latter has instead an inverse quadratic proportionality to the ML NA, and it is proportional to the detected photon wavelength. For a more in depth explanation and derivation of the DoF we refer to [30]. For translucent objects, this means that if they are in-focus, their shape is correctly refocused, but the objects outside of the DoF that are visible through them are imaged out of focus.

In case of super-resolved reconstructions, we assume that the refocused slice in the photographed scene has a smaller pixel size than the recorded light-field in the (s, t) coordinates. Equation (21) can be modified to remove this discrepancy, by adding a down-scaling operation after the projection of the matrix A_z . The down-scaling matrix U_n bins the value of $n \times n$ neighbouring pixels into a single value, where n is the ratio between the recorded light-field resolution and the slice resolution in the (s, t) coordinates. The resulting image formation process from Eq. (21) becomes:

$$U_n A_z \bar{\mathbf{x}}_z = \mathbf{b}. \quad (22)$$

The optical system's limited frequency response is also introduced in the form of a convolution matrix P_z (Toeplitz matrix), applied to the binned images [31]. This matrix performs the convolution of the binned images with the total PSF of the optical system. This PSF itself is the convolution of the PSFs generated by the lenses' limited size and the Nyquist limitation on the resolution. The ML and lenslet's PSFs correspond to the Airy functions defined by their size and the captured light wavelength [2, 11, 32]. They do not depend on z , and they are applied over the (s_i, t_i) and (u, v) coordinates respectively. The Nyquist limitation on the resolution corresponds to the integration over the corresponding portion to a given sub-aperture image on the ML [16]. It has the shape of a flat disk whose width depends on the distance z , refocusing distance z_0 and (u, v) resolution, and it is applied over the (s_i, t_i) coordinates.

By introducing the optical system's limited frequency response and super-resolution capabilities, the complete image formation process for one slice from Eq. (22) becomes:

$$P_z U_n A_z \bar{\mathbf{x}}_z = \mathbf{b}, \quad (23)$$

where U_n is the binning matrix for the chosen n super-resolution, P_z is the convolutional matrix that applies the optical system's PSF.

The matrices A_z and A_z^T can be computed explicitly by using the sets of unit vectors $\mathbf{E}_o = \{\mathbf{e}_{o,1}, \mathbf{e}_{o,2}, \dots\}$ and $\mathbf{L}_i = \{\mathbf{l}_{i,1}, \mathbf{l}_{i,2}, \dots\}$ respectively. The unit vectors \mathbf{E}_o correspond each to

a volume in the (s_o, t_o, z) coordinates, where every value is zero, except for the selected point of position $(s_{o,k}, t_{o,k}, z_k)$, which is one. The same goes for the unit vectors L_i in the coordinates (s_i, t_i, u, v) . These vectors are used in Eqs. (17)-(20) to compute a volume in the adjoint space. These resulting volumes are the columns of A_z and the rows of A_z^T when using unit vectors E_o , while they are the rows of A_z and the columns of A_z^T when using the unit vectors L_i . While this procedure can give mathematical insight about the projection operations, for practical applications, it is better to use existing efficient tomographic implementations, like the ASTRA toolbox [33].

2.3. Refocusing as a minimization problem

Refocusing is the procedure of retrieving a photograph $E_z(s_o, t_o)$ for a given refocusing distance $z = \alpha z_0$, from the light-field acquired at z_0 . It was proven in [20] that the simple back-projection of the sub-aperture images corresponds to the integration method for non super-resolved photographs described in [1]. According to Eq. (21), we obtain the stack of refocused images $\hat{\mathbf{x}}_z$, from Eqs. (18) and (20), as follows:

$$\hat{\mathbf{x}}_z = A_z^T \mathbf{b}, \quad (24)$$

where \mathbf{b} is again the vector representation of the recorded light-field. For super-resolved images, based on the forward model from Eq. (22), we update Eq. (24) to:

$$\hat{\mathbf{x}}_z = A_z^T U_n^T \mathbf{b}, \quad (25)$$

where U_n^T is the up-sampling matrix, transpose of the U_n down-sampling matrix.

The Fourier slice theorem method from [16] uses an approximation \tilde{A}_z^T to the back-projection matrix A_z^T which is computationally less expensive, resulting in a very similar refocusing problem:

$$\hat{\mathbf{x}}_z = \tilde{A}_z^T U_n^T \mathbf{b}. \quad (26)$$

When refocusing to only one slice, the method from [17] is equivalent to the one from [16], in the sense that it performs a Gaussian re-sampling of the Fourier space instead of using Kaiser-Bessel functions. This results in just a slightly different approximation matrix \tilde{A}_z^T for Eq. (26).

In this article, we formulate the refocusing reconstruction as a convex minimization problem, similarly to the high resolution image reconstruction from [10, 11]. We minimize the residual on the captured light-field, using Eq. (23):

$$\begin{aligned} \hat{\mathbf{x}}_z &= \underset{\mathbf{x}}{\operatorname{argmin}} \{ ||P_z U_n A_z \mathbf{x} - \mathbf{b}||_2^2 \} \\ \text{subject to: } &\mathbf{x} \geq \mathbf{0}, \end{aligned} \quad (27)$$

where the non-negativity constraint imposes the fact that there cannot be negative light scattering. This representation can be solved by existing iterative algorithms like SIRT and Conjugate Gradient Least-Squares (CGLS) [25]. The SIRT implementation for the solution of Eq. (27) can be found in Appendix A.

2.4. Regularization

The recent introduction of reconstruction algorithm classes like Chambolle-Pock [26, 27], offers the ability to generate algorithm instances that solve arbitrary problems characterized by convex and possibly non-smooth terms, in a straight-forward manner. This allows to surpass the implementations from [23] and [10], and to augment Eq. (27) with non-smooth convex terms

like the following:

$$\begin{aligned} \hat{\mathbf{x}}_z &= \underset{\mathbf{x}}{\operatorname{argmin}} \{ \|\mathbf{P}_z \mathbf{U}_n \mathbf{A}_z \mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{O} \mathbf{x}\|_1 \} \\ \text{subject to: } \mathbf{x} &\geq \mathbf{0}, \end{aligned} \quad (28)$$

where λ is a weighting parameter, and \mathbf{O} is an operator.

The l_1 -norm promotes sparsity [34], and we can choose an operator \mathbf{O} that returns a sparse representation of the solution vector \mathbf{x} when it presents certain characteristics. This can be used to impose prior knowledge on the reconstruction. For instance, natural images present a sparse wavelet decomposition, while this is not the case for white noise [34]. Thus, reducing the l_1 -norm of the wavelet decomposition for natural images, reduces the impact of white noise and possibly other sources of noise, like pixel value quantization and pixel surface integration.

Another popular choice for \mathbf{O} is the modulus of the point-wise gradient, whose l_1 -norm is known as Total Variation: $\operatorname{TV}(\cdot) = \|(|\nabla \cdot|)\|_1$. The TV term promotes isotropic gradient sparsity in the reconstructed images, thus favouring flat regions with sharp boundaries [27].

In this article, we propose to use either the TV term or the translation-invariant wavelet transform (SWT) [21,22]. The SWT is an overcomplete wavelet decomposition, that allows to remove the well-known block artifacts that characterize high compression levels of the traditional discrete wavelet transform (DWT), while preserving its sparsity properties. The CP implementation for the solution of Eq. (28) with a TV regularization term can be found in Appendix A.

2.5. Practical limitations

The model presented in section 2.2 suffers from a few practical limitations. Three main limitations are: lenslets are usually not perfectly aligned with the underlying sensor pixel grid; the ML does not have real pinholes, but it integrates over a finite region; and the ML causes geometric distortion.

The first limitation implies a mismatch of the (σ, τ) coordinates between the pixels under each lenslet. As a result, none of the presented methods, including the one from the literature can be directly applied to the collected data. This is routinely solved by a re-interpolation of the pixel grid from the area underneath each lenslet (lenslet footprint). The resulting re-sampled grid has the desired matching of (σ, τ) coordinates, but it suffers from a small penalty in directional resolution, as large as half of the (u, v) pixel-size.

The second corresponds to the Nyquist limitation on the resolution discussed in section 2.2. It results in a disk blur dependent on the refocusing distance from the original focal distance z_0 , and it can be mitigated through the inclusion of such blurring term in the projection model, as discussed in section 2.2.

The third limitation causes straight lines to be reconstructed as curved lines. This is a well-known problem in computer vision, and it is routinely solved by a check-board calibration of the camera, which returns a map of its distortion. Once the camera has been calibrated, the acquired images can be rectified through a common image distortion correction procedure. Reliable software for carrying out this procedure already exists. An example is the “Light Field Toolbox for Matlab” (LFT: <http://dgd.vision/Tools/LFTToolbox/>) from Donald G. Dansereau.

For a more in depth analysis of lenslet alignment with the underlying sensor, and the geometric distortion introduced by the ML, we refer to [35].

3. Reconstructions

We now address the reconstruction performance and applicability of the proposed methods. We first use a known phantom from [20] (the “logos” test case) to test the reconstruction fidelity of the refocused objects, for different up-scaling factors (Fig. 3). For brevity we only reconstruct the

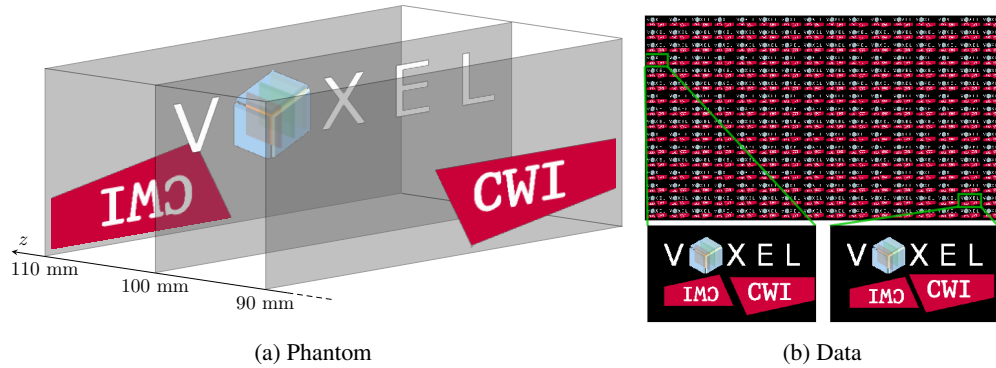


Fig. 3. Description of the logos synthetic case: (a) phantom used to simulate light field data, having a VOXEL logo at the acquisition focal distance $z = 100$ mm, and two CWI logos at the distances of 110 mm, and 90 mm; (b) simulated data, with two sub-aperture images for clarity.

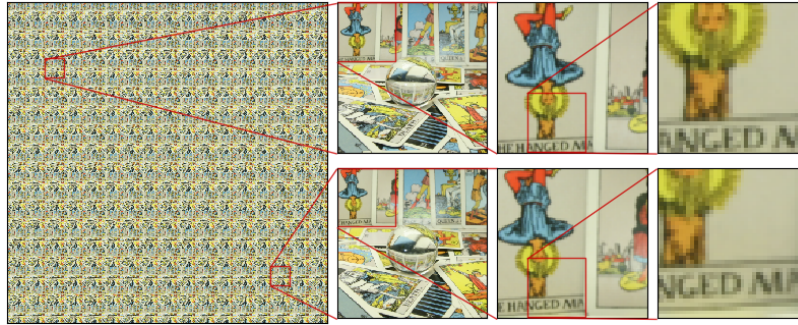
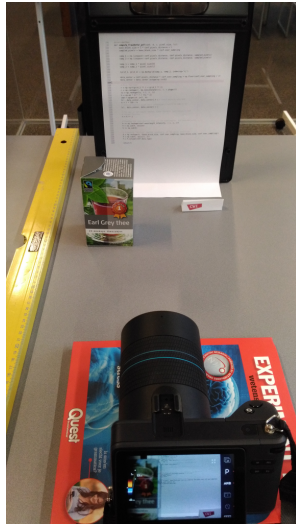


Fig. 4. Raw data from the Tarot Cards dataset. The data has been downsampled by a factor 4 in the (s, t) coordinates, and blurred with a known PSF.

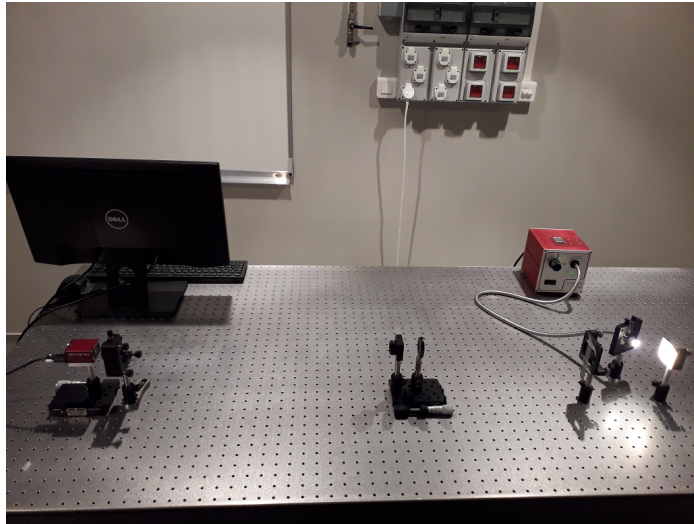
logos example for an ULF camera geometry. We then demonstrate the ability of our approach to reconstruct real-data for both ULF and FLF camera configurations.

For the ULF case, we first show reconstructions of two versions of the “Tarot Cards” dataset from the Stanford Light-Field Archive (<http://lightfield.stanford.edu/lfs.html>), originally acquired by *Andrew Adams* (data: Fig. 4). The two versions are used for different purposes: one is downsampled and blurred to simulate an MLA-based acquisition setup, and assess the PSF correction and super-resolution performance; the other is the original full-resolution version, which is used to show that the developed methods can work with full-resolution gantry setups. To further test the ULF case, we then show reconstructions of the “Tea” dataset, acquired with the setup in Fig. 5(a) using the Lytro Illum plenoptic camera (data: Fig. 6).

For the FLF case, we show the reconstruction of two datasets called “Letters” (micro-images: Fig. 7(a), sub-aperture images: Fig. 7(b)) and “Flower” (micro-images: Fig. 8(b), sub-aperture images: Fig. 8(c)), that we acquired experimentally with the dedicated setup in Fig. 5(b). The Letters dataset depicts vector art, that is easily recognizable by the human eye, and that is favoured by the TV-min constraint. The Flower dataset depicts natural imagery (https://github.com/pratulsrinivasan/Local_Light_Field_Synthesis), which is more complex and favoured by the SWT-min constraint.



(a) ULF setup



(b) FLF setup

Fig. 5. Experimental setups used for the acquisition of the ULF and FLF datasets “Tea”, “Letters”, and “Flower”: (a) ULF setup, using the Lytro Illum plenoptic camera; (b) FLF setup, built at Imagine Optic in Bordeaux (France).

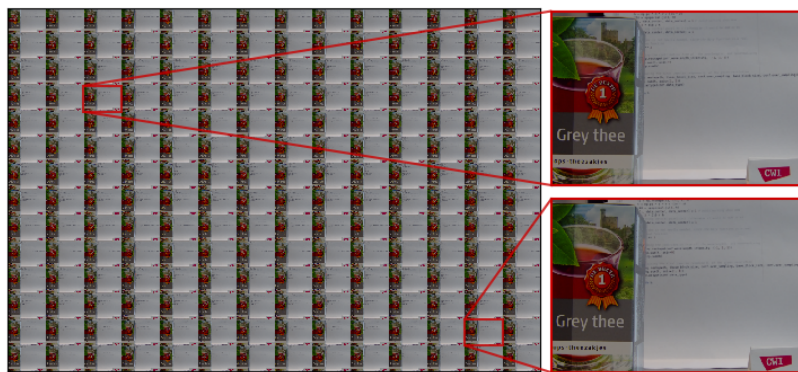


Fig. 6. Raw data from the Tea dataset acquired using a Lytro Illum plenoptic camera, with the setup from Fig. 5(a).

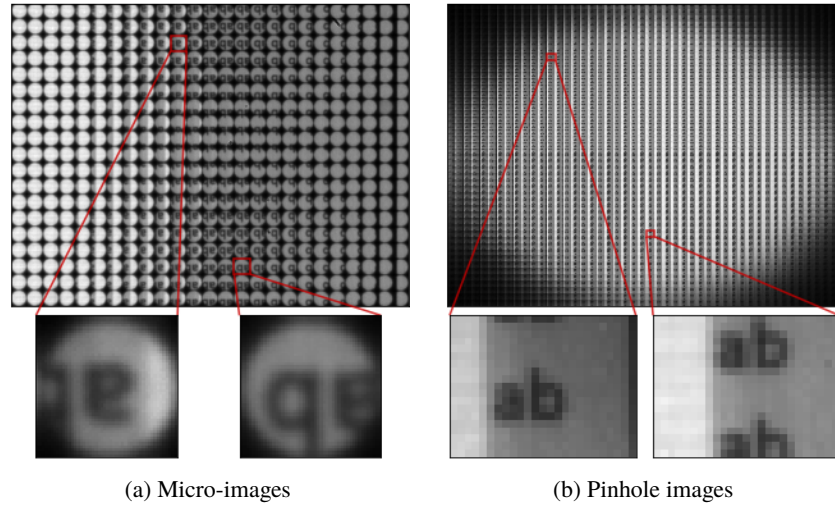


Fig. 7. Raw data for the Letters case: (a) Micro-image representation; (b) Sub-aperture representation. In both cases, two zoomed insets have been shown for clarity. The observed inversion between (a) and (b) is due to the fact that FLF data mixes spatial information with angular information at the micro-images [36], and that the lenslets perform an image space inversion of the images that form on their object space focal plane.

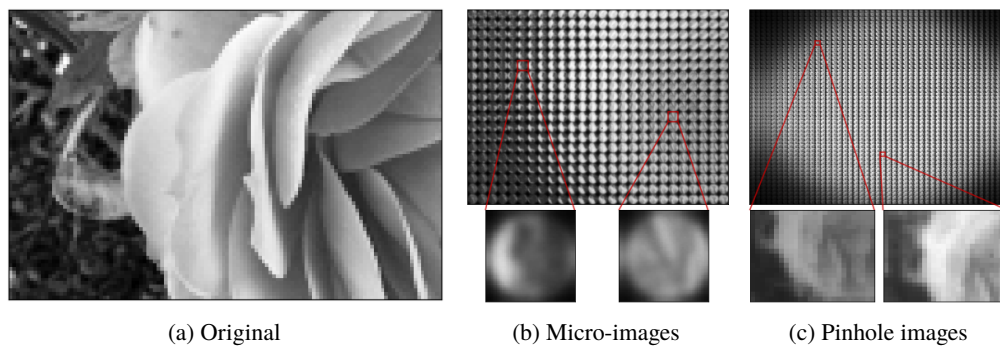


Fig. 8. Raw data for the Flower case: (a) Original (high resolution) image used for this test case; (b) Micro-image representation; (c) Sub-aperture representation. In both cases, two zoomed insets have been shown for clarity.

3.1. Data description

The logos scene is composed of three different objects positioned at 90, 100 and 110 mm respectively from the ML, as displayed in Fig. 3(a). The dataset is composed of 16×16 sub-aperture images, each composed of 512×256 pixels. The sub-aperture images for this case are presented in Fig. 3(b). The diameter d_1 of the main lens is 10 mm, while each lenslet has a diameter $d_2 = 16 \mu\text{m}$. The distance z_2 is $50 \mu\text{m}$, and the detector pixel size $1 \mu\text{m}$. The ML-to-MLA distance z_1 is 25 mm, so that the focal plane in the object space is at 100 mm ($|M| = 4$). In the creation of this dataset we applied the two PSFs due to Fraunhofer diffraction for the ML and the lenslets in the MLA. These PSFs were generated by computing the Fraunhofer diffraction pattern predicted for the lens aperture (d_1 and d_2 for ML and MLA respectively), detected light wavelength, and lens-to-target distance (z_1 and z_2 for ML and MLA respectively). This diffraction pattern was then integrated in steps equal to the underlying pixel size ($16 \mu\text{m}$ and $1 \mu\text{m}$ for ML and MLA respectively).

To enable the evaluation of the up-sampling refocusing reconstruction quality for the logos case, we also created two additional versions of such test case, where we binned (down-sampled) the sub-aperture images along the (s, t) coordinates by 2 and 4. The reconstruction of these down-sampled datasets was performed at the original dataset (s, t) resolution, turning it into a super-resolved reconstruction problem with corresponding up-sampling factors of 2 and 4, respectively. The resulting up-sampled refocused images could then be compared to the original (high resolution) phantom. In general, this means that for increasing sampling factors, the available data (s, t) resolution becomes worse, less data is available, and more operations are needed to produce images with the original (s, t) resolution. This procedure also resulted in the re-scaling of the expected lenslet sizes, detector pixel sizes, and PSF widths, according to the sampling factors.

The Tarot cards dataset has been acquired using a gantry setup (analogous to a camera array setup), which was built using Lego Mindstorms, and a Canon Digital Rebel XTi, with a Canon 10-22 mm lens. The cropped and registered dataset from the archive comes as a collection of 17×17 sub-aperture images, with a relatively high (s, t) resolution (1024×1024). The focal length of the main lens f_1 was set to 20 mm. We binned the sub-aperture images to a resolution of 256×256 , and blurred the light-field in the (u, v) coordinates using a different PSF per color channel. The three PSFs have been computed as the result of the Fraunhofer diffraction effects due to the lenslets, with the following parameters: lenslet diameter $d_2 = 20 \mu\text{m}$, distance $z_2 = 37 \mu\text{m}$ and detector pixel size $1.4 \mu\text{m}$. The parameters for the lenslets are the same as the ones in a well known plenoptic camera, commercially available under the name of *Lytro ILLUM*. The modified Tarots cards dataset is displayed in Fig. 4.

The Tea dataset (Fig. 6) has been acquired using a Lytro Illum plenoptic camera, with the setup in Fig. 5(a). Being the Lytro Illum a proprietary camera, we do not have the complete acquisition parameters. The raw data and metadata were extracted using the proprietary software “Lytro power tools”, property of Lytro. This software already performs the rectification and lenslet re-sampling procedures discussed in section 2.5.

The Letters and Flower datasets (Figs. 7 and 8) represent printed vector art and natural imagery, that was then photographed with the dedicated setup in Fig. 5(b), in the laboratories of Imagine Optic (Bordeaux, France). In this setup, we used the AC254-200-A-ML ML and MLA300-14AR-M MLA from Thorlabs, and the Stingray F-145 sensor from Allied Vision. The ML has focal distance $f_1 = 200$ mm, and the MLA is composed of $\sim 30 \times 30$ lenslets with $f_2 = 18.6$ mm, diameter $d_2 = 300 \mu\text{m}$, and built-in f -number equal to 62. The detector pixel size is $6.45 \mu\text{m}$. We chose a z_1 distance of 400 mm, which gave a z_0 distance of 400 mm and $|M| = 1$. We also chose a distance $b = 20.93$ mm, which corresponded in a distance $a = 167.4$ mm. The field-of-view (FoV) of the scene is $7.2 \text{ mm} \times 4.8 \text{ mm}$, while the DoF is $\approx 2.5 \text{ mm}$ [30]. The degradation of the micro-image quality observed in both Figs. 7 and 8 is due to the PSF of the

imaging system, and in particular to the large lenslet f -number, which negatively correlates to the imaging system spatial resolution [30].

For the reconstruction of all these test cases, the chosen regularization weight λ values have been selected empirically by trial and error. In the case of multi-channel data (e.g. RGB), we performed a single-channel reconstruction, where each channel has been reconstructed separately. However, the same λ value was used for the reconstruction of all the channels.

3.2. Performance analysis

To evaluate the reconstruction fidelity of the logos test case, we can directly compare the refocused regions with the phantom, and compute the *root mean square error* (RMSE) of each refocused image using the formula:

$$RMSE(\mathbf{x}_z) = \sqrt{\frac{\sum_n^N ((\mathbf{x}_z)_n - (\hat{\mathbf{x}}_z)_n)^2}{N}} \quad (29)$$

where the summation is over the selected N pixels, and the vector $\hat{\mathbf{x}}_z$ represents the expected solution at the given depth z . Lower values of the RMSE correspond to better refocusing performance.

3.3. Logos refocusing

The logos test case is a benchmark on the impact of both the PSF correction and the regularization, with respect to different up-sampling parameters. We compare the refocusing reconstruction results between a few different reconstruction approaches: integration (known also as “shift&sum”), the mathematically equivalent tomographic simple back-projection, Fourier slice theorem, SIRT with and without PSF correction, and CP using the Least Squares (l_2) data-divergence term and TV penalization (abbreviated to CP-LS-TV), both with and without PSF correction. The refocusing was performed on the CWI logo at distance 90 mm from the ML, corresponding to $\alpha = 0.9$ (as defined in section 2.3).

As mentioned in section 2.3, the choice of a simple Fourier slice theorem implementation from [16], over a filtered version from [17] is due to the fact that the two versions, for planar refocusing only differ by the introduction of a filtering step that is equivalent to a Gaussian resampling, instead of the Kaiser-Bessel resampling proposed in [16]. This means that on a noiseless phantom, the two methods should behave very similarly.

In Fig. 9 we present all the different reconstructions against the all-in-focus phantom, at the original phantom resolution (sampling factor = 1: down-sampling of sub-aperture images = 1, up-sampling of reconstruction = 1). It is apparent from the comparison of Figs. 9(b) to 9(d) against Figs. 9(e) to 9(h) that iterative methods can define better sharp transitions than traditional one-step refocusing methods, even in the presence of blurring. Then, from the comparison between Figs. 9(e) to 9(f) and Figs. 9(g) to 9(h), we see that the PSF correction greatly improves the recovery of sharp boundaries. From the comparison between Fig. 9(g) and Fig. 9(h), by looking close to the edge of the CWI target in the left inset of each figure, it is possible to notice that the additional regularization term helps in reducing the wavy type of artifacts due to the PSF reconstruction.

Figure 10(a) quantifies the performance of the different methods, by plotting the RMSEs of each reconstruction method for the down-sampling factors 1, 2 and 4 of the sub-aperture images in the original light-field, and corresponding up-sampling of the refocusing reconstructions. As also observed for Fig. 9, the one-step refocusing methods present similar reconstruction accuracy across the board, with the difference between the integration and the back-projection approach simply being due to the different interpolation method used in their implementations. Among the iterative methods, the PSF correction can improve the reconstruction accuracy by almost

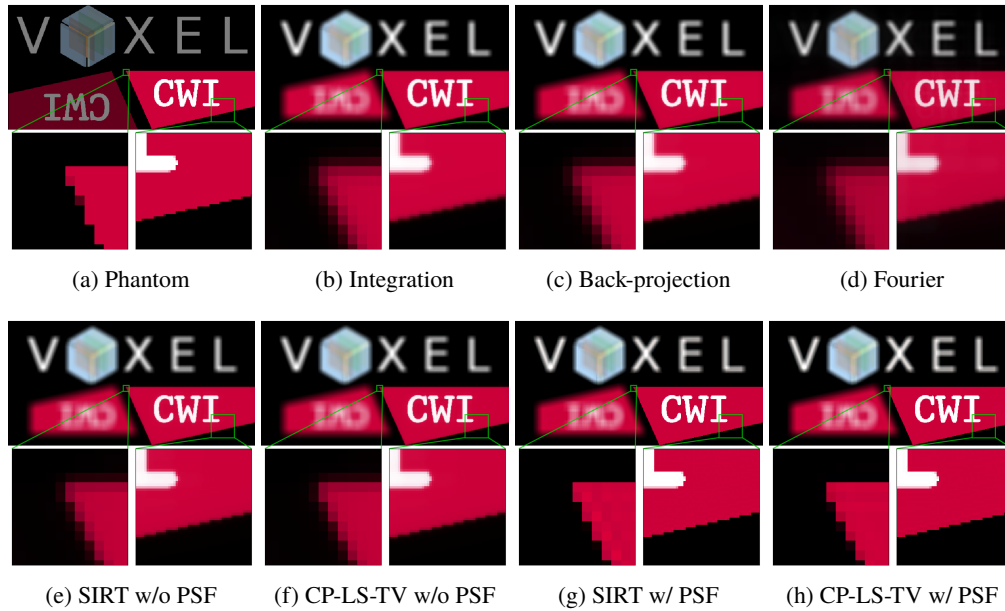


Fig. 9. Performance comparison of different refocusing approaches for the “logos” synthetic test case: (a) Phantom; (b) Integration; (c) Back-projection; (d) Fourier slice theorem; (e) SIRT without PSF; (f) Chambolle-Pock without PSF, with l_2 data-divergence term and TV regularization term with $\lambda = 1$; (g) SIRT with PSF; (h) Chambolle-Pock using PSF, l_2 data-divergence term, and TV regularization term with $\lambda = 1$.

an order of magnitude at the highest resolutions, while it becomes less important with higher detector binning factors, because the PSF effect tends to be substantially reduced [10]. This can be observed in particular for the sampling factor equal to 2, where the SIRT and CP-LS-TV reconstructions without PSF correction improve their refocusing performance with respect to the original dataset, because of a reduction in size of the lenslet PSF. The choice of the correct regularization (in this case TV-min) improves the result over non-regularized reconstructions, and it becomes relatively more important at higher sampling factors and with smaller PSFs.

Finally, in Fig. 10(b) we can observe the computational time required for every tested method at each down-sampling factor. Even if the focus of this article is not on the computational performance of the proposed method, we believe that it is worth giving an understanding of the performance penalties associated with each addressed aspect. These tests were carried out on a machine equipped with an Intel i7-6700K CPU, 64GB of DDR4 memory, and a NVIDIA GeForce GTX 970 GPU. The fastest method is the back-projection because it is performed using the ASTRA toolbox, which uses GPU acceleration. All the iterative methods also use the GPU acceleration offered by ASTRA for the forward and back-projection, but perform the rest of the operations on the CPU. This imposes data transfers across the PCIe bus (from GPU to CPU and vice versa) at each iteration. We expect to see substantial improvements by running these refocusing routines entirely on the GPU, by both avoiding bottlenecks due to the data transfers, and allowing to accelerate every other operation, including the PSF application. For the same number of iterations, the addition of the TV minimization term has a negligible impact on the refocusing performance. The PSF addition causes the highest performance penalty, which however depends on the size of the PSF: for increasing down-scaling factors, it has a progressively lower impact. The Fourier method is also showing better performance with higher down-sampling factors, because the initial 4D-FFT is performed on smaller data sizes [16].

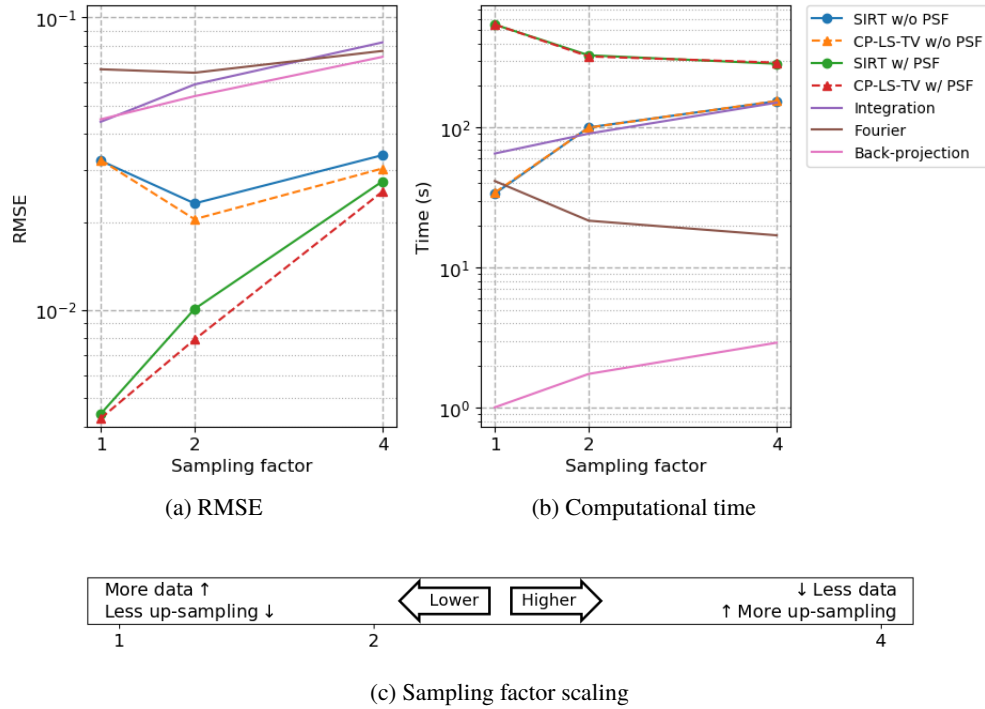


Fig. 10. Root-Mean-Square-Errors (RMSEs) and computational time for each reconstruction approach from Fig. 9 for the down-sampling factors 1, 2, and 4 of the sub-aperture images (the (s, t) coordinates of the light-field): (a) RMSE; (b) Computational time; (c) Scaling of sub-aperture image resolution and reconstruction up-sampling factor against the sampling factor.

3.4. Tarot cards refocusing

The Tarot cards test case shows the super-resolution and artifact correction potential of the proposed methods, over real-data, and in particular over a well understood and widely used dataset.

We first analyze the results of the downscaled and blurred dataset. In Fig. 11, we can observe the refocusing performance of the most significant reconstruction methods, for the two different up-sampling parameters of 1 and 4. For this test case, we chose to use: back-projection, Fourier slice theorem, SIRT with PSF, and CP-LS with SWT penalization based on Haar wavelets (CP-LS-SWT) and PSF. Again, the back-projection and Fourier based methods deliver comparable refocusing and feature-resolving performance. In Fig. 11(c), for up-scaling equal to 4, the SIRT reconstruction presents a better feature resolving performance than the one-step approaches, but also visible artifacts in the reconstruction. The SWT penalized reconstruction from Fig. 11(d) visibly preserves the higher feature resolving performance of the SIRT approach, while strongly reducing the said artifacts. The choice of the SWT penalization term over the TV term can be understood through the same reasoning detailed in section 2.4. The Tarot Cards dataset presents a variety of features, including shaded regions. These regions break the assumption of sharp boundaries associated to the TV minimization.

We now consider the reconstruction of the unmodified Tarots Cards dataset. It can be deduced from Fig. 12 that the approaches presented in this article allow to correctly reconstruct high resolution gantry data. By comparing the results of the different methods in Fig. 12, no difference

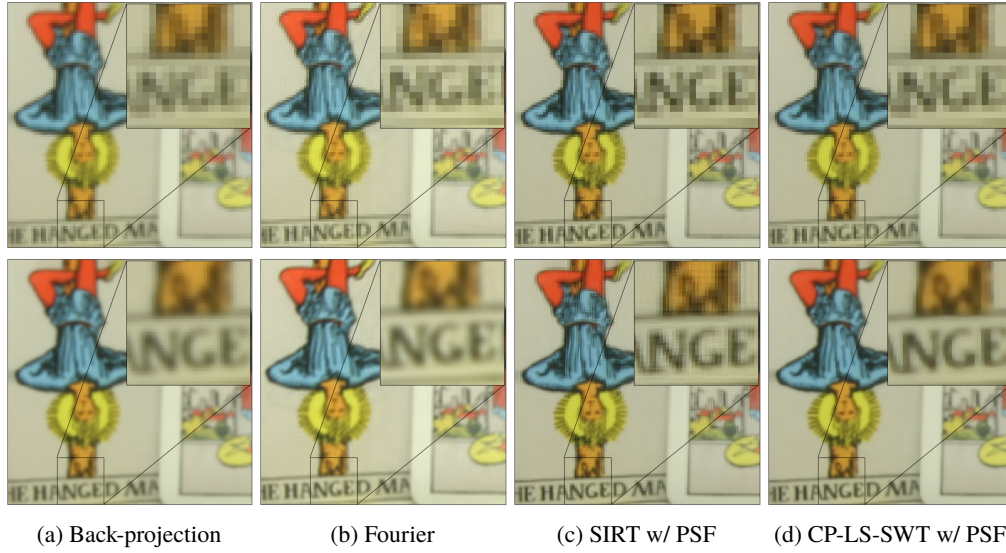


Fig. 11. Performance comparison of different refocusing approaches, at different levels of up-scaling (super-resolution) for the Tarot Cards case. First row shows reconstructions at up-sampling = 1, and second row up-sampling = 4. In the columns, instead, we have the different approaches: (a) Back-projection; (b) Fourier slice theorem; (c) SIRT with PSF; (d) Chambolle-Pock using PSF, l_2 data-divergence term, and SWT Haar regularization with $\lambda = 3$.

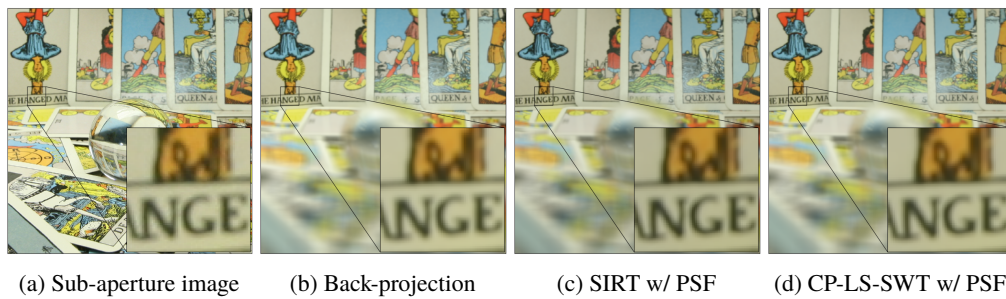


Fig. 12. Performance comparison of different refocusing approaches, for the full-resolution Tarot Cards case. We show one of the sub-aperture images on the side of different reconstruction approaches: (a) Central sub-aperture image; (b) Back-projection; (c) SIRT; (d) Chambolle-Pock using l_2 data-divergence term, and SWT Haar regularization with $\lambda = 3$.

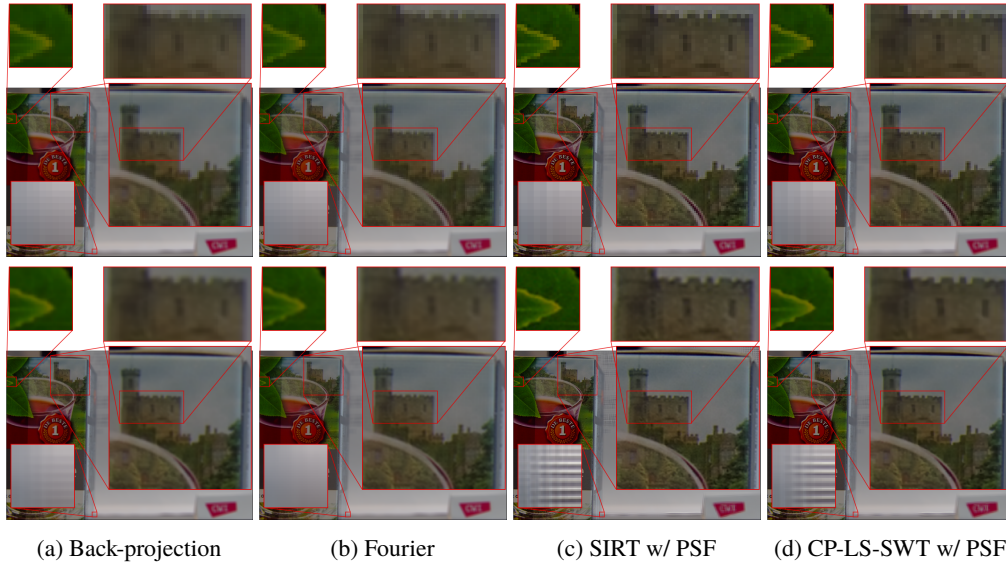


Fig. 13. Performance comparison of different refocusing approaches, at different levels of up-scaling (super-resolution) for the Tea case. First row shows reconstructions at up-sampling = 1, and second row up-sampling = 4. In the columns, instead, we have the different approaches: (a) Back-projection; (b) Fourier slice theorem; (c) SIRT with PSF; (d) Chambolle-Pock using PSF, l_2 data-divergence term, and SWT Haar regularization with $\lambda = 1$.

is visible between the reconstructions. This is due to the fact that, in this case, the back-projection method already returns a correct reconstruction. In fact, the blur visible in the reconstructions is also present in the sub-aperture images (Fig. 12(a)).

These high resolution gantry data reconstructions can be used to quantify the quality of the super-resolved refocusing reconstructions that used the modified Tarots Cards dataset. The measured RMSE values for the reconstructions in Fig. 11, against Fig. 12(b) are the following: 0.0363353, 0.103504, 0.0382393, 0.0169782 for the back-projection, Fourier, SIRT with PSF, and CP-LS with SWT penalization based on Haar wavelets (CP-LS-SWT) and PSF respectively. These results confirm the visual interpretation carried out earlier in this section.

3.5. Tea refocusing

The Lytro Illum plenoptic camera is a popular commercially available compound camera, that allows to easily acquire and extract single-shot light-fields, while maintaining a portable format. In Fig. 13, we show the reconstruction of the Lytro Illum data acquired with the setup in Fig. 5(a). As it was done for the Tarots Cards dataset, we performed reconstructions with up-sampling of 1 and 4, for the following methods: back-projection, Fourier slice theorem, SIRT with PSF, and CP-LS with SWT penalization based on Haar wavelets (CP-LS-SWT) and PSF.

This example shows that the methods developed in this article can be applied to the data acquired with this popular camera. It can be seen from the leaf and castle insets that once again both iterative approaches present a better feature resolving performance than the one-step approaches. Moreover, the SWT regularization is well behaved for the refocusing of natural images, and it helps in reducing artefacts.

In this example, we also focus our attention to the out-of-focus regions of the refocused image. In particular, we refer to the bottom inset, which magnifies an edge from the background. The super-resolved back-projected reconstruction (bottom of Fig. 13(a)) of this region is similar to

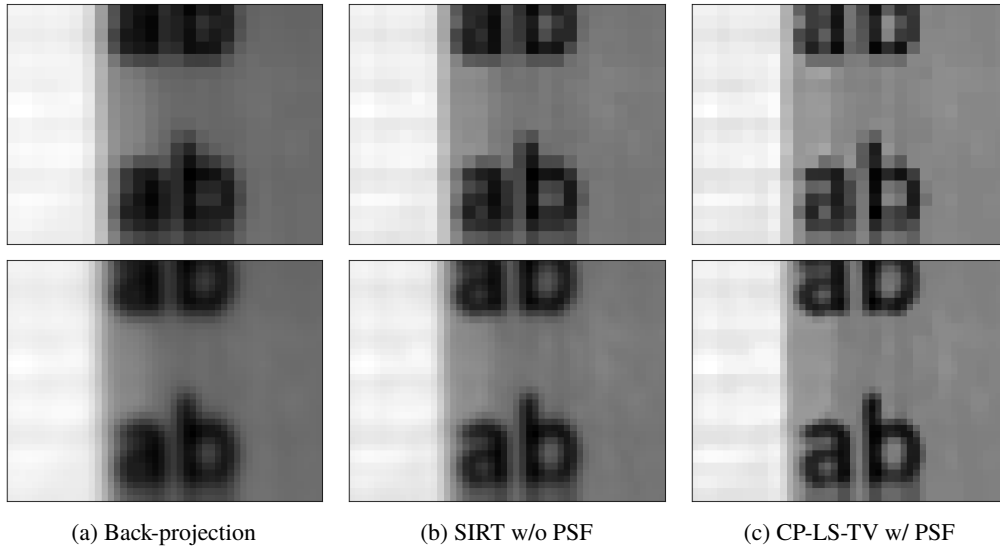


Fig. 14. Performance comparison of different refocusing approaches, at different levels of upscaling (super-resolution) for the Letters case. First row shows reconstructions at up-sampling = 1, and second row up-sampling = 2. In the columns, instead, we have the different approaches: (a) Back-projection; (b) SIRT without PSF; (c) Chambolle-Pock with PSF, l_2 data-divergence term, TV regularization, and $\lambda = 100$.

the corresponding Fourier version (bottom of Fig. 13(b)), but it also presents traces of a wavy pattern. The iterative methods used in Figs. 13(c) and 13(d) present a stronger artifact because they use the same back-projection operation as in Fig. 13(a), but iterate it multiple times. The observed artifact is due to inaccuracies in the numerical interpolation of the used back-projection implementation, which is more apparent for higher relative distances from the acquisition focal distance. In fact, this effect is not visible for the Tarot Cards test case (Fig. 12), and it deviates from the expected bokeh effect for out-of-focus regions. To remove it, the back-projection operation could be substituted by a Fourier type of back-projection.

3.6. Letters and Flower refocusing

The Letters and Flower test cases assess the ability of our approach to handle real-data produced with the FLF camera geometry. We show that every aspect of our approach, including super-resolution and regularization, can be carried over to this case.

In Fig. 14, we present the reconstruction of the FLF data from Fig. 7 for the following approaches: back-projection, SIRT without PSF and CP-LS-TV with PSF. For this case, the choice of the TV term is justified by the properties of the imaged objects: the printed letters, with sharp boundaries over a plain background, explicitly satisfy the prior knowledge of the TV minimization. We present both reconstructions with up-sampling equal to 1, and 2. From Fig. 14, it is clearly visible that the use of super-sampling and regularization techniques can improve the reconstruction of the smallest features.

The PSF used for this case is the result of both the theoretically predicted PSF due to Fraunhofer diffraction of the lenslets, and the Nyquist limitation on the resolution for the given refocusing distance (as seen in section 2.2). To compute the Nyquist PSF size, we used a computed refocusing distance $z \approx 430$ mm, an acquisition focus distance $z_0 = 400$ mm, and (u, v) resolution of ~ 0.175 mm. Figure 14(c), despite looking overall sharper than Fig. 14(b), still presents some blur. This is due to the fact that the PSFs were only theoretically computed, based on

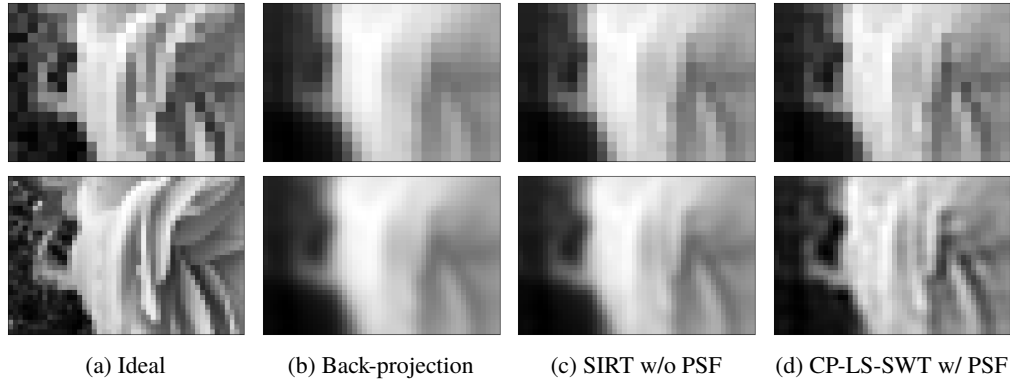


Fig. 15. Performance comparison of different refocusing approaches, at different levels of upscaling (super-resolution) for the flower case. First row shows reconstructions at up-sampling = 1, and second row up-sampling = 2. In the columns, instead, we have the different approaches: (a) Expected reconstruction for an ideal camera with infinite bandwidth optical system response; (b) Back-projection; (c) SIRT without PSF; (d) Chambolle-Pock with PSF, l_2 data-divergence term, SWT regularization, and $\lambda = 5$.

the estimated distances z , z_0 , and the measured b (which in turn determines a , and the (u, v) resolution). For higher reconstruction accuracy, a thorough calibration of the acquisition setup is required (section 2.5), including the experimental determination of the camera PSF.

In Fig. 15, we present the reconstruction of the FLF data from Fig. 8. Similarly to what was done for the Letters test case, we used the following: back-projection, SIRT without PSF and CP-LS-SWT with PSF. The notable difference from the Letters case is the choice of the SWT regularization term, over the TV term. As previously mentioned, natural imagery generally breaks the assumption of flat and sharp regions in the reconstructed images.

In Fig. 15(a), we can see the expected reconstruction with an ideal camera (with infinite bandwidth). However, in practice, as stated in [16], the acquired light-field is convolved with a low pass filter, which has the form of a 4D sinc function. This results in lower than ideal performance for each reconstruction method, but an increasing similarity to Fig. 15(a) indicates increasing reconstruction quality. Moreover, the images in Fig. 15(a), were obtained by downscaling the high resolution image from Fig. 8(a), and not by taking a picture of a printed version, as it was done for Figs. 8(b) and 8(c). As a consequence, especially the higher resolution version represents a purely ideal case, because it was not upsampled from its lower resolution version.

Despite requiring a more thorough calibration of the images and PSF, also in this case, the proposed methods improve on the back-projection approach. In particular, the super-resolution reconstruction using the theoretical PSF and SWT regularization allows to visibly recover more of the expected features, than the other considered approaches.

4. Conclusions and outlook

The proposed tomography based light-field refocusing method is defined in the direct-space, as opposed to previously proposed Fourier space tomography methods. We have shown that this formulation is applicable to both the ULF and FLF camera configuration, and it allows to both easily introduce advanced forward operator modeling, and advanced regularization techniques. This is due to the fact that it reduces the different camera geometries to small differences in the projection implementation, resulting in a unified refocusing strategy for all the available plenoptic acquisition schemes. One of the main results of this aspect is that it provides greater and wider applicability for future developments in the field. As an example, we showed that the introduction

of regularization terms in the refocusing reconstruction could be applied to both the ULF and FLF cases.

The TV and SWT regularization techniques also proved to be powerful tools for reducing the artifacts due to super-resolution reconstructions and PSF corrections. Newer and more advanced regularizations could help to further improve the reconstruction quality and push for higher super-resolution refocusing.

The proposed approach will allow in the future to introduce the optical system's aberration modeling in the forward projection operator, in the same fashion as for the PSF. This would result in yet higher refocusing accuracy and super-resolution. We believe that further effort could be devoted to determining the theoretically best achievable super-resolution for the light-field data and geometry at hand, once the PSF and aberrations have been taken into account. On another note, this article further confirms that computational improvements can help to minimize the impact of physical limitations imposed by the manufacturing of smaller optics and detectors with smaller pixel sizes.

Finally, we showed that the proposed method allows to tune the refocusing problem based on the available prior knowledge, but also that multiple advanced algorithms could be used to solve these problems. This means that the newest developments in the reconstruction algorithms could be ported to light-field refocusing from other fields like computed tomography, alongside with the newest developments in convex optimization and regularizations.

Appendix A

SIRT and CP-LS-TV refocusing implementations

Here we first report the operations needed to implement a SIRT algorithm that solves Eq. (27). Each iteration step can be computed as follows:

$$\begin{aligned} \mathbf{x}_z^{(k+1)} &= \text{pos} \left(\mathbf{x}_z^{(k)} + \mathbf{D}_2 \tilde{\mathbf{A}}_z^T \left(\mathbf{b} - \tilde{\mathbf{A}}_z \mathbf{x}_l^{(k)} \right) \right) \\ \tilde{\mathbf{A}}_z &= \mathbf{P}_z \mathbf{U}_n \mathbf{A}_z; \quad \tilde{\mathbf{A}}_z^T = \mathbf{A}_z^T \mathbf{U}_n^T \mathbf{P}_z^T \\ \mathbf{D}_1 &= \text{diag} \left(\frac{1}{|\tilde{\mathbf{A}}_z| \mathbf{1}} \right); \quad \mathbf{D}_2 = \text{diag} \left(\frac{1}{|\tilde{\mathbf{A}}_z^T| \mathbf{1}} \right) \end{aligned} \quad (30)$$

where the function $\text{pos}(\cdot)$ clips all the negative values to 0, k indicates the iteration step, $\mathbf{1}$ are vectors composed by only values 1 of the same sizes as the vectors \mathbf{x} and \mathbf{b} respectively, and the matrices \mathbf{D}_1 and \mathbf{D}_2 are diagonal matrices used to rescale the modified forward and backward operators $\tilde{\mathbf{A}}_z$ and $\tilde{\mathbf{A}}_z^T$ respectively.

For the solution of Eq. (28), we use the Chambolle-Pock type of algorithms, which are called primal-dual algorithms, because they solve two conjugate problems at the same time: the primal, defined by Eq. (28), and its dual, which is a maximization problem. If the solution exists and it is unique, it will be the same for both problems once the algorithm has reached convergence [27, 37]. We report here a CP preconditioned version of the said CP algorithm with an l_2 data divergence

term (LS), and TV penalization:

$$\begin{aligned}
 \mathbf{x}_z^{(0)} &= \mathbf{0}, \quad \bar{\mathbf{x}}^{(0)} = \mathbf{0}, \quad \mathbf{p}_d^{(0)} = \mathbf{0}, \quad \mathbf{p}_{tv}^{(0)} = \mathbf{0}, \quad \mathbf{D}_2 = \text{diag} \left(\frac{1}{|\tilde{\mathbf{A}}_z^T| \mathbf{1} + 4\lambda} \right) \\
 \text{for } l &:= [0, L) \\
 \mathbf{p}_d^{(l+1)} &= \frac{\mathbf{p}_d^{(l)} + \mathbf{D}_1 (\mathbf{A}_z \bar{\mathbf{x}}^{(l)} - \mathbf{b})}{\text{diag}(\mathbf{1}) + \mathbf{D}_1} \\
 \mathbf{p}_{tv}^{(l+1)} &= \frac{\mathbf{p}_{tv}^{(l)} + 1/2 \nabla \bar{\mathbf{x}}^{(l)}}{\max(\mathbf{1}, |\mathbf{p}_{tv}^{(l)} + 1/2 \nabla \bar{\mathbf{x}}^{(l)}|)} \\
 \mathbf{x}_z^{(l+1)} &= \text{pos} \left(\mathbf{x}^{(l)} - \mathbf{D}_2 \mathbf{A}^T \mathbf{p}_d^{(l+1)} + \lambda \mathbf{D}_2 \text{div}(\mathbf{p}_{tv}^{(l+1)}) \right) \\
 \bar{\mathbf{x}}^{(l+1)} &= \mathbf{x}^{(l+1)} + (\mathbf{x}^{(l+1)} - \mathbf{x}^{(l)})
 \end{aligned} \tag{31}$$

where \mathbf{p}_d and \mathbf{p}_{tv} are the detector and gradient terms of the dual problem respectively, the function $\text{div}(\cdot)$ computes the divergence of its input, the matrices $\tilde{\mathbf{A}}_z$, $\tilde{\mathbf{A}}_z^T$ and \mathbf{D}_1 are the same as for the SIRT algorithm, the matrix \mathbf{D}_2 has been adapted to also incorporate the TV term rescaling, and $\bar{\mathbf{x}}$ is a modified solution \mathbf{x}_z that helps in improving the convergence rate of the algorithm [27].

Funding

European Union's Horizon 2020 research and innovation program (VOXEL H2020-FETOPEN-2014-2015-RIA GA 665207).

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**, 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 the IEEE International Conference on Computer Vision, (IEEE, 2013), pp. 673–680.
4. T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in 2009 IEEE International Conference on Computational Photography (ICCP), (IEEE, 2009), pp. 1–9.
5. S. Wanner and B. Goldluecke, "Spatial and angular variational super-resolution of 4D light fields," Lect. Notes Comput. Sci. (including subseries Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) **7576 LNCS**, 608–621 (2012).
6. K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," ACM Transactions on Graph. **32**, 1 (2013).
7. Y. Wang, G. Hou, Z. Sun, Z. Wang, and T. Tan, "A simple and robust super resolution method for light field images," in 2016 IEEE International Conference on Image Processing (ICIP), (IEEE, 2016), pp. 1459–1463.
8. Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Light-Field Image Super-Resolution Using Convolutional Neural Network," IEEE Signal Process. Lett. **24**, 848–852 (2017).
9. M. Rossi and P. Frossard, "Geometry-Consistent Light Field Super-Resolution via Graph-Based Regularization," IEEE Transactions on Image Process. **27**, 4207–4218 (2018).
10. W. S. Chan, E. Y. Lam, M. K. Ng, and G. Y. Mak, "Super-resolution reconstruction in a computational compound-eye imaging system," Multidimens. Syst. Signal Process. **18**, 83–101 (2007).
11. 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**, 25418–25439 (2013).
12. S. A. Shroff and K. Berkner, "Image formation analysis and high resolution image reconstruction for plenoptic imaging systems," Appl. Opt. **52**, D22 (2013).
13. A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in IEEE International Conference on Computational Photography (ICCP), (IEEE, 2009), pp. 1–8.
14. T. G. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," J. Electron. Imaging **19**, 021106 (2010).
15. 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 (2005).
16. R. Ng, "Fourier slice photography," ACM Transactions on Graph. **24**, 735 (2005).

17. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear Volumetric Focus for Light Field Cameras," *ACM Transactions on Graph.* **34**, 1–20 (2015).
18. T. Georgiev, G. Chunev, and A. Lumsdaine, "Superresolution with the focused plenoptic camera," in *Computational Imaging IX*, C. A. Bouman, I. Pollak, and P. J. Wolfe, eds. (2011).
19. C. Herzog, O. de La Rochefoucauld, G. Dovillaire, X. Granier, F. Harms, X. Levecq, E. Longo, L. Mignard-Debise, and P. Zeitoun, "Comparison of reconstruction approaches for plenoptic imaging systems," in *Unconventional Optical Imaging*, C. Fournier, M. P. Georges, and G. Popescu, eds. (SPIE, 2018), May, p. 104.
20. N. Viganò, H. Der Sarkissian, C. Herzog, O. de la Rochefoucauld, R. van Liere, and K. J. Batenburg, "Tomographic approach for the quantitative scene reconstruction from light field images," *Opt. Express* **26**, 22574 (2018).
21. M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr., "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Process. Lett.* **3**, 10–12 (1996).
22. A. K. Louis, P. Maass, and A. Rieder, *Wavelets: Theory and Applications*, Pure and Applied Mathematics (Wiley, 1997).
23. S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis Mach. Intell.* **24**, 1167–1183 (2002).
24. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, 1988).
25. A. V. der Sluis and H. V. der Vorst, "SIRT-and CG-type methods for the iterative solution of sparse linear least-squares problems," *Linear Algebr. its Appl.* (1990).
26. A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.* **40**, 120–145 (2010).
27. 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**, 3065–3091 (2012).
28. 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**, A1 (2018).
29. T. M. Buzug, *Computed Tomography* (Springer-Verlag, 2008).
30. M. Martínez-Corral and B. Javidi, "Fundamentals of 3d imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems," *Adv. Opt. Photonics* **10**, 512 (2018).
31. G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 28 (Johns Hopkins University, 1996).
32. J. W. Goodman, *Introduction to Fourier Optics*, Electrical and Computer Engineering: Communications and Signal Processing (McGraw-Hill, 1996), 2nd ed.
33. 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**, 250–253 (2011).
34. G. Kutyniok, *Compressed Sensing* (Cambridge University, 2012).
35. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2013), pp. 1027–1034.
36. A. Lumsdaine, T. G. Georgiev, and G. Chunev, "Spatial analysis of discrete plenoptic sampling," *Proc. SPIE* **8299**, 829909 (2012).
37. S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University, 2004).