



**HAL**  
open science

## A Fourier Disparity Layer representation for Light Fields

Mikaël Le Pendu, Christine Guillemot, Aljosa Smolic

► **To cite this version:**

Mikaël Le Pendu, Christine Guillemot, Aljosa Smolic. A Fourier Disparity Layer representation for Light Fields. IEEE Transactions on Image Processing, In press, pp.5740 - 5753. 10.1109/TIP.2019.2922099 . hal-02130555

**HAL Id: hal-02130555**

**<https://hal.science/hal-02130555>**

Submitted on 15 May 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Fourier Disparity Layer representation for Light Fields

Mikael Le Pendu, Christine Guillemot, and Aljosa Smolic

**Abstract**—In this paper, we present a new Light Field representation for efficient Light Field processing and rendering called Fourier Disparity Layers (FDL). The proposed FDL representation samples the Light Field in the depth (or equivalently the disparity) dimension by decomposing the scene as a discrete sum of layers. The layers can be constructed from various types of Light Field inputs including a set of sub-aperture images, a focal stack, or even a combination of both. From our derivations in the Fourier domain, the layers are simply obtained by a regularized least square regression performed independently at each spatial frequency, which is efficiently parallelized in a GPU implementation. Our model is also used to derive a gradient descent based calibration step that estimates the input view positions and an optimal set of disparity values required for the layer construction. Once the layers are known, they can be simply shifted and filtered to produce different viewpoints of the scene while controlling the focus and simulating a camera aperture of arbitrary shape and size. Our implementation in the Fourier domain allows real time Light Field rendering. Finally, direct applications such as view interpolation or extrapolation and denoising are presented and evaluated.

**Index Terms**—Light Fields, Fourier domain, rendering, refocusing, view interpolation, denoising.

## I. INTRODUCTION

Light Fields are commonly represented as 4 dimensional functions with 2 spatial and 2 angular dimensions [1], [2]. They can be seen as 2D arrays of images (called sub-aperture images), each having an unlimited depth of field, and differing from their neighbour images only by a slight shift of the view angle. The sampling in the angular dimensions is key in Light Field imaging [3]. In particular, densely sampled Light Fields make it possible to directly render images with shallow depth of field while controlling the focus depth. Such rendering, often referred to as Light Field refocusing, does not require knowledge of the scene’s geometry. It is usually performed either by shifting and averaging the sub-aperture images [4] or by selecting a 2D slice in the 4D Fourier domain [5]. However, a dense angular sampling comes at the expense of very high requirements in terms of capture, storage, processing power and memory. A too sparse angular sampling, on the other hand, does not allow for a smooth transition between viewpoints and causes angular aliasing in the refocused images, characterized by sharp structures in the out of focus regions. The importance of a dense angular sampling is clearly shown by the vast literature on viewpoint interpolation. Several approaches exist including depth image

This project has been supported in part by the Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776 and in part by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM).

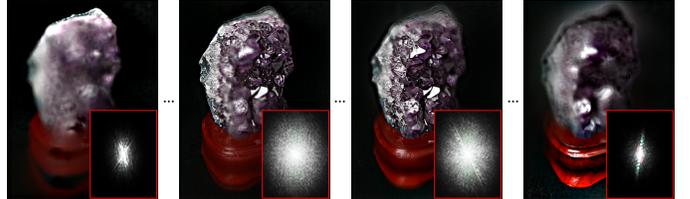


Fig. 1. Fourier Disparity Layer representation. For the visualization, the layers are shown in the spatial domain (i.e. after inverse Fourier Transform). The magnitude spectrum of each layer is also shown in the red boxes. Note that the FDL representation differs visually from a focal stack: the out-of-focus regions tend to disappear while the regions in focus have enhanced contrasts.

based rendering techniques [6]–[11], deep learning methods either exploiting a depth map estimation [12], [13] or not [14]–[16], and approaches leveraging sparsity priors of the Light Field data in a transformed domain [17]–[20]. However, although viewpoint interpolation greatly simplifies the capture of dense Light Fields, it also increases the amount of data to store and process for the final rendering application.

Alternatively, Light Fields can be represented as a focal stack, that is, a set of shallow depth of field images (e.g. photos taken with a wide aperture) with different focusing depths. This representation has the advantage of allowing an unlimited angular density with few images because the sampling is performed on the depth dimension instead of 2 angular dimensions. However, for rendering tasks such as simulating a different camera aperture size or shape, or a change of viewpoint, the common approach is to first convert the focal stack into the 4D representation. For instance, Levin and Durand [21] retrieve sub-aperture images by the deconvolution of shifted and averaged focal stack images. A similar deconvolution technique is used in [22] to first synthesize the 4D Light Field from a focal stack in order to render images with arbitrary aperture shapes. More recent methods have also been proposed to reconstruct the 4D Light Field from a focal stack either in the spatial domain using depth from focus [23], or via optimization in the Fourier domain [24], [25].

The main motivation of this work is to simplify the handling of Light Fields through the definition of a compact representation which, unlike focal stacks, can be directly used for any Light Field rendering task. In this aim, we propose the Fourier Disparity Layer (FDL) representation illustrated in Fig. 1. It can be easily constructed either from a 4D Light Field, a focal stack, or even a hybrid Light Field combining sub-aperture and wide aperture images with varying parameters (i.e. focusing depth, aperture, point of view). Given the 2D Fourier transform of each input image with their parameters (i.e. angular coordinates, aperture, focus), the FDL model

is constructed using linear optimization. For each frequency component, a linear least squares problem is solved to determine the corresponding Fourier coefficients of the different layers, each layer being associated to a given disparity value. The layers can then be directly used for real time rendering. For instance, sub-aperture images are obtained by shifting the layers proportionally to their associated disparity value and by averaging them. This is directly implemented in the Fourier domain as a simple linear combination with frequency-dependent coefficients. More general rendering with arbitrary point of view, aperture shape and size, and focusing depth is performed with the same computational complexity without the need to first reconstruct the 4D Light Field.

In the case where the input is a set of sub-aperture images, we propose a gradient descent based calibration method to determine their angular coordinates as well as the optimal set of disparity values. The formulation of the optimization problems for the calibration and the layer construction are closely related. Nevertheless, we define two regularization schemes with different properties to better suit each situation.

Additionally, we demonstrate the effectiveness of our approach for several direct applications. First, when the input is a sparse set of sub-aperture images, view interpolation and extrapolation is obtained by constructing the FDL representation and by rendering views at intermediate angular coordinates with an infinitely small aperture. In a second application, the same viewpoints as the input are rendered to produce a denoised result. For this use case, we present a possible extension of the model where the shift applied to each layer is not constrained to be proportional to the angular coordinates of the view to reconstruct. The relaxed model allows a more accurate representation of occlusions and non-Lambertian effects in the scene.

Since the computational complexity is a key aspect motivating the need for a new Light Field representation, our implementation makes efficient use of the GPU at every step of the processing chain (i.e. calibration, layer construction, rendering). The proposed algorithms are built upon simple linear algebra operations performed independently at each frequency component, which makes our approach particularly suitable for GPU parallelization.

In summary, the contributions are:

- Definition of the Fourier Disparity Layer representation and its construction from other Light Field representations (e.g. sub-aperture images, focal stack, combination of focal stack images and sub-aperture images).
- Calibration method jointly determining the input view positions and disparity values of the layers.
- Fast and advanced Light Field rendering from the FDL representation with simultaneous control over the viewpoint, aperture size, aperture shape and focusing depth.
- Analysis of other application scenarios: view interpolation and denoising.

## II. RELATED WORK

Related Light Field representations have been used in the design of several Light Field displays [26]–[28]. These

displays reproduce the Light Field using a stack of light attenuating LCD layers placed in front of a backlight. Thanks to the distance separating the LCD panels in the display, the image perceived depends on the observer’s position and is proportional to the product of the layers. This layer representation has similarities with the one presented in this paper, and it can be constructed either from the Light Field views [26], [27] or from a focal stack [28]. The main difference however, is that, because of the physics of the light attenuating LCD layers, the sub-aperture images of the Light Field are reconstructed as a product of the layers’ pixels instead of a sum. Hardware limitations also impose constraints on the layer representation. For instance, the number of layers is generally small (e.g. 3 to 5), which is often insufficient to accurately represent the whole Light Field. Furthermore, the layers must only have positive pixel values in order to be displayed on the LCD panels. This constraint is not required in our model, which allows us to efficiently construct the layers in the Fourier domain.

Similarly to the FDL method proposed in this paper, Alonso et al. [25] construct layers by an optimization in the Fourier domain. However, their method is limited to a focal stack input. In this configuration, the problem is well conditioned because the input images already contain dense angular information and each constructed layer is associated to the focusing depth of one of the focal stack images. Hence, no regularization scheme was considered for this application. The method we propose is more generic as it can also construct the layers from sub-aperture images. Therefore, specific regularization strategies are studied, which allows us to address a much larger range of applications including calibration, view interpolation, denoising, etc.

Finally, the proposed FDL representation directly relates to the dimensionality gap Light Field prior described by Levin and Durand [21]. It states that the support of the Light Field data in the 4D Fourier domain is a 3D manifold which was later characterized as a hypercone in [29]. By additionally considering the limited depth range of a scene, Dansereau et al. [29] determined that the frequency-domain support of the Light Field forms a hyperfan. They define this shape as the intersection of the hypercone with a dual fan previously described in [30]. In this paper, we derive the FDL representation from the dimensionality gap prior assuming a discrete set of depths instead of a continuous range. For the discrete depth case, we show formally in Section IV-A that this prior is itself derived from the assumption of a non-occluded Lambertian scene. This is a limitation for any method directly enforcing the dimensionality gap prior. For instance, as observed in [21], [29], in the reconstructed sub-aperture images, occluding objects may appear transparent near the occlusion boundaries. However, semi-transparent objects and reflections on flat surfaces are accurately reproduced, which is particularly challenging for depth image based rendering methods. We also present in Section V-C a possible generalization of our layer model to allow a better representation of other non-Lambertian effects and occlusions.

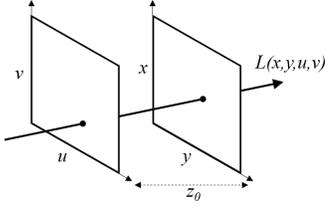


Fig. 2. Two-plane parameterization. The focal plane  $(x, y)$  is parallel to the camera plane  $(u, v)$  and located at a distance  $z_0$ .

### III. LIGHT FIELD NOTATIONS

Let us first consider the 4D representation of Light Fields in [1] and Lumigraph in [2] parameterized with two parallel planes, as illustrated in Fig. 2. The 4D representation describes the radiance along rays by a function  $L(x, y, u, v)$  where the pairs  $(x, y)$  and  $(u, v)$  respectively represent spatial and angular coordinates. For simplicity of notation, we consider a 2D Light Field  $L(x, u)$  with one spatial dimension  $x$  and one angular dimension  $u$ , but the generalization to a 4D Light Field  $L(x, y, u, v)$  is straightforward.

In this paper, we use the notion of disparity instead of depth. Given the depth  $z_0$  of the focal plane in Fig. 2, a depth  $z$  can be directly converted into a disparity  $d$  with  $d = \frac{z_0 - z}{z}$  (i.e. objects at depth  $z_0$  from the camera plane have zero disparity). Refocusing the Light Field then consists in defining a new Light Field  $L'(x, u) = L(x - us, u)$ . The refocus parameter  $s$  is defined such that the regions of disparity  $d = s$  in the Light Field  $L$ , have a disparity equal to zero in the Light Field  $L'$ . A refocused image, noted  $B^s$ , is then formed by refocusing the Light Field with parameter  $s$  and integrating the light rays over the angular dimension as described in [4]:

$$B^s(x) = \int_{\mathbb{R}} L(x - us, u) \psi(u) du. \quad (1)$$

where  $\psi(u)$  represents the aperture of the imaging system. In the case where  $B^s$  is captured by a camera with an aperture area  $\mathcal{A}$  (typically a disk), then  $\psi$  is defined by:

$$\psi(u) = \begin{cases} 1 & \text{if } u \in \mathcal{A} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

In this paper, we use a more general version of Eq. (1) where the image can be observed at any position  $u_0$  on the camera plane with any given aperture  $\psi$ :

$$B_{u_0}^s(x) = \int_{\mathbb{R}} L(x - us, u_0 + u) \psi(u) du. \quad (3)$$

Note that when the aperture area is infinitely small, the function  $\psi$  is equal to the dirac delta function  $\delta$ . In this case, the image formed by Eq. (3) is the sub-aperture image noted  $L_{u_0}$  and defined by  $L_{u_0}(x) = L(x, u_0)$ .

The different notations used in the article are summarized in Table I.

TABLE I  
TABLE OF NOTATION

Symbols	Description
$L(x, u)$	Light Field with spatial coordinate $x$ and angular coordinate $u$ .
$\omega_x, \omega_u$	Respectively spatial and angular frequencies.
$L_{u_0}$	Sub-aperture image at position $u_0$ ( $L_{u_0}(x) = L(x, u_0)$ ).
$B_{u_0}^s$	Image with refocus parameter $s$ and position $u_0$ .
$\psi$	Aperture function.
$u_0$	Angular coordinate of the view to reconstruct.
$u_l$	Angular coordinate of a known input view.
$d_k$	Disparity value in the Light Field.
$\Omega_k^u$	Region of disparity $d_k$ in an image of view position $u$ .
$\delta$	Dirac delta function
$\hat{f}$	Fourier transform of a function $f$ .
$\mathbf{A}^\top$	Transpose of the matrix $\mathbf{A}$ .
$\overline{\mathbf{A}}$	Complex conjugate of $\mathbf{A}$ (without transpose).
$\mathbf{A}^*$	Conjugate transpose of $\mathbf{A}$ ( $\mathbf{A}^* = \overline{\mathbf{A}^\top}$ ).
$\mathbf{A}_{j,k}$	Element of $\mathbf{A}$ on the $j^{\text{th}}$ row and $k^{\text{th}}$ column.
$\mathbf{A}_k$	$k^{\text{th}}$ column of matrix $\mathbf{A}$ .
$\mathbf{x}_k$	$k^{\text{th}}$ element of vector $\mathbf{x}$ .

### IV. FOURIER DISPARITY LAYER REPRESENTATION

#### A. Light Field Prior and FDL Representation

For the derivations, we assume that the scene is Lambertian, without occlusion, and can be divided into  $n$  spatial regions  $\Omega_k$  with constant disparity  $d_k$ . Formally, this can be written:

$$\forall k \in \llbracket 1, n \rrbracket, \forall (x, u) \in \Omega_k \times \mathbb{R}, L(x - ud_k, u) = L(x, 0). \quad (4)$$

Here, the spatial regions  $\Omega_k$  are defined for the central view at  $u = 0$ . From this assumption, we prove in Appendix A that the Fourier transform of the Light Field can be decomposed as follows:

$$\hat{L}(\omega_x, \omega_u) = \sum_k \delta(\omega_u - d_k \omega_x) \hat{L}^k(\omega_x), \quad (5)$$

where  $\hat{L}^k$  is defined by

$$\hat{L}^k(\omega_x) = \int_{\Omega_k} e^{-2i\pi x \omega_x} L(x, 0) dx. \quad (6)$$

Each function  $\hat{L}^k$  can be interpreted as the Fourier transform of the central view only considering the region  $\Omega_k$  of disparity  $d_k$ . Hence, we call these functions Fourier Disparity Layers (FDL). One can note from Eq. (5) that the Light Field information is entirely contained in the Fourier Disparity Layers and their associated disparity values  $d_k$ . We will show that the layers  $\hat{L}^k$  can be constructed either from sub-aperture or wide aperture images without knowing the regions  $\Omega_k$ . In other words, our method does not require a disparity map. However, the disparity values  $d_k$  are necessary. Hence, a method for estimating these values is also presented in Section V-B.

Another interpretation of Eq. (5) is that the Light Field is sparse in the Fourier domain and the non-zero values are located at frequencies such that  $\omega_u = d_k \omega_x$  for each disparity value  $d_k$  of the Light Field. This is similar to the Light Field prior referred to as dimensionality gap prior in [21], and resulting in the definition of the hyperfan filter in

[29]. However, these previous works consider a continuous disparity range, which does not allow for a practical layer representation, since it would result in an infinite number of layers.

### B. FDL Decomposition of Images

*Sub-aperture images:* Here, we derive the relationship between the FDL representation of the Light Field and the Fourier Transform  $\hat{L}_{u_0}$  of a sub-aperture image  $L_{u_0}$ . Note that in  $\hat{L}_{u_0}$ , the Fourier Transform only applies to the spatial dimension. Hence,  $\hat{L}_{u_0}$  is obtained from the Light Field's spectrum  $\hat{L}$  by applying the inverse Fourier transform in the angular dimension as follows:

$$\hat{L}_{u_0}(\omega_x) = \int_{-\infty}^{+\infty} e^{+2i\pi u_0 \omega_u} \hat{L}(\omega_x, \omega_u) d\omega_u. \quad (7)$$

Using the FDL decomposition of  $\hat{L}(\omega_x, \omega_u)$  from Eq. (5), the transformed sub-aperture image can be directly derived:

$$\hat{L}_{u_0}(\omega_x) \quad (8)$$

$$= \sum_k \hat{L}^k(\omega_x) \int_{-\infty}^{+\infty} e^{+2i\pi u_0 \omega_u} \delta(\omega_u - d_k \omega_x) d\omega_u \quad (9)$$

$$= \sum_k e^{+2i\pi u_0 d_k \omega_x} \hat{L}^k(\omega_x). \quad (10)$$

*General image model:* Now, considering the general model in Eq. (3) the Fourier Transform of an image  $B_{u_0}^s$  is:

$$\begin{aligned} \hat{B}_{u_0}^s(\omega_x) &= \int_{-\infty}^{+\infty} e^{-2i\pi x \omega_x} \left[ \int_{-\infty}^{+\infty} L(x - us, u_0 + u) \psi(u) du \right] dx \\ & \quad (11) \end{aligned}$$

$$= \int_{-\infty}^{+\infty} \psi(u) \int_{-\infty}^{+\infty} e^{-2i\pi(x+us)\omega_x} L(x, u_0 + u) dx du \quad (12)$$

$$= \int_{-\infty}^{+\infty} e^{-2i\pi u s \omega_x} \psi(u) \hat{L}_{u_0+u}(\omega_x) du. \quad (13)$$

Using the FDL decomposition of  $\hat{L}_{u_0+u}$  from Eq. (10), we obtain:

$$\begin{aligned} \hat{B}_{u_0}^s(\omega_x) &= \int_{-\infty}^{+\infty} e^{-2i\pi u s \omega_x} \psi(u) \sum_k e^{+2i\pi(u_0+u)d_k \omega_x} \hat{L}^k(\omega_x) du \\ & \quad (14) \end{aligned}$$

$$= \sum_k e^{+2i\pi u_0 d_k \omega_x} \hat{L}^k(\omega_x) \int_{-\infty}^{+\infty} e^{+2i\pi u(d_k-s)\omega_x} \psi(u) du \quad (15)$$

$$= \sum_k e^{+2i\pi u_0 d_k \omega_x} \hat{\psi}(\omega_x(s - d_k)) \cdot \hat{L}^k(\omega_x). \quad (16)$$

One can easily verify that in the particular case of an infinitely small aperture such that  $\psi = \delta$ , we have  $\hat{\psi} = 1$ , thus Eq. (16) becomes equivalent to the sub-aperture case in Eq. (10).

## V. CONSTRUCTION OF THE LAYERS

The layer construction is performed from a set of  $m$  images noted  $\{B_j\}_{j \in \llbracket 1, m \rrbracket}$  that follow the model of Eq. (3) (e.g. images captured by cameras on the same plane and oriented perpendicularly to this plane). Each image  $B_j$  is associated with its angular coordinate  $u_j$ , aperture function  $\psi_j$ , and refocus parameter  $s_j$ . Thanks to the genericity of this model, different forms of input data may be used. For example, in the case of a set of sub-aperture images, the aperture function  $\psi_j$  of each image is equal to the dirac delta function  $\delta$ . Note that, in this case, the refocus parameter  $s_j$  has no influence in the equations (e.g. see Eq. (10)) and is not required. Another notable example is that of a focal stack input (i.e. images with the same angular coordinates  $u_j$  and aperture functions  $\psi_j$ , but with different refocus parameters  $s_j$ ).

The layer construction problem is formulated in the general case in the next sub-section, assuming all the image parameters are known. Then, in sub-section V-B, we show how to determine the positions  $u_j$  of the views and the layer's disparity values  $d_k$  when the input data is a set of sub-aperture images.

### A. Problem Formulation

The results of Eqs. (10) and (16) show that for a fixed spatial frequency  $\omega_x$  in the Fourier domain, the images  $B_j$  can be simply decomposed as linear combinations of the Fourier Disparity Layers. Therefore, the FDL representation of the Light Field can be learned by linear regression for every coefficient in the discrete Fourier domain.

We first compute the discrete Fourier Transform  $\hat{B}_j$  of each image  $B_j$  and we construct, for each frequency component  $\omega_x$ , a vector  $\mathbf{b}$  such that  $\mathbf{b}_j = \hat{B}_j(\omega_x)$ . Given the disparity value  $d_k$  of each layer ( $k \in \llbracket 1, n \rrbracket$ ), we also construct the matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$  with:

$$\mathbf{A}_{jk} = e^{+2i\pi u_j d_k \omega_x} \hat{\psi}_j(\omega_x(s_j - d_k)). \quad (17)$$

By defining the vector  $\mathbf{x}$  with  $\mathbf{x}_k = \hat{L}^k(\omega_x)$ , the FDL decomposition in Eq. (16) is reformulated as  $\mathbf{A}\mathbf{x} = \mathbf{b}$ . A simple layer construction method then consists in solving an ordinary linear least squares problem independently for each frequency component  $\omega_x$ . In practice, however, this may be an ill-posed problem. It is typically the case when the number of available images is lower than the number of layers required to represent the scene (i.e.  $m < n$ ). More generally, depending on the input configuration, and for some frequency components, the matrix  $\mathbf{A}$  may be ill-conditioned. This results in over-fitting and extreme noise amplification when the layer model is used to render new images of the scene (e.g. view interpolation or extrapolation). In order to avoid this situation, we include a Tikhonov regularization term in the problem formulation:

$$\mathbf{x} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{\Gamma}\mathbf{x}\|_2^2, \quad (18)$$

where  $\mathbf{\Gamma}$  is the Tikhonov matrix, and  $\lambda$  is a parameter controlling the amount of regularization. The problem (18), has a well-known closed form solution:

$$\mathbf{x} = (\mathbf{A}^* \mathbf{A} + \lambda \mathbf{\Gamma}^* \mathbf{\Gamma})^{-1} \mathbf{A}^* \mathbf{b}, \quad (19)$$

where  $*$  is the Hermitian transpose operator (i.e. complex conjugate of the transposed matrix).

In our implementation of the layer construction, the Tikhonov matrix is defined according to the  $2^{nd}$  order view regularization scheme presented in subsection V-D1 (see Eq. (29)). It encourages smooth variations between close viewpoints generated from the FDL model, which is intuitively a desirable property for a Light Field.

### B. FDL Calibration

We additionally propose a calibration method that jointly estimates the layers' disparity values  $d_k$  and the angular coordinates  $u_j$  of the input images  $B_j$ . For simplicity, the calibration is restricted to the case of sub-aperture images with infinitely small aperture such that  $\psi_j = \delta$ . In this case, Eq. (17) has a simpler expression  $\mathbf{A}_{jk} = e^{+2i\pi u_j d_k \omega_x}$ . Note that in the general configuration where the aperture is unknown, both the aperture functions  $\psi_j$  and the refocus parameters  $s_j$  would also need to be estimated. However, that generalization is out of the scope of this paper and is left for future work.

In what follows we express  $\mathbf{A}$  as a matrix function  $\mathbf{A}: \mathbb{R}^{m \times n} \rightarrow \mathbb{C}^{m \times n}$  such that, for a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{A}(\mathbf{M})_{j,k} = e^{+2i\pi \mathbf{M}_{j,k}}$ . The calibration problem can then be stated in a similar way as the layer construction problem (18) by treating the calibration parameters as unknowns in the minimization. However, these parameters do not depend on the frequency. Hence, the function to minimize is expressed as a sum over the  $Q$  frequency components  $\omega_x^q$  ( $Q$  is equal to the number of pixels in each input image):

$$\min_{\mathbf{u}, \mathbf{d}} \sum_{q=1}^Q \left( \|\mathbf{A}(\omega_x^q \mathbf{u} \mathbf{d}^\top) \mathbf{x}^q - \mathbf{b}^q\|_2^2 + \lambda \|\Gamma \mathbf{x}^q\|_2^2 \right), \quad (20)$$

where the input view positions  $u_j$  and the disparity values  $d_k$  are arranged in the column vectors  $\mathbf{u}$  and  $\mathbf{d}$  respectively ( $\mathbf{u} \mathbf{d}^\top \in \mathbb{R}^{m \times n}$ ). The vectors  $\mathbf{x}^q$  and  $\mathbf{b}^q$  contain the Fourier coefficients of, respectively, the disparity layers and the input images at the frequency  $\omega_x^q$  (i.e.  $\mathbf{x}_k^q = \hat{L}^k(\omega_x^q)$  and  $\mathbf{b}_j^q = \hat{B}_j(\omega_x^q)$ ). Unlike the layer construction problem, the matrix  $\Gamma$  is defined according to the  $2^{nd}$  order layer regularization approach detailed in subsection V-D2 (see Eq. (30)).

In order to solve this problem, we perform a gradient descent along the vectors of parameters  $\mathbf{u}$  and  $\mathbf{d}$ . At each iteration, the current estimate of  $\mathbf{u}$  and  $\mathbf{d}$  is first used to update the layers values in each vector  $\mathbf{x}^q$  using Eq. (19). The layers values are then used to compute the gradients  $\nabla \mathbf{u}$  and  $\nabla \mathbf{d}$  of the objective function in Eq. (20) along  $\mathbf{u}$  and  $\mathbf{d}$  respectively.

By differentiation with respect to each element of  $\mathbf{u}$  and  $\mathbf{d}$ , one can show that the gradients are expressed as:

$$\nabla \mathbf{u}_j = \sum_k \mathbf{d}_k \nabla \mathbf{P}_{j,k} \quad (21)$$

$$\nabla \mathbf{d}_k = \sum_j \mathbf{u}_j \nabla \mathbf{P}_{j,k}, \quad (22)$$

with  $\mathbf{P} = \mathbf{u} \mathbf{d}^\top$ , and each column  $k$  of the corresponding gradient matrix  $\nabla \mathbf{P}$  is:

$$\nabla \mathbf{P}_k = 4\pi \sum_q \text{Im} \left( \omega_x^q \mathbf{x}_k^q \cdot \overline{\mathbf{A}(\omega_x^q \mathbf{P})_k} \circ (\mathbf{A}(\omega_x^q \mathbf{P}) \mathbf{x}^q - \mathbf{b}^q) \right), \quad (23)$$

where  $\circ$  is the Hadamard product (i.e. element-wise multiplication), and  $\text{Im}$  is the imaginary part. Note that the matrix  $\mathbf{A}(\omega_x^q \mathbf{P})$  was previously constructed to determine  $\mathbf{x}^q$ , and can thus be re-used in Eq. (23) to efficiently compute the gradients.

The computations are further accelerated by performing a stochastic gradient descent, where only a small subset of the  $Q$  frequency components  $\omega_x^q$  is selected randomly at each iteration for the gradients computation. In our implementation, subsets of 4096 frequency components were selected. Therefore, the computational cost per iteration does not depend on the image resolution.

Finally, given the gradients, the updated vectors  $\mathbf{u}'$  and  $\mathbf{d}'$  are then computed as

$$\mathbf{u}' = \mathbf{u} - \alpha \frac{\nabla \mathbf{u}}{\epsilon + \|\nabla \mathbf{u}\|_2}, \text{ and } \mathbf{d}' = \mathbf{d} - \alpha \frac{\nabla \mathbf{d}}{\epsilon + \|\nabla \mathbf{d}\|_2}, \quad (24)$$

where  $\epsilon$  is a small value encouraging the stability of the algorithm when the gradients become small (i.e. when  $\mathbf{u}$  and  $\mathbf{d}$  are close to an optimum). In our experiments a fixed value  $\alpha = 0.2$  was used.

### C. Calibration of a Relaxed FDL Model

Due to the assumptions of non-occluded Lambertian Light Field in the problem definition, the FDL representation may introduce distortions in occluded regions or for strong non-Lambertian effects. In order to better cope with this limitation, we propose a relaxation of the FDL model. In the original FDL calibration (see previous subsection), the parameters in  $\mathbf{u}$  and  $\mathbf{d}$  form a parameter matrix  $\mathbf{P} = \mathbf{u} \mathbf{d}^\top$ , where each element  $\mathbf{P}_{j,k}$  corresponds to the shift applied to the  $k^{th}$  layer to approximate the  $j^{th}$  input sub-aperture image. Since  $\mathbf{u}$  and  $\mathbf{d}$  are column vectors, the rank of the shift parameter matrix  $\mathbf{P}$  is equal to 1. Hence, the shift applied to each layer is proportional to the angular coordinates of the view to reconstruct. However, this property is not suitable for non-Lambertian surfaces and occlusions. Therefore, in the relaxed FDL model, we remove the rank 1 constraint by searching directly for the matrix  $\mathbf{P}$  instead of  $\mathbf{u}$  and  $\mathbf{d}$  in the calibration algorithm. For that purpose, the gradient descent is applied using the gradients  $\nabla \mathbf{P}$  defined in Eq. (23). In order to obtain satisfying convergence, the matrix  $\mathbf{P}$  is initialized by the solution of the original rank 1 constrained problem.

The advantage of the relaxed model for handling occlusions or strongly non-Lambertian effects is illustrated in Section VIII-B for the Light Field denoising application. Note that, for this application, we only need to reconstruct the input views for which the shift parameters were obtained in the matrix  $\mathbf{P}$ . Hence, at each frequency  $\omega_x^q$ , the Fourier coefficients of the reconstructed views are given by  $\mathbf{A}(\omega_x^q \mathbf{P}) \mathbf{x}^q$ . However, for more general rendering tasks (e.g. view interpolation, refocusing), the rank 1 constraint is necessary in order to apply the rendering equation (16). Therefore, in the rest of the paper, we consider the original FDL model unless otherwise stated.

### D. Regularization Schemes

In order to keep the layer construction and calibration problems in Eqs. (18) and (20) easy to solve, we have used a Tikhonov regularization. The definition of the Tikhonov matrix  $\Gamma$  depends on the intended objective. In the most basic form, classical  $l_2$  regularization is obtained by simply taking  $\Gamma$  equal to the identity matrix  $\mathbf{I}_n$ . This prevents the values in  $\mathbf{x}$  (i.e. Fourier coefficients of the disparity layers) from taking too high values, which reduces the noise, but also results in a loss of details. Furthermore, for the calibration problem, using the  $l_2$  regularization may not accurately estimate the disparity distribution of the Light Field as illustrated by the calibration results of Fig. 3. The sorted disparity values shown in Fig. 3(b) can be interpreted as an estimation of the cumulative disparity distribution.

In this section, we present two schemes referred to as  $2^{nd}$  order view regularization and  $2^{nd}$  order layer regularization which are better suited to the layer construction and the calibration respectively.

1) *2<sup>nd</sup> Order View Regularization*: For the layer construction, we want to encourage smooth variations between close viewpoints generated from the layers. For that purpose, at each frequency, we penalize the second derivative of sub-aperture images generated from the model with respect to the angular coordinate. From the expression of sub-aperture images in Eq. (10), the second derivative at a coordinate  $u_0$  is given by:

$$\left. \frac{\partial^2 \hat{L}_u(\omega_x)}{\partial u^2} \right|_{u=u_0} = -4\pi^2 \sum_k w_x^2 d_k^2 e^{+2i\pi u_0 d_k \omega_x} \hat{L}^k(\omega_x). \quad (25)$$

Let us now consider a set  $\mathcal{R}$  of angular coordinates to regularize. From Eq. (25), the  $2^{nd}$  order view regularization then consists in constructing a matrix  $\Gamma$  where each row  $l$  is associated to an angular coordinate  $u_l \in \mathcal{R}$  and is defined by  $\Gamma_{l,k} = w_x^2 d_k^2 e^{+2i\pi u_l d_k \omega_x}$ . For simplicity, we ignore the constant factor  $-4\pi^2$  in the definition of  $\Gamma$  since it can be accounted for in the regularization parameter  $\lambda$ . Intuitively, in order to apply the regularization at every coordinate in the camera plane, one would need to define a matrix  $\Gamma$  with an infinite number of rows (i.e. infinite set  $\mathcal{R}$ ). Although this might seem impractical, we show in what follows that it can be done by observing that the solution to the regularized problem in Eq. (19) only requires the knowledge of  $\Gamma^* \Gamma$  of finite size  $n \times n$ .

First, let us take a continuous interval  $\mathcal{R} = [-r/2, r/2]$  of size  $r$ , instead of a discrete set. Then, each column  $\Gamma_k$  must be interpreted instead as a function  $\Gamma_k : u \mapsto w_x^2 d_k^2 e^{+2i\pi u d_k \omega_x}$ . Thus, each element of  $\Gamma^* \Gamma$  is given by the following inner product:

$$[\Gamma^* \Gamma]_{k_1, k_2} = \langle \Gamma_{k_1}, \Gamma_{k_2} \rangle = \int_{-r/2}^{r/2} \Gamma_{k_1}(u) \cdot \overline{\Gamma_{k_2}(u)} du, \quad (26)$$

$$= \omega_x^4 d_{k_1}^2 d_{k_2}^2 \int_{-r/2}^{r/2} e^{2i\pi u (d_{k_1} - d_{k_2}) \omega_x} du, \quad (27)$$

$$= \omega_x^4 d_{k_1}^2 d_{k_2}^2 r \text{sinc}(r(d_{k_1} - d_{k_2}) \omega_x). \quad (28)$$

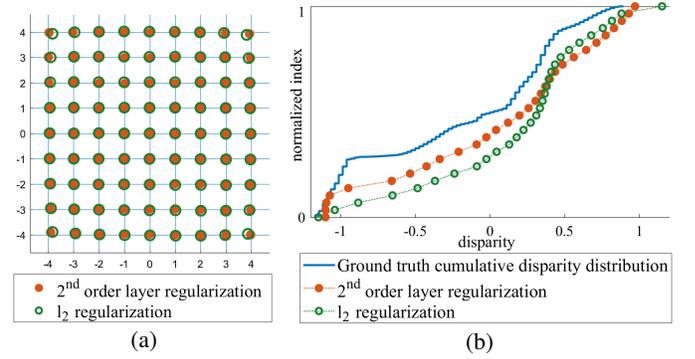


Fig. 3. Example of calibration results for the synthetic Light Field ‘papillon’ [31] with 30 layers: (a) Estimated angular coordinates of the views (the ground truth is the 9x9 regular grid). (b) Sorted disparity values. The vertical axis represents the normalized index of each of the sorted values. The blue curve was obtained by sorting all the pixels of the ground truth disparity map in order to visualize the cumulative disparity distribution.

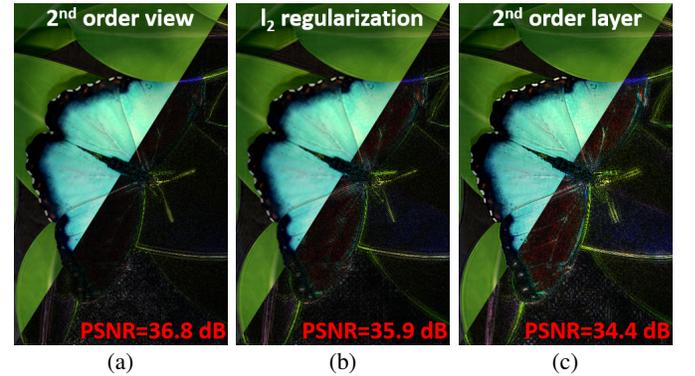


Fig. 4. Central view of the Light Field ‘papillon’ interpolated by constructing a FDL from the 4 corner views with: (a)  $2^{nd}$  order view regularization, (b)  $l_2$  regularization, (c)  $2^{nd}$  order layer regularization. The bottom right part of each image is the error from the ground truth (magnified 5x). The given PSNR is averaged over all the interpolated views. For the comparison, the regularization parameter  $\lambda$  was adjusted in each case to obtain the best PSNR.

For convenience, we replace  $\Gamma^* \Gamma$  in Eq. (19) by the scaled version  $\widetilde{\Gamma^* \Gamma} = \frac{\Gamma^* \Gamma}{r}$ , so that the amount of regularization remains independent of the size of the integration domain in Eq. (26) (i.e. area of the portion of the camera plane covered by the regularization). In our implementation of the  $2^{nd}$  order view regularization, we consider the full camera plane (i.e.  $\mathcal{R} = (-\infty, \infty)$ ) by making  $r$  tend towards the infinity. In this case,  $\widetilde{\Gamma^* \Gamma}$  simply becomes a diagonal matrix such that:

$$[\widetilde{\Gamma^* \Gamma}]_{k,k} = \omega_x^4 d_k^4. \quad (29)$$

Note that for  $\omega_x = 0$ , Eq. (29) gives  $\widetilde{\Gamma^* \Gamma} = 0$ , which does not produce any regularization. Therefore, in practice, a small value  $\epsilon$  is added to the diagonal elements of  $\widetilde{\Gamma^* \Gamma}$ . This is equivalent to adding a  $l_2$  regularization term  $\epsilon \cdot \|\mathbf{x}\|_2^2$  to the problem (18).

The advantage of the proposed  $2^{nd}$  order view regularization over the simple  $l_2$  regularization is demonstrated in Fig. 4 for the view interpolation problem.

2) *2<sup>nd</sup> Order Layer Regularization*: For the calibration step, however, it is preferable to use a regularization that does not depend on the parameters  $d_k$ . Otherwise, the expression of the gradients  $\nabla \mathbf{d}$  would need to take the regularization term into account, which can result in more complex gradient computations and a slower convergence of the algorithm.

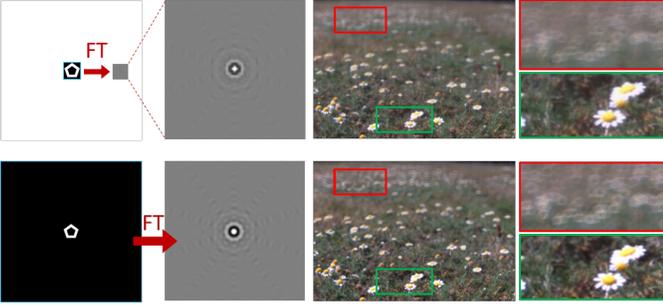


Fig. 5. Top row: no padding. Bottom row: with zero-padding. From left to right: Aperture image; Real part of the Fourier transform of the aperture image (up-sampled with linear interpolation for the top row for the comparison); Rendered result with magnified details. Zero-padding removes the aliasing in the Fourier transform of the aperture, which better preserves contrasts in the final rendered image (best viewed zoomed in).

Instead, we encourage smooth variations between successive layers by defining  $\Gamma$  as a discrete approximation of the second-order differential operator as follows:

$$\Gamma = \begin{bmatrix} -2 & 1 & & & & & \\ 1 & -2 & 1 & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & \\ & & & & & 1 & -2 \end{bmatrix}. \quad (30)$$

The  $2^{nd}$  order layer regularization thus penalizes large differences between neighboring layers, which results in a more uniform distribution of the disparity values. As shown in Fig. 3(b), the calibration using the  $l_2$  regularization (i.e.  $\Gamma = \mathbf{I}_n$ ) tends to find too many disparity values close to the dominant disparity (between 0 and 0.5 in the figure) and may underestimate other parts of the light field (e.g. with disparities close to -1). The  $2^{nd}$  order layer regularization attenuates that effect by encouraging a more uniform distribution. Note, however, that for the layer construction, smoothness along the layers is not desirable and produces artifacts along the edges in the images rendered from the FDL model as shown in Fig. 4(c).

## VI. LIGHT FIELD RENDERING

### A. Implementation

Knowing the layers and their disparities, any view with arbitrary aperture and focus can be rendered in the Fourier domain by applying the FDL decomposition equation (16) and by computing the inverse Fourier transform. The interpretation in the spatial domain is that each layer is shifted (i.e. multiplied by  $e^{+2i\pi u_0 d_k \omega_x}$  in the Fourier domain) and filtered (i.e. multiplied by  $\hat{\psi}(\omega_x(s - d_k))$  in the Fourier domain).

However, except for specific aperture shapes (e.g. square or dirac), the Fourier transform  $\hat{\psi}$  of the aperture function has no analytical expression. Hence, in our implementation, an approximation is obtained by drawing the aperture shape as an image, and by taking its discrete Fourier transform. A linear interpolation is used to determine  $\hat{\psi}(\omega_x(s - d_k))$  from the discrete frequency samples. In order to increase the accuracy of this approximation, the aperture image is zero-padded before computing its Fourier transform, which

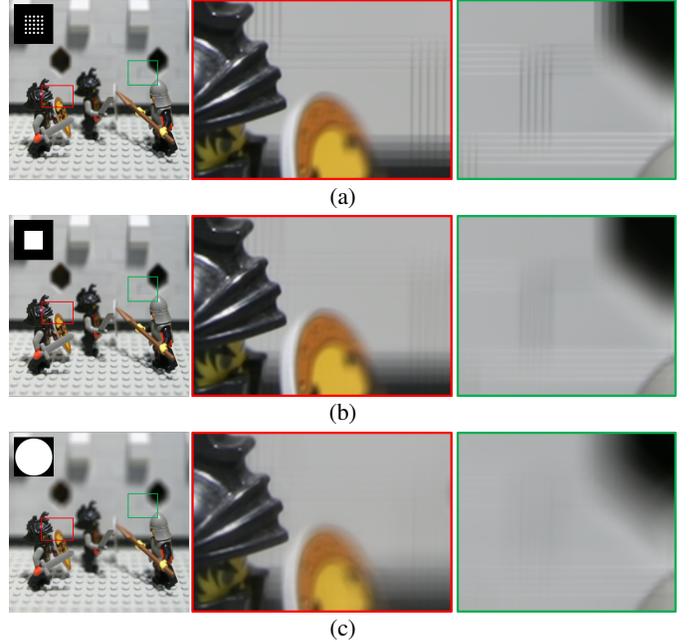


Fig. 6. Refocused image using 5x5 views of the Light Field ‘Lego Knights’ from the Stanford dataset [32]: (a) Shift and Sum method [4], (b) Our method using a square aperture covering the angular coordinates of the input 5x5 views. (c) Our method using a larger circular aperture. Angular aliasing is reduced in our method by considering all the angular coordinates within the aperture.

increases the resolution in the spectral domain. The effect of the padding is illustrated in Fig. 5 showing that the resulting increased spectral domain resolution allows a better preservation of the contrasts in the rendered images. Note also that the spatial resolution of the aperture image can be taken arbitrarily large without affecting the aperture size in the rendered image. In practice, we control the aperture size by replacing  $\hat{\psi}(\omega_x(s - d_k))$  in Eq. (16) by  $\hat{\psi}(f\omega_x(s - d_k))$  with a scaling factor  $f$ . For example, taking  $f = 0$  simulates a sub-aperture image. On the other hand, taking a large factor  $f$  produces a shallow depth of field effect without affecting the complexity of the method.

### B. Visual Results and Discussion

In Fig. 6, we show a comparison of our approach with the conventional Shift and Sum refocusing method [4]. Thanks to the regularization used in the layer construction (see Section V-D), the light field is extended angularly, which allows rendering images with reduced angular aliasing as in Fig. 6(b). As shown in Fig. 6(c), the aperture can also be taken larger than the baseline of the input Light Field from which the FDL model is constructed. The possibility to change the shape of the aperture also gives better control over the bokeh (i.e. the quality of the out-of-focus regions).

Another important and often overlooked factor influencing the bokeh is the color space of the input image data. The refocusing Eqs. (1) and (3) assume that the Light Field rays  $L(x, y, u, v)$  are proportional to luminance data. However, images are traditionally represented in a non-linear color space (e.g. gamma corrected) in order to account for the non-linearity of the human visual system as well as typical displays.

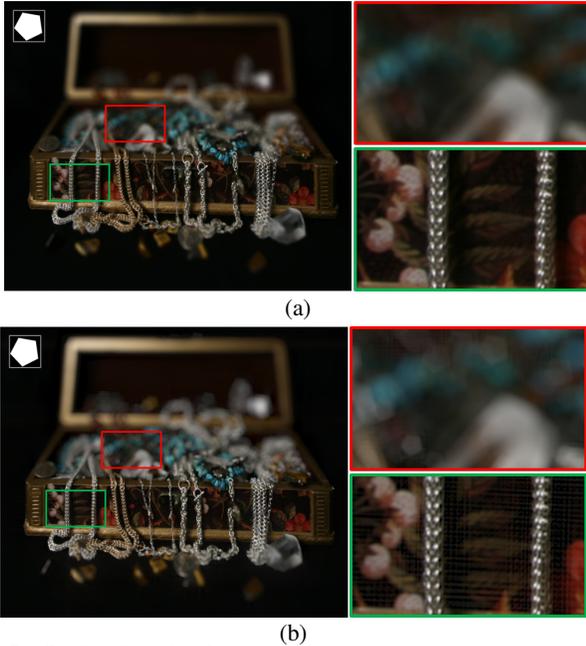


Fig. 7. Rendered result when the processing is performed in (a) gamma corrected space, (b) linear space. Linear space processing results in brighter out-of-focus regions where the bokeh shape (i.e. aperture shape shown in the top left corner) appears more clearly. However, it also increases the artifacts (see details in the green rectangles).

This non-linearity affects the appearance of the refocused images. More realistic bokeh can be achieved by applying inverse gamma correction to the input images before the layer construction, and by applying gamma correction back after the rendering step, as shown in Fig. 7. One limitation of this approach, however, is that computing the layers in a linear color space also produces more visible artifacts. This is due to the fact that solving the layer construction problem (18) in a linear color space does not take the non-linearity of human perception of luminance into account.

A demonstration of our rendering application is available in the supplementary materials. It shows that our approach can be used for controlling simultaneously the viewpoint, the aperture shape and size, and the focus in real-time.

## VII. COMPLEXITY

In this section, we analyse the complexity of the different processing steps proposed in our approach.

First, regarding the spatial resolution, our implementation takes advantage of the symmetry property of the Fourier transform for real signals. Given a real valued function  $g$ , its Fourier transform  $\hat{g}$  is such that  $\hat{g}(-\omega) = \hat{g}(\omega)$ . Hence, for the layer construction and the rendering algorithms, only half of the discrete spectrum must be computed. The remaining frequencies are directly obtained by copying and taking the complex conjugate of the first half. This symmetry property is also used in the calibration step by selecting random frequencies only in one half of the spectrum for the stochastic gradient descent. Regarding memory requirements, using only half of the frequencies in the Fourier representation compensates the fact that complex numbers require twice as much memory as real numbers.

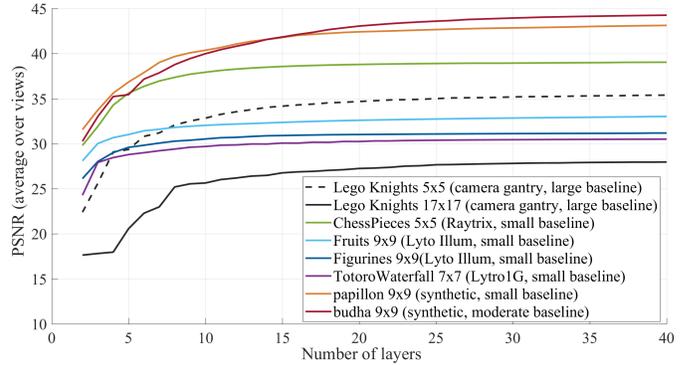


Fig. 8. Reconstruction PSNR (average over views) for different Light Fields with respect to the number of layers used in the FDL model. For the Light Field ‘Lego Knights’, two different results are shown using either the full 17x17 views or only the 5x5 central views.

In addition to the spatial resolution, another important factor to consider for the complexity is the number of layers used. In order to analyse the number of layers needed for our model, we have performed an experiment where Light Field sub-aperture images are used for FDL calibration and construction with varying numbers of layers. The input images are then reconstructed via FDL rendering. Several Light Fields from various datasets were tested including natural Light Fields captured with a first generation Lytro camera (‘Totoro Waterfall’ [33]), a Lytro Illum camera (‘Fruits’ and ‘Toys’ [34]), a Raytrix R8 camera (first frame of the Light Field ‘ChessPieces’ [35]), a traditional camera moving on a gantry (‘Lego Knights’ [32]), and synthetic Light Fields (‘papillon’ and ‘buddha’ [31]). Fig. 8 presents the average Peak Signal to Noise Ratio (PSNR) of all the reconstructed sub-aperture images with respect to the number of layers. The saturation of the PSNR curves show that 15 to 30 layers are generally sufficient, and using more layers does not significantly change the results. Light Fields with a large baseline, and thus large differences between extreme viewpoints (e.g. Lego 5x5, Lego 17x17), typically require more layers than smaller baseline Light Fields. However a very large range of viewpoints also implies large occlusions, which results in a significant loss when reconstructing the sub-aperture images, as observed with Lego 5x5 and Lego 17x17. This makes the analysis difficult for very large baseline Light Fields which may still benefit from more than 30 layers in the ideal non-occluded case. Note that for the small baseline Light Fields captured with Lytro cameras, the relatively low PSNR is due to inaccuracies in the input data (e.g. noise, color differences between views) which are reduced in our reconstruction. More details and results on the reduction of noise and color inconsistencies are presented in section VIII-B.

The computing times for the different processing steps are presented in Table II for all the Light Fields in Fig. 8. The table also provides the number of iterations needed for the calibration since it has a large impact on the computing time. For this experiment, 30 layers were used although less layers would be sufficient for some of the Light Fields (e.g. see ‘ChessPieces’, ‘Toys’, ‘Totoro Waterfall’ in Fig. 8). The processing times were measured using an Intel Core i7-

TABLE II

COMPUTING TIMES FOR OUR MATLAB IMPLEMENTATION USING AN INTEL CORE I7-7700 CPU AND AN NVIDIA GEFORCE GTX 1080 GPU. FOR EVERY LIGHT FIELD 30 LAYERS WERE USED.

Input Light Field (resolution)	Calibration time (#iter)	Layer construction	Render
<b>Lego Knights</b> (1024x1024x5x5)	9.2 s (184)	2.2 s	35 ms
<b>Lego Knights</b> (1024x1024x17x17)	56 s (403)	7.2 s	35 ms
<b>Chess Pieces</b> (1920x1080x5x5)	5.8 s (90)	4.3 s	53 ms
<b>Fruits</b> (625x434x9x9)	5.1 s (70)	0.9 s	12 ms
<b>Toys</b> (625x434x9x9)	6.2 s (91)	0.9 s	11 ms
<b>Totoro WaterFall</b> (379x379x7x7)	3.8 s (72)	0.4 s	7 ms
<b>papillon</b> (768x768x9x9)	7.8 s (109)	1.9 s	20 ms
<b>buddha</b> (768x768x9x9)	8.1 s (113)	1.9 s	19 ms

7700 CPU and a Nvidia GeForce GTX 1080 GPU. All the proposed steps were implemented in Matlab using the parallel computing toolbox in order to process the different frequencies in parallel on the GPU.

Although the processing time for the calibration and layer construction steps is affected by the number of input views (i.e. angular resolution), the rendering time only depends on the spatial resolution and the number of layers. For example, the same rendering time is obtained for ‘Lego Knights’ using either 5x5 or 17x17 input views. This is particularly advantageous when rendering images with a large aperture size that are traditionally obtained by averaging a large number of sub-aperture images with the Shift and Sum algorithm [4]. Note that similarly to our method, the Fourier Slice refocusing algorithm [5] first transforms the Light Field in an intermediate representation in order to perform fast refocusing. In that method the refocusing complexity does not depend on the number of input views. However, since the intermediate representation is the 4D Fourier transform of the Light Field, the memory requirement remains proportional to the number of views. Furthermore, it does not allow changing the aperture or the viewpoint.

## VIII. DIRECT APPLICATIONS

### A. View Interpolation and Extrapolation

As shown in Fig. 6, our approach reduces the angular aliasing and allows extending the aperture size for an input Light Field consisting of a sparse set of sub-aperture images. Equivalently, the layer construction can be seen as a view interpolation and extrapolation method since sub-aperture images with arbitrary angular coordinates can be rendered from the FDL model.

1) *Comparison and Combination with Sparsity Based Approach*: For evaluating the view interpolation capability of the method, we have compared our results with the recent method of Vagharshakyan et al. [20] that enforces the sparsity of the

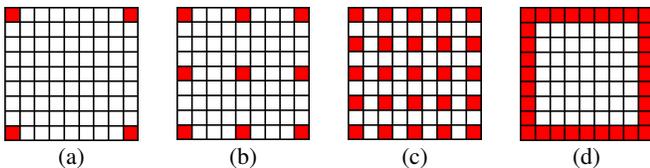


Fig. 9. Tested sampling configurations of the input views (shown in red) for the view interpolation: (a) ‘2x2’, (b) ‘3x3’, (c) ‘5x5’, (d) ‘border’.

TABLE III

VIEW INTERPOLATION RESULTS FOR THE LIGHT FIELD ‘PAPILLON’. THE RESULTS GIVEN ARE THE AVERAGE PSNR OF THE INTERPOLATED VIEWS AND THE TOTAL RECONSTRUCTION TIME (INCLUDING CALIBRATION FOR THE FDL METHOD).

Method \ Input views	2x2	3x3	5x5	border
<b>Shearlet</b> [20]	34.8 dB 143 mn	40.8 dB 157 mn	41.8 dB 185 mn	37.2 dB 105 mn
<b>FDL</b>	36.8 dB <b>5.4 s</b>	40.9 dB <b>7.8 s</b>	<b>42.9 dB</b> <b>8.9 s</b>	<b>43.1 dB</b> <b>8.8 s</b>
<b>Shearlet(full)+FDL</b>	35.5 dB 143 mn	41.4 dB 157 mn	42.6 dB 185 mn	40.1 dB 105 mn
<b>Shearlet(border)+FDL</b>	<b>38.4 dB</b> 53 mn	<b>42.0 dB</b> 53 mn	42.7 dB 53 mn	x

epipolar images of the Light Field in the shearlet domain. For the comparison, the synthetic Light Field ‘papillon’ was used with several sampling configurations of the input views as illustrated in Fig. 9. The average PSNR of the interpolated views and the computation times are shown in Table III. The table also includes the results for two schemes combining the FDL and the Shearlet approaches. In the ‘Shearlet(full)+FDL’, the full Light Field is reconstructed with [20]. Then, a FDL model is computed from the input views and the previously reconstructed intermediate views. The final reconstruction is obtained with FDL rendering. The ‘Shearlet(border)+FDL’ scheme is similar, but only the views at the periphery of the Light Field are interpolated using [20]. Our experiments with the Shearlet method were performed using the author’s implementation of the epipolar image reconstruction with 100 iterations. The reconstruction of the full Light Field was performed by scanning the epipolar images in the ‘direct’ order as detailed in [20].

The results in Table III show that our FDL view interpolation performs particularly well in the ‘border’ configuration. The Shearlet method, on the other hand, does not fully take advantage of this configuration as it processes each epipolar image independently using only the input views located on the corresponding row or column in the Light Field. In the other configurations, a higher PSNR is also observed with the FDL method. However, when the input Light Field is too sparse (e.g. 2x2), it produces ringing artifacts as shown in Fig. 10. In this situation, the Shearlet method better preserves the edges than our FDL method, but it tends to blur the fine details and introduce color distortions, which explains the lower PSNR. The best strategy for such sparse inputs is then to combine the two methods with the ‘Shearlet(border)+FDL’ scheme which does not produce ringing artifacts. In comparison with the Shearlet method alone, this scheme better keeps the details and color consistency as shown in Fig. 10. It is also significantly faster since the reconstruction of the interior views is performed using our FDL approach which requires a negligible computing time compared to the Shearlet method.

2) *Comparison with Deep Learning Approaches*: We have additionally performed comparisons with three state-of-the-art deep learning approaches [12], [14], [16]. The method of Kalantari et al. [12] uses two Convolutional Neural Networks (CNN) where the first one determines a disparity map, while the second one refines the disparity compensated views. Yeung et al. [16] have proposed a CNN defining an end-to-end mapping between sparse input views and the dense Light

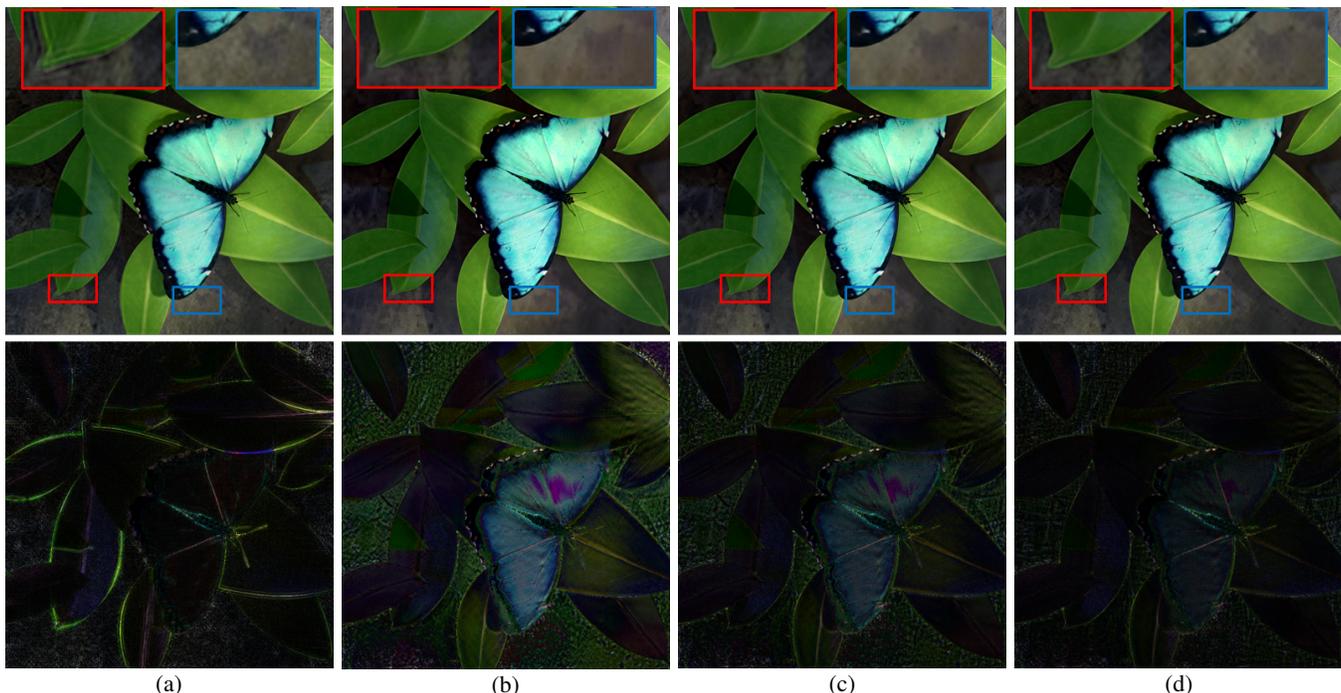


Fig. 10. Interpolation of the central view from the 4 corner views of the Light Field ‘papillon’ (i.e. 2x2 configuration) using: (a) only FDL, (b) only Shearlet [20], (c) ‘Shearlet(full)+FDL’, (d) ‘Shearlet(border)+FDL’. The bottom images show the residual computed with the ground truth central view (magnified 5x). A video of the results is available in the supplementary materials.

Field. Finally, the method of Wu et al. performs super-resolution of the epipolar images. For the comparison, the natural Light Fields ‘Fruits’ and ‘Toys’ [34] as well as ‘Friends 1’ and ‘Vespa’ [36] were used with gamma correction. They were captured with a Lytro Illum camera, similarly to the Light Fields used for the training in the tested deep learning approaches. The PSNR results and average runtimes are presented in Table IV for the configurations with 2x2 input and 8x8 output views (2x2→8x8) and with 3x3 input and 7x7 output views (3x3→7x7). Note that for the method of Yeung et al. [16], the available CNN model only applies to the 2x2→8x8 configuration. Although the method of Kalantari et al. [12] was designed for 2x2 input views, we have used it in the 3x3→7x7 configuration by treating the quadrants of the Light Field separately. Furthermore, in [12] and [16], the output Light Field is cropped in the spatial dimensions. For the comparison, the same crop (12 pixels on each side) was applied to compute all the PSNR values in Table IV.

The CNNs used in [12] and [16] are particularly efficient for the 2x2→8x8 configuration for which they were designed and trained. In this case, they obtain better results than our FDL approach which is penalized by ringing artifacts, similarly to the results shown in Fig. 10 (a). However, the performances of Kalantari et al’s method drop in the 3x3→7x7 configuration, indicating that the model learned for 2x2 input views does not generalize well to more dense inputs. Our FDL model, on the other hand, provides significantly better interpolation in this case thanks to the higher density of input views, leading to a better conditioning of the FDL construction problem. While Yeung et al. have also reported in [16] high quality results for the 3x3→7x7 scenario, an adapted version of their model must be designed and trained for each configuration.

TABLE IV  
VIEW INTERPOLATION RESULTS FOR NATURAL LIGHT FIELDS. THE RESULTS GIVEN ARE THE AVERAGE PSNR OF THE INTERPOLATED VIEWS.

		Fruits	Toys	Friends 1	Vespa	average runtime
Yeung et al. [16]	2x2→8x8	33.1	32.9	37.5	34.4	40s
Kalantari et al. [12]	2x2→8x8	34.2	33.2	37.1	34.4	10mn30s
	3x3→7x7	31.7	32.1	36.4	34.0	7mn
Wu et al. [14]	2x2→8x8	28.3	32.0	36.6	32.4	19mn
	3x3→7x7	35.3	35.7	39.4	37.7	13mn30s
FDL	2x2→8x8	32.4	32.5	36.8	32.1	6s
	3x3→7x7	35.9	35.3	40.0	36.9	6s

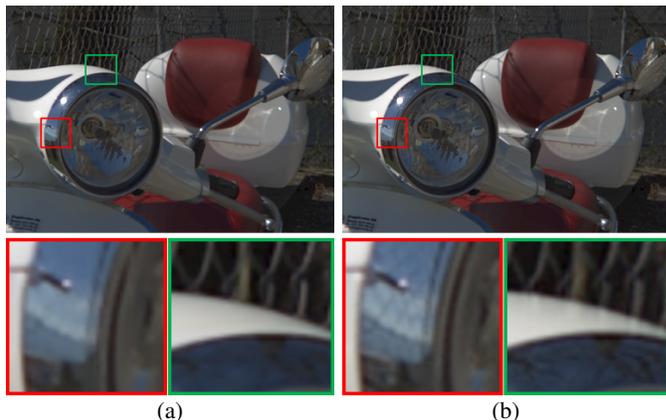


Fig. 11. Interpolated view (5,5) of the Light Field ‘Vespa’ from the configuration 3x3→7x7: (a) Wu et al. (b) FDL. Our FDL method produces more detailed results (see red boxes). However, textures of occluded regions become apparent on the occluding objects (see green boxes).

Our PSNR results are more comparable to those of Wu et al. [14]. Visual comparisons in Fig. 11 additionally reveals that our method better preserves the details. However, highly textured occluded areas become apparent on the fore-

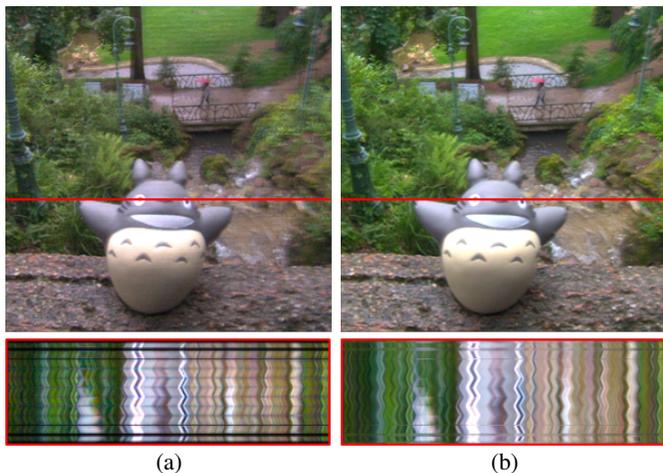


Fig. 12. Simultaneous denoising and color correction of the Light Field ‘TotoroWaterfall’ (captured with a Lytro camera): (a) Original (b) Processed with the FDL method. The top image corresponds to a view on the side of the Light Field. Epipolar images corresponding to the horizontal red line are shown on the bottom to visualize the color consistency between views.

ground, making the borders of occluding objects appear semi-transparent. Note also that for the tested Light Fields and the two configurations  $2 \times 2 \rightarrow 8 \times 8$  and  $3 \times 3 \rightarrow 7 \times 7$ , the average computing time of [14] was about 16 minutes against only 6 seconds for our method (including the calibration step).

### B. Denoising and Other Filtering Effects

Another direct application of our FDL construction algorithm is the Light Field denoising problem. Denoising is naturally obtained by constructing a FDL model from a Light Field, and by rendering images using the same angular coordinates as the input sub-aperture images. The noise is filtered in this approach thanks to the model definition that enforces the Light Field dimensionality gap prior [21] (see Section IV-A). Further noise reduction is also allowed by the proposed regularization that better ensures consistency between views. Note that for the same reasons, the method also ensures color and illumination consistency along the angular dimensions. Hence it can serve simultaneously as a denoising and color correction method for Light Fields with variations of color and illumination between the sub-aperture images. In practice, this is particularly useful for Light Fields captured with plenoptic cameras (e.g. Lytro), as shown in figure 12.

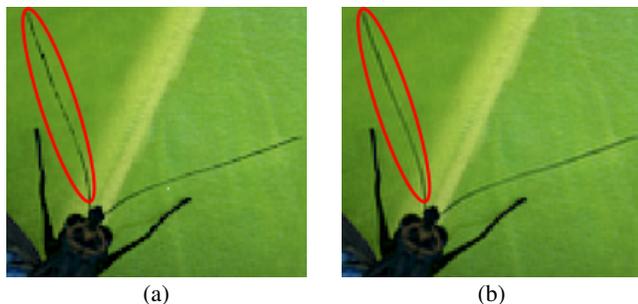


Fig. 13. Spatial anti-aliasing effect of the FDL model: (a) Original central view (detail of the Light Field ‘papillon’), (b) Reconstructed view from the FDL model. The effect is particularly visible on the butterfly’s antennae.

TABLE V  
DENOISING RESULTS FOR THE LIGHT FIELD ‘TAROT’ USING  $11 \times 11$  INPUT VIEWS AND NOISE WITH STANDARD DEVIATION  $\sigma = 10$  AND  $\sigma = 50$ .

	$\sigma = 10$		$\sigma = 50$		average runtime
	PSNR	SSIM	PSNR	SSIM	
noisy input	28.1	0.787	14.2	0.283	
LFBM5D [38]	39.3	0.983	32.7	0.937	~5h30mn
HF4D [29]	31.2	0.888	22.3	0.603	~40s
FDL (original)	32.0	0.945	26.6	0.785	~1mn30s
FDL (relaxed)	33.4	0.946	26.1	0.760	~2mn15s

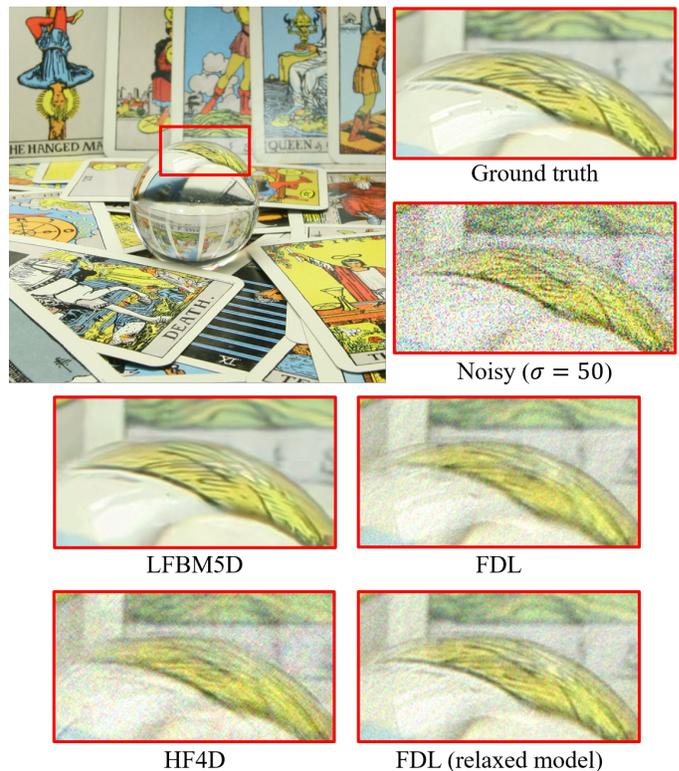


Fig. 14. Denoising results for the Light Field ‘Tarot’. A view on the corner of the Light Field is shown. For the test, the Light Field was corrupted by gaussian noise with standard deviation  $\sigma = 50$ .

Similarly, by ensuring consistency between multiple views, the FDL model and its regularization act as a spatial anti-aliasing filter in the case of spatially aliased Light Fields (see Fig. 13). This situation typically arises with Light Fields captured with lenslet-based plenoptic cameras because of the low lenslet resolution. It has been shown in [37] that such Light Fields are spatially non-bandlimited and that despite the low spatial sampling, higher frequencies than the Nyquist frequency can be recovered in theory. However, we construct the layers at the same spatial resolution as the input views. Therefore, while an anti-aliasing effect is produced, these higher frequencies are not recovered in our method. Constructing higher resolution layers to fully exploit the 4D Light Field sampling would allow Light Field super-resolution and is left for future work.

For Light Fields with a wide baseline, our processing may also introduce distortions due to the assumptions of non-occluded Lambertian Light Field. In this case, better results

can be obtained using the relaxed version of the FDL model presented in Section V-C. Denoising results are presented in Fig. 14 for the Light Field ‘Tarot’ that exhibits strong non-Lambertian effects. The reflections in the crystal ball are better preserved by the relaxed FDL model than the original one. For the comparison, the results obtained with the LFBM5D [38] and the Hyperfan 4D filter (HF4D) [29] are also shown. The corresponding PSNR and SSIM results are given in Table V.

The best results were obtained with the LFBM5D method that completely removes the noise while preserving most of the details in the image. However, this method performs heavy processing and typically requires several hours for a Light Field. The HF4D method is faster since it simply consists in multiplying the 4D Fourier Transform of the Light Field by a 4D filter. This hyperfan filter is designed to attenuate the frequencies outside of the theoretical region of support of the Light Field in the 4D Fourier domain, under the assumption of a non-occluded Lambertian scene. With comparable processing times, our FDL approach removes more noise thanks to the linear optimization used instead of the direct filtering in the HF4D method. Furthermore, our model can be more easily generalized for the case of scenes with occlusions or non-Lambertian surfaces, as shown with the proposed relaxation.

Note however that for a high level of noise (i.e.  $\sigma = 50$ ), the relaxation of the FDL model results in lower objective scores despite the better preservation of the non-Lambertian surface observed in Fig. 14. The reason is that the relaxed model also preserves more noise from the input data. Hence the choice of the original or the relaxed version of the model is a tradeoff between the noise level and the amount of occlusions or non-Lambertian effects.

## IX. CONCLUSION AND PERSPECTIVES

We have presented a new representation for Light Fields called Fourier Disparity Layers. The light information in the scene is decomposed into several layers, each corresponding to a depth plane parameterized by a disparity value. The decomposition step is formulated in the Fourier domain using a simple linear least square problem per-frequency, hence allowing fast processing with GPU parallelization. We have demonstrated the advantages of the FDL representation for several Light Field processing tasks. Those include real time rendering, view interpolation, denoising, as well as a calibration step that determines the angular coordinates of the input sub-aperture images along with an optimal set of disparity values for each layer.

The computational efficiency of our layer construction method coupled with its flexibility regarding the type of input images opens perspectives for an even larger range of applications. For example, our supplementary video presents results of FDL model construction from a focal stack, where the camera aperture and the focus parameter of each image are known. More generally, the method could even take advantage of a hybrid capture system taking images with different apertures, focusing depth, and varying positions on the camera plane. Furthermore, since our approach does not require a specific pattern of the input view positions, it can greatly simplify the

capture of Light Fields by removing the need for the different viewpoints to be regularly sampled on a grid.

However, despite this flexibility, we have shown that the quality of the FDL model produced with our method depends on the input configuration. For example, while very accurate view interpolation is obtained from the sub-aperture images located on the borders of the Light Field, a too sparse viewpoint sampling may lead to ringing artifacts. Furthermore, for Light Fields with a large baseline, the FDL model is more likely to produce visible errors due to large occlusion areas or non-Lambertian effects. In the paper, we have partially addressed these limitations. In the case of a too sparse viewpoint sampling, a combination of our view interpolation with a state of the art approach was shown to outperform either of the two methods taken individually. A better handling of occlusions and non-Lambertian effects was also obtained, for the denoising application, using a relaxed version of the model.

In the aim of further extending the applicability of the FDL approach, future work may focus on generalizing the relaxed model to view interpolation, or including an additional prior directly in the layer construction (e.g. sparsity in the Shearlet domain similarly to [20]) to better cope with very sparse or very noisy Light Fields. A generalization of the calibration to wide aperture images would also be a valuable tool to facilitate the creation of Light Fields from less conventional input data such as focal stacks.

## APPENDIX A PROOF OF SPARSITY PRIOR

The spatial regions  $\Omega_k$  are defined for the central view at  $u = 0$ . The corresponding regions  $\Omega_k^u$  can also be defined for any view  $u$  by  $\Omega_k^u = \{x \in \mathbb{R} \mid x + ud_k \in \Omega_k\}$ .

The expression in Eq. (4) is then formulated as:

$$\forall k \in \llbracket 1, n \rrbracket, \forall (x, u) \in \Omega_k^u \times \mathbb{R}, L(x, u) = L(x + ud_k, 0). \quad (31)$$

In the assumption of a non-occluded Light Field, the regions  $\Omega_k^u$  are such that  $\forall u \in \mathbb{R}, x \in \mathbb{R} \setminus \bigcup_k \Omega_k^u, L(x, u) = 0$ , and for any fixed  $u \in \mathbb{R}$ , the sets  $\Omega_k^u$  are pairwise disjoint. Hence, the Fourier Transform of the Light Field is given by:

$$\begin{aligned} \hat{L}(\omega_x, \omega_u) &= \iint_{\mathbb{R}^2} e^{-2i\pi(x\omega_x + u\omega_u)} L(x, u) dx du \\ &= \int_{-\infty}^{+\infty} e^{-2i\pi u\omega_u} \sum_k \left[ \int_{\Omega_k^u} e^{-2i\pi x\omega_x} L(x, u) dx \right] du. \end{aligned} \quad (32)$$

Using Eq. (31), and by change of variable we obtain:

$$\begin{aligned} \hat{L}(\omega_x, \omega_u) &= \int_{-\infty}^{+\infty} e^{-2i\pi u\omega_u} \sum_k \left[ \int_{\Omega_k^u} e^{-2i\pi x\omega_x} L(x + ud_k, 0) dx \right] du \\ &= \int_{-\infty}^{+\infty} e^{-2i\pi u\omega_u} \sum_k \left[ \int_{\Omega_k} e^{-2i\pi(x-ud_k)\omega_x} L(x, 0) dx \right] du. \end{aligned} \quad (33)$$

By re-arranging the terms the result is:

$$\begin{aligned} & \hat{L}(\omega_x, \omega_u) \\ &= \sum_k \left[ \int_{-\infty}^{+\infty} e^{-2i\pi u(\omega_u - d_k \omega_x)} du \int_{\Omega_k} e^{-2i\pi x \omega_x} L(x, 0) dx \right] \\ &= \sum_k \left[ \delta(\omega_u - d_k \omega_x) \hat{L}^k(\omega_x) \right]. \end{aligned} \quad (34)$$

where we define  $\hat{L}^k(\omega_x) = \int_{\Omega_k} e^{-2i\pi x \omega_x} L(x, 0) dx$ .  $\square$

## REFERENCES

- [1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. ACM, 1996, pp. 31–42.
- [2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96, 1996, pp. 43–54.
- [3] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 2000, pp. 307–318.
- [4] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-Held Plenoptic Camera," CSTR 2005-02, Tech. Rep., Apr. 2005.
- [5] R. Ng, "Fourier slice photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744, Jul. 2005.
- [6] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and Chintan, "Spatio-angular resolution tradeoffs in integral photography," in *Proc. Eurograph. Symp. Rendering*, 2006, pp. 263–272.
- [7] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions of Pattern analysis and machine intelligence*, vol. 36, no. 3, 2013.
- [8] S. Pujades, F. Devernay, and B. Goldluecke, "Bayesian view synthesis and image-based rendering principles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 3906–3913.
- [9] O. S.-H. G. Chaurasia, S. Duchêne and G. Drettakis, "Depth synthesis and local warps for plausible image-based navigation," *ACM Trans. Graph.*, vol. 32, 2013.
- [10] Z. Zhang, Y. Liu, and Q. Dai, "Light field from micro-baseline image pair," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3800–3809.
- [11] E. Penner and L. Zhang, "Soft 3d reconstruction for view synthesis," *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, vol. 36, no. 6, 2017.
- [12] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. on Graphics (Proceedings of SIGGRAPH Asia 2016)*, vol. 35, no. 6, 2016.
- [13] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "Deepstereo: Learning to predict new views from the worlds imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5515–5524.
- [14] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on epi," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.
- [15] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 57–65.
- [16] H. W. F. Yeung, J. Hou, J. Chen, Y. Y. Chung, and X. Chen, "Fast light field reconstruction with deep coarse-to-fine modeling of spatial-angular clues," in *ECCV*, 2018.
- [17] L. Shi, H. Hassaneh, A. Davis, D. Katabi, and F. Durand, "Light field reconstruction using sparsity in the continuous fourier domain," *ACM Trans. on Graphics*, vol. 34, no. 1, pp. 1–13, 2014.
- [18] K. Takahashi, S. Fujita, and T. Fujii, "Good group sparsity prior for light field interpolation," in *IEEE Conf. on Image Processing (ICIP)*, Sep. 2017, pp. 1447–1451.
- [19] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Accelerated shearlet-domain light field reconstruction," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1082–1091, Oct. 2017.
- [20] —, "Light field reconstruction using shearlet transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 133–147, Jan. 2018.
- [21] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 1831–1838.
- [22] K. Kodama and A. Kubota, "Efficient reconstruction of all-in-focus images through shifted pinholes from multi-focus images for dense light field synthesis and rendering," *IEEE Trans. on Image Processing*, vol. 22, no. 11, pp. 4407–4421, 2014.
- [23] A. Mousnier, E. Vural, and C. Guillemot, "Partial light field tomographic reconstruction from a fixed-camera focal stack," *CoRR*, vol. abs/1503.01903, 2015.
- [24] F. Pérez, A. Pérez, M. Rodriguez, and E. Magdaleno, "Lightfield recovery from its focal stack," *J Math Imaging Vis*, vol. 56, pp. 573–590, 2016.
- [25] J. R. Alonso, A. Fernández, and J. A. Ferrari, "Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack," *Appl. Opt.*, vol. 55, no. 9, pp. 2380–2386, Mar. 2016.
- [26] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays," *ACM Trans. Graph.*, vol. 30, no. 4, 2011.
- [27] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 31, no. 4, pp. 1–11, 2012.
- [28] K. Takahashi, Y. Kobayashi, and T. Fujii, "From focal stack to tensor light-field display," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4571–4584, Sep. 2018.
- [29] D. G. Dansereau, D. L. Bongiorno, O. Pizarro, and S. B. Williams, "Light field image denoising using a linear 4d frequency-hyperfan all-in-focus filter," in *Proc. SPIE*, vol. 8657, Feb. 2013, pp. 8657 – 8657 – 14.
- [30] D. Dansereau and L. T. Bruton, "A 4-d dual-fan filter bank for depth filtering in light fields," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 542–549, Feb. 2007.
- [31] S. Wanner, S. Meister, and B. Goldlücke, "Datasets and benchmarks for densely sampled 4d light fields," in *Vision, Modeling & Visualization*, 2013, pp. 225–226, dataset available at [http://lightfieldgroup.iwr.uni-heidelberg.de/?page\\_id=713](http://lightfieldgroup.iwr.uni-heidelberg.de/?page_id=713), accessed 01-10-2018.
- [32] "The stanford light field archive," <http://lightfield.stanford.edu/lfs.html>, accessed: 01-10-2018.
- [33] X. Jiang, M. L. Pendu, R. A. Farrugia, and C. Guillemot, "Light field compression with homography-based low-rank approximation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1132–1145, Oct. 2017, dataset available at <https://www.irisa.fr/temics/demos/lightField/CLIM/DataSoftware.html>, accessed 01-10-2018.
- [34] M. Le Pendu, X. Jiang, and C. Guillemot, "Light field inpainting propagation via low rank matrix completion," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1981–1993, Apr. 2018, dataset available at <https://www.irisa.fr/temics/demos/lightField/CLIM/DataSoftware.html>, accessed 01-10-2018.
- [35] "Light field video dataset captured by a r8 raytrix camera," <https://www.irisa.fr/temics/demos/lightField/CLIM/Datasets/RaytrixR8Dataset-5x5/index.html>, accessed: 01-10-2018.
- [36] M. Rerabek and T. Ebrahimi, "New light field image dataset," 2016, dataset available at <https://mmspg.epfl.ch/downloads/epfl-light-field-image-dataset>, accessed 05-04-2019.
- [37] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 16:1–16:19, Mar. 2015.
- [38] M. Alain and A. Smolic, "Light field denoising by sparse 5d transform domain collaborative filtering," in *IEEE International Workshop on Multimedia Signal Processing (MMSP 2017)*, Oct. 2017.