High Resolution Light Field Recovery with Fourier Disparity Layer Completion, Demosaicing, and Super-Resolution

Mikael Le Pendu, and Aljosa Smolic

Abstract—In this paper, we present a novel approach for recovering high resolution light fields from input data with many types of degradation and challenges typically found in lenslet based plenoptic cameras. Those include the low spatial resolution, but also the irregular spatio-angular sampling and color sampling, the depth-dependent blur, and even axial chromatic aberrations. Our approach, based on the recent Fourier Disparity Layer representation of the light field, allows the construction of high resolution layers directly from the low resolution input views. High resolution light field views are then simply reconstructed by shifting and summing the layers. We show that when the spatial sampling is regular, the layer construction can be decomposed into linear optimization problems formulated in the Fourier domain for small groups of frequency components. We additionally propose a new preconditioning approach ensuring spatial consistency, and a color regularization term to simultaneously perform color demosaicing. For the general case of light field completion from an irregular sampling, we define a simple iterative version of the algorithm. Both approaches are then combined for an efficient super-resolution of the irregularly sampled data of plenoptic cameras. Finally, the Fourier Disparity Layer model naturally extends to take into account a depth-dependent blur and axial chromatic aberrations without requiring an estimation of depth or disparity maps.

Index Terms—Light Fields, Fourier Disparity Layers, Super-Resolution, Demosaicing, Completion, Demultiplexing, Plenoptic Camera.

1 Introduction

Light fields enhance the traditional 2D images with two additional angular dimensions representing a variation of the viewpoint along a plane. The recent advances in computational photography have made it possible to capture dense light fields, allowing for a continuous change of viewpoint as well as post-capture refocusing. However, capturing a dense and high quality light field from a real scene remains a challenging task. A common approach consists in capturing first a sparse light field either with a rig of cameras or with a single moving camera. View synthesis algorithms can then be used to simulate a dense angular sampling from the sparse set of captured views.

Lenslet based plenoptic cameras offer a promising alternative by capturing a dense light field in a single shot with a portable device. A microlens array placed in front of the sensor splits the light rays coming from different angles, hence forming many images of the scene from different viewpoints varying within the camera aperture. In the typical setting where the sensor is placed at the microlenses' focal distance [1], the angular sampling density is only limited by the sensor resolution. However, the spatial resolution is much lower than in conventional cameras, since each view captures only one pixel per microlens. While a higher spatial sampling can be achieved in the focused plenoptic camera setting [2] this comes at the expense of the angular resolution. Furthermore, the extraction of high quality views form the plenoptic RAW data is particularly challenging. For example, conventional 2D demosaicing is insufficient to retrieve accurate colors from the bayer color pattern applied on the sensor. Methods specifically taking the light field structure into account are required [3], [4], [5]. Additionally, due to manufacturing difficulties, the size of each microlens may not be a multiple of the pixel size, and a perfect alignment of the microlens array with the sensor is unlikely, which results in an irregular sampling in the spatio-angular 4D space. Some camera designs also use an hexagonal array of microlenses in order to optimize the spatial sampling, which further requires resampling the extracted views from an hexagonal to a square grid of pixels. Finally, limitations of the optics introduce several artefacts such as depth-dependent blur and chromatic aberrations.

In the traditional imaging pipeline for plenoptic cameras, the demosaicing, the angular and the spatial resampling are treated sequentially, with successive interpolation steps [6], [7], [8], [9]. Spatial super-resolution is generally considered as a separate problem and does not take into account the errors accumulated in the early view extraction steps. Furthermore, while light field super-resolution is a very active topic, only few methods consider a depth-dependent blur model.

In this paper, we propose a super-resolution method that can simultaneously address the various types of degradation of real captures in a single optimization framework. Our method is based on the Fourier Disparity Layer model (FDL) [10] in which a set of layers, constructed in the Fourier domain, can be shifted and summed to form any view of the light field. Unlike [10], however, we construct the FDL that optimally reconstructs views with an incomplete spatial sampling. The missing samples are thus naturally completed thanks to the inter-view dependencies captured

• *M. Le Pendu and A. Smolic are with the School of Computer Science, Trinity College Dublin.*

by the model. A novel pre-conditioning approach further ensures spatial consistency. We show that when the missing pixels form a regular pattern (e.g. super-resolution), the problem is expressed as simple systems of linear equations in the Fourier domain, allowing for fast computations. For the general case of completion with an irregular sampling, we also present an iterative version of our algorithm. In both cases, color demosaicing is addressed using a regularization term in the FDL optimization to exploit inter-component correlations. Furthermore, thanks to the flexibility of the FDL model, the depth-dependent blur and axial chromatic aberrations can be taken into account and compensated for, without increasing the computational complexity.

In our experiments, we first evaluate the performance of our super-resolution algorithm with respect to the state-of-the-art in light field super-resolution. Then, in the more challenging scenario of high resolution light field extraction from plenoptic camera RAW data, we apply our full optimization framework combining completion and super-resolution along with demosaicing and correction of depth-dependent blur and chromatic aberrations. Our experiments with Lytro Illum camera data reveal finer details and an increased depth of field compared to existing approaches, including the official Lytro Desktop software.

In summary, the contributions of the paper are:

- Theoretical analysis of the relationship between frequencies in the case of missing pixel data.
- Fast FDL super-resolution algorithm considering small subsets of frequencies. Unlike the FDL construction in [9] that treats frequencies independently, the proposed algorithm can recover the high frequencies from the aliasing in the lower ones.
- Derivation of a pre-conditioning matrix enforcing spatial consistency.
- Iterative version of the algorithm for the more general light field completion problem.
- Definition of a color regularization term for improved color demosaicing.
- Application to high quality view extraction from real lenslet RAW data using the full optimization framework and further leveraging the flexibility of the FDL model to correct depth-dependant blur and axial chromatic aberrations.

2 Related Work

2.1 Light Field super-resolution

Our work is essentially related to the topic of light field super-resolution. Fundamental work in this area has been proposed by Liang and Ramamoorthi [11], studying an advanced image formation model for lenslet based cameras. It accounts for physical parameters such as the angular sensitivity of the photosensors as well as their spatial extent and that of the microlenses. Derivations from the model allow them to characterize the depth-dependent filtering effect of the combined microlenses and sensor. Their analysis confirms the empirical observation that, despite this filtering, spatial aliasing remains in the captured light fields, hence allowing for effective super-resolution exploiting the multiple views.

Although we do not consider the full image formation model of [11], our method can be used to tackle a depth-dependent blur. This aspect is often ignored in super-resolution, with some exceptions however, such as the method of Bishop and Favaro [12], using an image formation model to derive a Bayesian framework.

The models in [11] and [12] do not consider the Bayer array that applies a red, green, or blue filter on each pixel. Therefore, a color demosaicing step must be applied first to retrieve the full color information at each pixel. Demosaicing errors may then be further increased by deconvolution. An alternative approach is used in [3], [4], [13], [14], where the color components are super-resolved separately, therefore avoiding a propagation of demosaicing artifacts. However, unlike our method, they do not exploit the strong correlations between RGB components, or only partially, in the method of Yu et al. [4], for recovering the red and blue components. To our knowledge, none of the light field demosaicing methods simultaneously compensate for chromatic aberrations. Furthermore, most other works in light field super-resolution essentially focus on inverting a given downsampling operator with fixed blur kernel, therefore ignoring depth-dependent blur as well as irregular sampling and color issues.

Another challenging aspect of light field super-resolution is the need for accurate depth estimation, which is difficult to perform from the input aliased low resolution data. Several methods including [12], [15], [16], [17] rely on disparity information obtained separately from the super-resolution algorithm. On the other hand, Alain and Smolic [18] estimate correspondences between views that are refined within the super-resolution algorithm. While providing high quality results, this type of approach is very computationally intensive. Thanks to a graph-based prior capturing the light field structure, the method of Rossi and Frossard [19] only requires a fast and approximate disparity estimation. However, the global graph-based optimization itself has a high computational complexity. By contrast, our method is fast and only requires a single disparity value per layer of the FDL model, reflecting the disparity distribution in the scene. These values are efficiently obtained with the FDL calibration in [10].

Finally, a very recent body of work explores approaches fully based on machine learning [20], [21], [22], [23], [24], [25], [26], [27]. While reaching very high performance by implicitly learning both spatial and angular image priors, the current methods are limited in their applications due to the impossibility to adjust parameters such as the blur kernel, or the number of input views. In these methods, the training dataset consists of original light fields with their associated downsampled version. The model learnt is thus specific to the downsampling used, and the training process must be repeated to address another blur filter.

2.2 Fourier Disparity Layers

Our work relies on the Fourier Disparity Layer representation of the light field [10]. In essence, the FDL model consists of a set of layers, each layer mostly containing the texture information at a given depth in the scene. Each layer is associated to a disparity value representing its depth, so that any view of the light field can be reconstructed as a

TABLE 1
Table of notation

Symbols	Description
L_j	Light field view of index j ($j \in [\![1, m]\!]$).
l_k	Layer of index k in the FDL model ($k \in [\![1, n]\!]$).
μ_j	Mask indicating missing pixels in view L_j.
ϕ, ψ	Respectively spatial and angular blur filters.
u_j	Angular coordinate of input view L_j.
d_k	Disparity value associated with the layer l_k.
S	Super-resolution factor.
N, N_0	Spatial resolution of the images respectively with and without super-resolution. ($N = S \cdot N_0$).
q, r, t	Indices for discrete frequency components ($q = r + tN_0$) $q \in [\![0, N-1]\!]$, $r \in [\![0, N_0 - 1]\!]$, $t \in [\![0, S-1]\!]$.
ω_q	Spatial frequency of index q ($\omega_q = q/N$).
$\hat{f}$	Fourier transform of a function f.
$\overline{z}$	Conjugate of a complex number z.
$\mathbf{A}$, $\mathbf{x}$	A matrix (bold uppercase) and vector (bold lowercase).
$\mathbf{A_q}$, $\mathbf{x_q}$	Respectively a matrix and a vector with index q.
$(\mathbf{A_q})_{j,k}$	Element of a matrix $\mathbf{A_q}$ on the j^{th} row and k^{th} column.
$(\mathbf{x_q})_k$	k^{th} element of a vector $\mathbf{x_q}$.
$\mathbf{A}^\top$	Transpose of a matrix $\mathbf{A}$.
$\mathbf{A}^*$	Conjugate transpose of $\mathbf{A}$ ($\mathbf{A}^* = \overline{\mathbf{A}^\top}$).

sum of the layers shifted according to their disparity value and the view position. The FDL model can be determined in order to optimally reconstruct a set of input light field views. It was shown in [10] that in the Fourier domain, this global optimization consists in simple linear least square problems expressed independently at each spatial frequency component, which is ideal for GPU parallel computations. A calibration method was also derived to determine the disparity value of each layer and the input view coordinates which are required for the layer construction.

Several applications such as real-time refocusing, view interpolation, or denoising naturally arise from this approach. However, it does not address super-resolution, demosaicing, or more generally, completion of light fields since the input views with missing samples cannot be transformed in the Fourier domain. Therefore, a generalization to these applications is not straightforward, and it is the main contribution of our paper.

3 General Problem Formulation

For simplicity, we consider a 2D spatio-angular slice $L(x, u)$ of the traditional 4D light field, where x and u represent respectively a spatial and an angular dimension. Generalization to the 4D parameterization as defined in [28] is straightforward due to the separability of the Fourier transform. The main notations of the paper are summarized in Table 1.

A view of index j at angular coordinate u_j is defined by $L_j(x) = L(x, u_j)$ and its Fourier transform is $\hat{L}_j(\omega)$, where ω is the spatial frequency. The Fourier Disparity Layer model consists of a set of n layers represented in the Fourier domain and noted $\hat{l}_k(\omega)$ with $k \in [\![1, n]\!]$. Each layer is associated to a disparity value d_k. A view L_j of the light field is then predicted in the Fourier domain by the FDL model as:

$$\hat{L}_j(\omega) = \sum_{k=1}^{n} e^{2i\pi u_j d_k \omega} \hat{l}_k(\omega), \tag{1}$$

where the multiplication by $e^{+2i\pi u_j d_k \omega}$ is interpreted in the spatial domain as a shift by $u_j \cdot d_k$. Intuitively, it corresponds to the displacement of a point of disparity d_k between the reference view (at $u = 0$) and the view L_j.

The problem is now to construct the FDL from m input views L_j ($j \in [\![1, m]\!]$) with discrete spatial sampling.

3.1 Case of complete view sampling

To introduce the matrix notations that will be used throughout the paper, let us first state the problem as in [10], assuming that each input view forms a band-limited signal without missing sample. The discrete Fourier transform is obtained at every frequency $\omega_q = q/N$, where N is the number of pixels in each view and $q \in [\![0, N-1]\!]$. According to Eq. (1), the FDL construction then consists in solving the following linear problem for each frequency ω_q:

$$\mathbf{x_q} = \arg\min_{\mathbf{x}} \|\mathbf{A_q x} - \mathbf{b_q}\|_2^2 + \lambda \|\mathbf{x}\|_2^2, \tag{2}$$

where λ is the l_2 regularization parameter and where we define the matrices $\mathbf{A_q} \in \mathbb{C}^{m \times n}$, and vectors $\mathbf{b_q} \in \mathbb{C}^{m \times 1}$ and $\mathbf{x_q} \in \mathbb{C}^{n \times 1}$ with their elements indexed by j and k:

$$(\mathbf{A_q})_{jk} \triangleq e^{2i\pi u_j d_k \omega_q}, \tag{3}$$

$$(\mathbf{b_q})_j \triangleq \hat{L}_j(\omega_q), \tag{4}$$

$$(\mathbf{x_q})_k \triangleq \hat{l}_k(\omega_q). \tag{5}$$

Note that a more elaborate regularization term was used in [10], but we will consider a l_2 regularization here for simplicity. This problem has a well-known solution:

$$\mathbf{x_q} = \left(\mathbf{A_q}^* \mathbf{A_q} + \lambda \mathbf{I_n}\right)^{-1} \mathbf{A_q}^* \mathbf{b_q}, \tag{6}$$

where $\mathbf{I_n}$ is the identity matrix of size $n \times n$, and $\mathbf{A_q}^*$ denotes the conjugate transpose of $\mathbf{A_q}$.

3.2 Case of incomplete view sampling

Now, in the case where the input views contain unknown samples, their discrete Fourier transform is unknown. However, by defining a mask μ_j such that $\mu_j(x) = 0$ if $L_j(x)$ is unknown, then $\mu_j(x) \cdot L_j(x)$ is known for all x and its discrete Fourier transform can be computed numerically. We can then redefine the vectors $\mathbf{b_q}$ from Eq. (4) with $(\mathbf{b_q})_j = \widehat{[\mu_j L_j]}(\omega_q)$. Knowing that the multiplication in the spatial domain corresponds to a circular convolution in the discrete Fourier domain, and using Eq. (1), the relation with the FDL model can then be obtained:

$$(\mathbf{b_q})_j = \frac{1}{N} \sum_{q'=0}^{N-1} \hat{L}_j(\omega_{q'}) \cdot \hat{\mu}_j((\omega_q - \omega_{q'})_{\mathrm{mod}\,1}) \tag{7}$$

$$= \frac{1}{N} \sum_{q'=0}^{N-1} \sum_{k=1}^{n} e^{2i\pi u_j d_k \omega q'} \hat{l}_k(\omega_{q'}) \cdot \hat{\mu}_j((\omega_q - \omega_{q'})_{\mathrm{mod}\,1}). \tag{8}$$

where $a_{\mathrm{mod}\,b}$ represents the remainder of the euclidean division of a by b. Then, we want to find the disparity layers that verify this equation for every $q \in [\![0, N-1]\!]$ and $j \in [\![0, m]\!]$.

In order to express this linear relation in matrix form, let us first define the diagonal matrices $\mathbf{M_{q,q'}}$ with their diagonal terms

$$(\mathbf{M_{q,q'}})_{j,j} = \frac{1}{N} \cdot \hat{\mu}_j((\omega_q - \omega_{q'})_{\mathrm{mod}\,1}) \tag{9}$$

From Eq. (8), we can then write the FDL construction problem as the following linear optimization:

$$\mathbf{x} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_2^2, \tag{10}$$

where $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{x}$ are defined by:

$$\mathbf{A} \triangleq \begin{pmatrix} \mathbf{M_{0,0}A_0} & \dots & \mathbf{M_{0,N-1}A_{N-1}} \\ \vdots & & \vdots \\ \mathbf{M_{N-1,0}A_0} & \dots & \mathbf{M_{N-1,N-1}A_{N-1}} \end{pmatrix}, \tag{11}$$

$$\mathbf{b} \triangleq \left(\mathbf{b_0}^\top, \dots, \mathbf{b_{N-1}}^\top\right)^\top, \tag{12}$$

$$\mathbf{x} \triangleq \left(\mathbf{x_0}^\top, \dots, \mathbf{x_{N-1}}^\top\right)^\top. \tag{13}$$

Solving this problem constructs the FDL directly in the Fourier domain while ignoring the contribution in the spatial domain of the unknown pixels. Therefore, reconstructing the views from the resulting FDL using Eq. (1) naturally completes the missing pixels. However, while the problem has a solution similar to Eq. (6), the size of the matrix $\mathbf{A}$ is $mN \times nN$, which makes the matrix inversion impractical (the number of pixels N in each view can be very large).

In Section 4, we derive practical algorithms from these results. In particular, we show in Section 4.1 how Eq. (10) can be split into smaller size problems in the case of super-resolution, where the unknown pixels form a regular grid.

3.3 Depth-Dependent Deconvolution

Previously, we have assumed that the known pixels where not filtered (e.g. blurred in the capture or in a downsampling process). Let us now generalize the FDL construction for the case of an input light field filtered before the sampling by spatial and angular filters. Since the filtering is performed before sampling, we consider in this section the continuous light field domain $(x, u) \in \mathbb{R}^2$. Hence, for the derivations in this section, we use the notation $L_u(x) = L(x, u)$ for a view u instead of $L_j(x) = L(x, u_j)$.

Let us note L^{filt} the light field filtered by a spatial filter ϕ and an angular filter ψ. The Fourier transform $\hat{L}^{filt}_{u_0}$ of a view $L^{filt}_{u_0}$ of the filtered light field is given by:

$$\hat{L}^{filt}_{u_0}(\omega) = \int_{-\infty}^{+\infty} \hat{L}_{u_0-u}(\omega)\hat{\phi}(\omega)\psi(u)\mathrm{d}u, \tag{14}$$

where the angular filtering is performed as a convolution with ψ in the angular dimension and as a Fourier domain multiplication by $\hat{\phi}$ in the spatial dimension. Using the FDL decomposition in Eq. (1), we obtain:

$$\hat{L}^{filt}_{u_0}(\omega) = \int_{-\infty}^{+\infty} \sum_{k=1}^{n} e^{2i\pi(u_0-u)d_k\omega}\hat{l}_k(\omega)\hat{\phi}(\omega)\psi(u)\mathrm{d}u, \tag{15}$$

$$= \sum_{k=1}^{n} \hat{l}_k(\omega)e^{2i\pi u_0 d_k\omega}\hat{\phi}(\omega) \int_{-\infty}^{+\infty} e^{-2i\pi u d_k\omega}\psi(u)\mathrm{d}u, \tag{16}$$

$$= \sum_{k=1}^{n} e^{2i\pi u_0 d_k\omega}\hat{\phi}(\omega)\hat{\psi}(d_k\omega)\hat{l}_k(\omega). \tag{17}$$

Therefore, the spatio-angular filter $\phi \cdot \psi$ of the light field corresponds to the depth-dependent filter $\hat{\phi}(\omega)\hat{\psi}(d_k\omega)$ for a Fourier Disparity Layer $\hat{l}_k$ of disparity d_k. The deconvolution is directly obtained by solving the FDL optimization problem (i.e. Eq. (2) without or Eq. (10) with completion) and by replacing the definition of $\mathbf{A_q}$ in Eq. (3) by:

$$(\mathbf{A_q})_{jk} \triangleq e^{2i\pi u_j d_k \omega_q}\hat{\phi}(\omega_q)\hat{\psi}(d_k\omega_q). \tag{18}$$

4 Algorithmic Solutions

4.1 Super-resolution problem

In this section, we show that the completion problem greatly simplifies in the case of super-resolution thanks to the highly repetitive pattern of missing pixels. After deriving the new problem formulation, we present a preconditioning approach that further ensures spatial consistency of the results. Finally, we propose a simple post-processing step.

4.1.1 Derivations from the General Completion Problem

Here, we note N_0 the input spatial resolution and S the super-resolution factor. The spatial resolution of the target light field is then $N = SN_0$. Spatial frequencies ω_q are sampled as $\omega_q = q/(SN_0)$, with $q \in [\![0, SN_0 - 1]\!]$. For super-resolution, we can define the same mask μ for all the views as:

$$\mu(x) = \begin{cases} S & \text{if } x_{\bmod S} = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Its discrete Fourier transform $\hat{\mu}$ is then:

$$\hat{\mu}(\omega_q) = \sum_{x=0}^{SN_0-1} \mu(x) \cdot e^{-2i\pi x \frac{q}{SN_0}} \tag{20}$$

$$= S \sum_{x=0}^{N_0-1} e^{-2i\pi x \frac{q}{N_0}} = \begin{cases} SN_0 & \text{if } q_{\bmod N_0} = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Therefore, the matrices $\mathbf{M_{q,q'}}$ defined in Eq. (9) become:

$$\mathbf{M_{q,q'}} = \begin{cases} \frac{SN_0}{N}\mathbf{I_m} = \mathbf{I_m} & \text{if } (q-q')_{\bmod N_0} = 0, \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{22}$$

Using Eqs. (11) and (22), We can now express the construction matrix $\mathbf{A}$ for the super-resolution problem:

$$\mathbf{A} = \begin{pmatrix} \mathbf{D_0} & \dots & \mathbf{D_{S-1}} \\ \vdots & \ddots & \vdots \\ \mathbf{D_0} & \dots & \mathbf{D_{S-1}} \end{pmatrix}, \tag{23}$$

where $\mathbf{D_t}$ (with index $t \in [\![0, S-1]\!]$) is a block-diagonal matrix formed by (see illustration in Fig. 1):

$$\mathbf{D_t} = \mathrm{diag}\left(\mathbf{A_{tN_0}}, \mathbf{A_{tN_0+1}}, ..., \mathbf{A_{tN_0+N_0-1}}\right) \tag{24}$$

Furthermore, computing the discrete Fourier Transform of μL_j gives:

$$(\mathbf{b_q})_j = \widehat{[\mu L_j]}(\omega_q) = S \sum_{x=0}^{N_0-1} L_j(Sx) \cdot e^{-2i\pi x \frac{q}{N_0}} \tag{25}$$

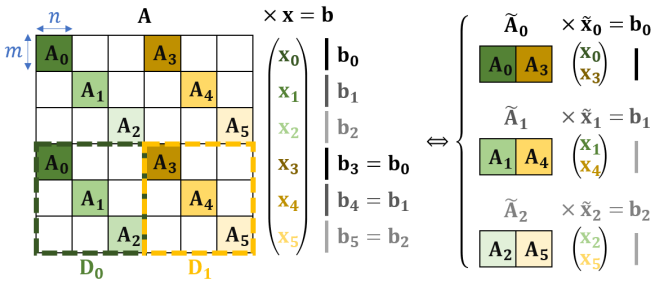

Fig. 1. Simplification of the $mSN_0 \times nSN_0$ super-resolution problem into N_0 small problems of size $m \times nS$. The Illustration shows the case of a super-resolution factor $S = 2$ and an input resolution $N_0 = 3$. In practice, N_0 can be very large (number of pixels of each input view).

Therefore, for indices $r \in [\![0, N_0 - 1]\!]$ and $t \in [\![0, S - 1]\!]$, we have $\mathbf{b_r} = \mathbf{b_{r+tN_0}}$. From this result, and given the block-diagonal structure of the matrices $\mathbf{D_t}$, the problem can then be split into N_0 smaller independent problems of the form:

$$\tilde{\mathbf{x}}_\mathbf{r} = \arg\min_{\mathbf{x}} \left\| \tilde{\mathbf{A}}_\mathbf{r} \mathbf{x} - \mathbf{b_r} \right\|_2^2 + \lambda \left\| \mathbf{x} \right\|_2^2, \tag{26}$$

$$\text{with} \quad \tilde{\mathbf{A}}_\mathbf{r} \triangleq \left(\mathbf{A_r}, \mathbf{A_{r+N_0}}, \ldots, \mathbf{A_{r+(S-1)N_0}} \right), \tag{27}$$

$$\text{and} \quad \tilde{\mathbf{x}}_\mathbf{r} \triangleq \left(\mathbf{x_r}^\top, \mathbf{x_{r+N_0}}^\top, \ldots, \mathbf{x_{r+(S-1)N_0}}^\top \right)^\top. \tag{28}$$

The high resolution FDL, $\hat{l}_k(\omega_q)$ is thus obtained for all frequency indices $q = r + tN_0$ using Eqs. (28) and (5). To facilitate understanding of these derivations, the matrix structure and simplification process are illustrated in Fig. 1.

Note that only small groups of S frequencies are involved in each problem of index r. Since $\tilde{\mathbf{A}}_\mathbf{r} \in \mathbb{C}^{m \times nS}$, the problem size is independent of the input image resolution N_0, making the direct resolution of the linear system practical. Furthermore in Eq. (25), $L_j(Sx)$ corresponds to the low resolution version of the image L_j. The vectors $\mathbf{b_r}$ are then obtained by computing directly the Fourier transform of the low resolution input views. This further accelerates the implementation.

4.1.2 Preconditioning

In practice, depending on the disparity values, the number of views available and their angular coordinates, the super-resolution problem may be ill conditioned. In this situation, the l_2 regularization term in Eq. (26) prevents the frequency components from taking extreme values, but it does not take into account global properties of the frequency spectrum in natural images. In particular, high frequencies are expected to have lower complex magnitudes on average than lower frequencies. In Eq. (26) the frequencies ω_r, ω_{r+N_0}, ..., $\omega_{r+(S-1)N_0}$ are given the same weight although they are very distant in the frequency spectrum. As a result, solving the problem tends to produce peaks of magnitude at high frequencies located at ω_{N_0}, ..., $\omega_{(S-1)N_0}$ due to the peak around the frequency zero (i.e. $\mathbf{b_0}$) naturally contained in the low resolution input. The consequence for the reconstructed views is that the recovered pixels take too low values (see Fig. 2).

In order to solve this issue and ensure spatial consistency, we propose a preconditioning method that weights the respective contributions of the different frequency components. Mathematically, the preconditioned problem is expressed as:

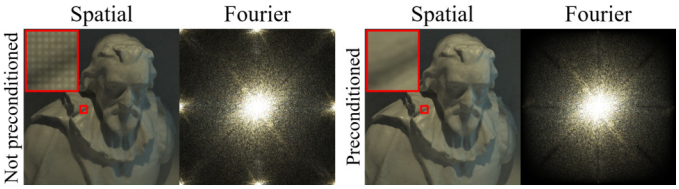

Fig. 2. Super-resolution without preconditioning (left) and with preconditioning (right). Without preconditionning, the recovered pixels have too low values and the Fourier spectrum displays peaks at high frequencies.

$$\begin{aligned} \tilde{\mathbf{y}}_\mathbf{r} &= \arg\min_{\mathbf{y}} \left\| \tilde{\mathbf{A}}_\mathbf{r} \mathbf{P}_\mathbf{r} \mathbf{y} - \mathbf{b_r} \right\|_2^2 + \lambda \left\| \mathbf{y} \right\|_2^2, \\ \tilde{\mathbf{x}}_\mathbf{r} &= \mathbf{P}_\mathbf{r} \tilde{\mathbf{y}}_\mathbf{r}. \end{aligned} \tag{29}$$

where $\mathbf{P_r} \in \mathbb{C}^{nS \times nS}$ is a diagonal preconditioning matrix.

Now, let us consider the worst case where only one view noted L_1 is available. It can be shown that when solving the problem without preconditioning with Eq. (26), the unknown pixels are completed with zeros since there is no other view providing information on these pixels. In the preconditioned problem, we derive the matrices $\mathbf{P_r}$ to change this default behaviour so that in the single view case, the FDL obtained is equal to a reference FDL noted l^{ref} produced as follows:

1) Upsample the view L_1 with an interpolation filter η,
2) Apply the simple FDL construction in Eq. (2) from the already upsampled image.

The interpolation filter η ensures spatial consistency, thus avoiding the too dark completed pixels when the super-resolution problem is ill conditioned. In our implementation we choose the bilinear filter since bilinear interpolation does not affect the known pixels, hence avoiding loss of information in the cases where the problem is already well conditioned.

Since this analysis is based on the single view case, no depth information can be retrieved with multiple layers in the FDL model. Therefore, we can derive $\mathbf{P_r}$ for a simplified single layer FDL model. We show that a direct formula can be obtained for the matrices $\mathbf{Pr}$ satisfying the equality between the reference FDL $\hat{l}^{ref}$ and the result $\hat{l}$ of the preconditioned super-resolution. We provide here the outline of the derivations but more details are given in the supplementary materials.

For a given frequency index $r \in [\![0, N_0 - 1]\!]$, and a single layer k with disparity d_k, we define condensed notations in Table 2. Note that $b = \mathbf{b_r}$ is defined as a scalar since there is only one view.

The goal is then to determine p_t such that $x_t = z_t$. First, we want to find the expression of x_t (i.e. FDL coefficients

TABLE 2
Condensed scalar notations for a given frequency index $r \in [\![0, N-1]\!]$ and for a single layer k.

$\forall t \in [\![0, S-1]\!]$,	$p_t \triangleq (\mathbf{P_r})_{t+1,t+1}$
$a_t \triangleq e^{2i\pi d_k \omega_{r+tN_0}}$	$h_t \triangleq \hat{\eta}(\omega_{r+tN_0})$
$f_t \triangleq \hat{\phi}(\omega_{r+tN_0})\hat{\psi}(d_k\omega_{r+tN_0})$	$x_t \triangleq \hat{l}_k(\omega_{r+tN_0}) = (\tilde{\mathbf{x}}_\mathbf{r})_{t+1}$
$b \triangleq \widehat{[\mu L_1]}(\omega_{r+tN_0})$ $(= \mathbf{b_{r+tN_0}} = \mathbf{b_r})$	$z_t \triangleq \hat{l}_k^{ref}(\omega_{r+tN_0})$

obtained by solving the preconditioned super-resolution in Eq. (29)). Given $\mathbf{P_r}$, Eq. (29) has the general solution:

$$\tilde{\mathbf{x}}_\mathbf{r} = \mathbf{P_r}\left(\mathbf{P_r}^*\tilde{\mathbf{A}}_\mathbf{r}^*\tilde{\mathbf{A}}_\mathbf{r}\mathbf{P_r} + \lambda\mathbf{I_S}\right)^{-1}\mathbf{P_r}^*\tilde{\mathbf{A}}_\mathbf{r}^*\mathbf{b_r}. \tag{30}$$

Note that in the single view case, the matrix $\tilde{\mathbf{A}}_\mathbf{r}$ becomes a row vector. Therefore, the matrix inversion can be further derived using the Sherman-Morrison formula (see details in supplementary materials). The solution obtained is then:

$$\tilde{\mathbf{x}}_\mathbf{r} = \frac{\mathbf{P_r}\mathbf{P_r}^*\tilde{\mathbf{A}}_\mathbf{r}^*}{\lambda + \tilde{\mathbf{A}}_\mathbf{r}\mathbf{P_r}\mathbf{P_r}^*\tilde{\mathbf{A}}_\mathbf{r}^*}\mathbf{b_r}. \tag{31}$$

Using the notations in Table 2 and the definitions of $\tilde{\mathbf{A}}_\mathbf{r}$ and $\tilde{\mathbf{x}}_\mathbf{r}$ in Eqs. (27) and (28), we can then rewrite Eq. (31):

$$\forall t \in [\![0, S-1]\!],\ x_t = \frac{|p_t|^2 \cdot \overline{a_t f_t}}{\lambda + \sum_{t'=0}^{S-1}\left(|p_{t'}|^2 \cdot |f_{t'}|^2\right)} \cdot b \tag{32}$$

Now, let us find the expression of z_t (i.e. reference FDL coefficients). Upsampling an image with filter η is performed by first upsampling with the mask μ (i.e. set intermediate pixels to zero), and then applying the filter η. Hence, at each frequency ω_q with $q = r + tN_0$, the upsampled view is given by $\hat{\eta}(\omega_q)\widehat{[\mu L_1]}(\omega_q) = h_t b$. The reference FDL is then obtained by solving the simple FDL construction in Eq. (2) where the input $\mathbf{b_q}$ is replaced by $h_t b$. Using the general solution in Eq. (6) expressed with the scalar notations, we obtain (see details in supplementary materials):

$$z_t = \frac{\overline{a_t f_t}}{\lambda + |f_t|^2} \cdot h_t b \tag{33}$$

By enforcing the equality $x_t = z_t$, we obtain the following equations for the diagonal elements p_t of the preconditioning matrix $\mathbf{P_r}$:

$$\forall t \in [\![0, S-1]\!],\ |p_t|^2 \cdot \left(\frac{\lambda + |f_t|^2}{h_t}\right) - \sum_{t'=0}^{S-1}|p_{t'}|^2 \cdot |f_{t'}|^2 = \lambda. \tag{34}$$

We show in the supplementary materials that this problem is solved by using:

$$p_t = \sqrt{\frac{\lambda \cdot h_t}{(\lambda + |f_t|^2) \cdot \left(1 - \sum_{t'=0}^{S-1}\left(h_{t'}|f_{t'}|^2/(\lambda + |f_{t'}|^2)\right)\right)}}, \tag{35}$$

The diagonal pre-conditioning matrix is then formed as $\mathbf{P_r} = \mathrm{diag}(p_0, ..., p_{S-1})$. More generally, for the case of multiple layers, the size of $\mathbf{P_r}$ is $nS \times nS$ instead of $S \times S$. We directly generalize from the single layer case by computing for each layer k the disparity-dependent filter coefficients f_t (see Table 2). The corresponding values p_t are then computed with Eq. (35) and arranged in the diagonal matrix $\mathbf{P_r}$ according to the layer index k and frequency index t.

4.1.3 Post-Processing: Residual Back-Projection

Because of the l_2 regularization term, some loss of information occurs when solving the super-resolution optimization problem. As a result, downsampling the super-resolved light field (with the same filters ϕ and ψ used in the model) does not recover exactly the original low resolution input. While this might be a desirable filtering effect (e.g. noise reduction, brightness consistency between views) and is necessary for an effective super-resolution, it may also cause some artifacts (e.g. around occlusions areas). Therefore, we additionally propose a simple post-processing step, similar to the back-projection approach commonly used in super-resolution, to ensure that the super-resolved light field matches the low resolution input.

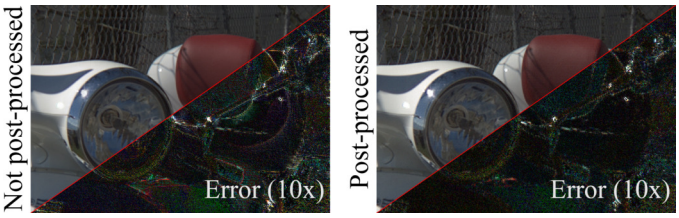

Fig. 3. Super-resolution results on the natural light field 'Vespa'. Left: without post-processing (PSNR=36.25dB). Right: with post-processing (PSNR=37.59dB). The bottom right part of each image represents to the absolute error with respect to the ground truth (magnified 10x).

For that purpose, at each frequency ω_r of the low resolution input ($r \in [\![0, N_0 - 1]\!]$), we compute the residual $\mathbf{b_r} - \tilde{\mathbf{A}}_\mathbf{r}\tilde{\mathbf{x}}_\mathbf{r}$ between the original light field views and the ones obtained with super-resolution and downsampling. For each view, the residual is simply upsampled by zero-padding in the Fourier domain (i.e. unchanged for frequencies below ω_{N_0}, and set to zero for frequencies between ω_{N_0} and ω_N). Finally, the residual is deconvolved with the spatial filter ϕ (i.e. division by $\hat{\phi}$ in the Fourier domain), and added back to the view. Note that only the spatial filter can be inverted for the residual deconvolution since the post-processing is performed independently for each view. An example of result is shown in Fig. 3.

4.2 Algorithm for General Completion

When the input views have an irregular spatial sampling, the general completion problem cannot be simplified. Instead of solving directly the very large problem presented in Section 3.2, we propose an iterative algorithm (see Algorithm 1). The algorithm iteratively filters the light field thanks to a FDL model (steps 2 and 3), and updates only the unknown pixels by replacing them with their filtered values (step 5). Note that forward and inverse Fourier transforms are needed (steps 1 and 4) since the filtering is performed in the Fourier domain while the update is performed in the pixel domain.

Similarly to the original problem in Section 3.2, this approach constructs the FDL by only constraining the values

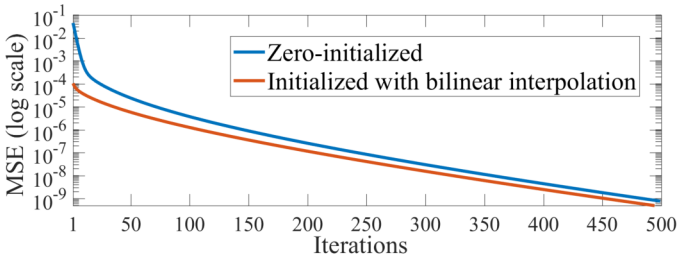

Fig. 4. Convergence of the iterative completion algorithm in the particular case of x2 super-resolution. The reference for computing the mean square error (MSE) is the result of the direct super-resolution method from Section 4.1. The MSE tends to zero, showing the convergence to the same solution as the direct super-resolution.

Fig. 5. Demosaicing test where each view of a synthetic light field was masked by the Bayer color pattern. Left: one view of the ground truth light field, Center: FDL super-resolution applied independently to RGB components, Right: FDL super-resolution with color regularization.

Algorithm 1: Iterative FDL completion

Data: Set of incomplete views L_j with an initialization of the unknown pixels.

while *not converged* **do**

1. Compute Fourier transform $\hat{L}_j$ of each view.
2. Construct a FDL model from the current views (e.g. with Eq. (2)).
3. Reconstruct the views $\hat{L}'_j$ from the FDL model.
4. Compute inverse Fourier Transforms L'_j.
5. Update only the unknown pixels of each view L_j with the corresponding pixels in L'_j.

of the known pixels in the input data. We show in Fig. 4 that in the particular case of super-resolution (i.e. regular grid of known pixels), this algorithm converges to the same solution as the direct algorithm presented in Section 4.1.

4.3 Demosaicing Problem

In the case of color images with a color mosaic filter (e.g. bayer pattern), only one of the red green or blue components is known at each pixel. While the super-resolution or completion algorithms may be applied independently to each component, improved results can be obtained by considering the strong correlations between color components. These correlations are generally exploited in demosaicing methods using an assumption of either color difference constancy or color ratio constancy [29]. Following the color difference constancy assumption, the intensity variations (e.g. image gradients) of the red, green and blue components should be similar. This prior can be encouraged in our optimization approach by introducing a regularization term that penalizes, for each layer l^k, the differences between the intensity variations of its red, green and blue components.

In order to define the color regularization, let us first reformulate the simple FDL construction problem in Eq.(2) for multiple components. For a color index $c \in \{R, G, B\}$, we note l_k^c and L_j^c the corresponding color component of the layer l_k and view L_j respectively. Here, we define the matrices of input views and layers coefficients $\mathbf{B_q} \in \mathbb{C}^{m\times 3}$ and $\mathbf{X_q} \in \mathbb{C}^{n\times 3}$ as $(\mathbf{B_q})_{j,c} \triangleq L_j^c(\omega_q)$ and $(\mathbf{X_q})_{k,c} \triangleq l_k^c(\omega_q)$. The vectors $\mathbf{b_q}$ and $\mathbf{x_q}$ from Eqs. (4) and (5) can then be replaced by the matrices $\mathbf{B_q}$ and $\mathbf{X_q}$ in the original problem of Eq. (2). Now, we can introduce a color regularization term of the form $\left\|\mathbf{X_q}\mathbf{G_q}^\top\right\|_2^2$, where $\mathbf{G_q}$ is a matrix computing the differences in intensity variations between RGB components. In our implementation, we use:

$$\mathbf{G_q} = \hat{\gamma}(\omega_q)\begin{pmatrix} -2 & 0.5 & 1.5 \\ 2 & 0 & -2 \end{pmatrix}, \tag{36}$$

where γ is the filter for computing intensity variations. For our implementation, we use the discrete Laplacian filter to represent variations in both horizontal and vertical directions for 2D images with a single filter. The weights in Eq. (36) were determined empirically to regularize more strongly the red and blue components than the green component that is usually more reliable.

The color regularized FDL construction (without completion or super-resolution) is then:

$$\min_{\mathbf{X}} \|\mathbf{A_q X} - \mathbf{B_q}\|_F^2 + \lambda \|\mathbf{X}\|_F^2 + \lambda_{col} \left\|\mathbf{X}\mathbf{G_q}^\top\right\|_F^2, \tag{37}$$

where λ_{col} is the color regularization parameter and $\|.\|_F$ denotes the Frobenius norm. Because of the right multiplication by $\mathbf{G_q}$, the problem cannot be directly solved in this matrix form. It must be expressed equivalently in vectorized form using the vectorization operator vec and the Kronecker product $\otimes$ as follows:

$$\begin{aligned}\min_{\mathrm{vec}(\mathbf{X})} \|(\mathbf{I_3} \otimes \mathbf{A_q})\mathrm{vec}(\mathbf{X}) - \mathrm{vec}(\mathbf{B_q})\|_2^2 + \lambda \|\mathrm{vec}(\mathbf{X})\|_2^2 + \\ + \lambda_{col} \|(\mathbf{G_q} \otimes \mathbf{I_n})\mathrm{vec}(\mathbf{X})\|_2^2 .\end{aligned} \tag{38}$$

Therefore, we obtain a least squares problem with Tikhonov regularization that can be solved efficiently. Note, however, that due to the Kronecker product $\mathbf{I_3}\otimes\mathbf{A_q}$, the size of the problem is multiplied by 3 compared to the version without color regularization.

In order to perform demosaicing, the completion algorithm presented in Section 4.2 can be applied by replacing the simple FDL construction in step 2, by the color regularized problem of Eq. (38). Alternatively, when the known pixels of each view form a regular pattern, the color regularization can be adapted to the super-resolution problem in Section 4.1. For that purpose, we can similarly reformulate the super-resolution for multiple color components by arranging the layer's coefficients in the matrices $\tilde{\mathbf{X}}_\mathbf{r} \triangleq \left(\mathbf{X_r}^\top, \mathbf{X_{r+N_0}}^\top, ..., \mathbf{X_{r+(S-1)N_0}}^\top\right)^\top$ for each frequency index $r \in [\![0, N_0 - 1]\!]$ (the matrices $\tilde{\mathbf{B}}_\mathbf{r}$ can be defined accordingly). The color regularization term then becomes $\lambda_{col}\left\|(\tilde{\mathbf{G}}_\mathbf{r} \otimes \mathbf{I_n})\mathrm{vec}(\tilde{\mathbf{X}}_\mathbf{r})\right\|_2^2$ with $\tilde{\mathbf{G}}_\mathbf{r}$ defined as

$$\tilde{\mathbf{G}}_\mathbf{r} \triangleq \left((\mathbf{G_r} \otimes \mathbf{e_1})^\top, ..., (\mathbf{G_{r+tN_0}} \otimes \mathbf{e_{t+1}})^\top, ...\right)^\top, \tag{39}$$

where $\mathbf{e_t}$ is a $1 \times S$ row vector with element t equal to 1 and all the other elements equal to 0.

We show in Fig. 5 the effect of the color regularization for the demosaicing problem in a synthetic example where the input light field was generated by applying the traditional Bayer pattern to each view.

5 Application: RAW Light Field Extraction

In this section, we show how the methods presented in Section 4 can be used to extract high quality light field views from the RAW data of plenoptic cameras. In particular, we

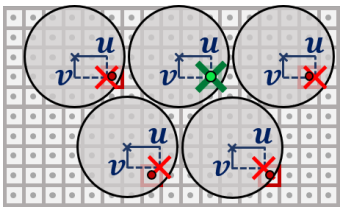

Fig. 6. RAW sampling with imperfect alignment between sensor and microlenses. The view at angular coordinates (u, v) has only one pixel known accurately (at the green cross location). For most microlenses the exact (u, v) position (red crosses) falls between several pixels.

study the case of the unfocused plenoptic cameras [1] where the image formed behind each microlens corresponds to an angular patch at a given spatial coordinate.

5.1 Initial Views and Masks Extraction

In a first step, the spatio-angular samples must be extracted from the RAW data. In the ideal case where the microlens array and the sensor are perfectly aligned and where the distance between the microlense's centers is a multiple of the pixel size, the number of views sampled in the RAW data is equal to the number of pixels behind each microlens. However, due to imperfect alignement, the angular patches formed behind each microlens have different angular samplings. As a result, at a given angular coordinate (i.e. view position), only a few pixels may be available (see Fig. 6). Therefore, a large number of views must be extracted in order to contain all the pixels originally captured.

Conventional view extraction methods [6], [9] use 2D interpolation (e.g. bicubic, bilinear) to extract a number of views corresponding to the number of samples of each angular patch, which results in loss of captured information. Instead, we sample the views with half pixel precision and use an interpolation to the nearest pixel. We additionally extract a mask for each view indicating the positions of the pixels known accurately. The mask is defined by assigning each pixel of the sensor only to the closest extracted view. In order to obtain an initial estimation of the complete color information, we perform a 2D demosaicing of the RAW data using the method in [30] before extracting the views.

5.2 Calibration and Chromatic Aberrations

Although the Fourier Disparity Layer construction does not require a disparity map, it relies on the disparity value d_k of each layer l_k and the angular coordinate u_j of each input view L_j. In order to determine the set of parameters ensuring an optimal reconstruction, we use the FDL calibration proposed in [10] applied to the light field initially extracted as described in Section 5.1 (we do not use the mask here).

We additionally address the problem of chromatic aberrations by performing the calibration independently for each color component. In particular, axial chromatic aberrations of the main lens result in different colors being in focus at different depths. In the light fields captured by plenoptic cameras, the disparity depends on the the depth in focus during the capture. Hence, axial chromatic aberrations can be removed by constructing the layers with different disparity values for the red, green and blue components. Then, the corrected views can be rendered from the FDL model using Eq. (1) with the same parameters for all the

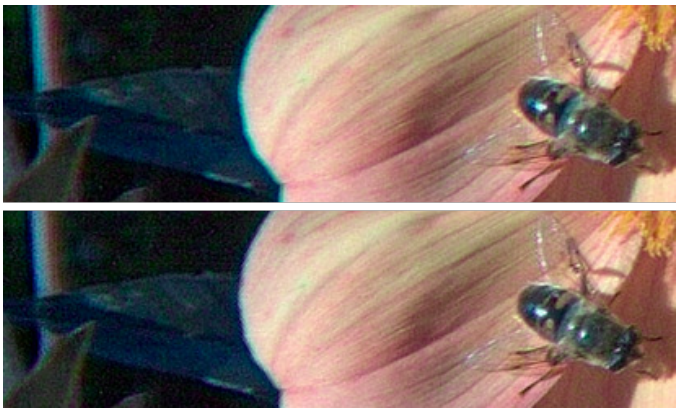

Fig. 7. Correction of axial chromatic aberrations. Top: without correction. Bottom: with correction (FDL constructed from parameters calibrated per-component). An external view is shown here since axial chromatic aberrations increase for views with high angular coordinates (i.e. further away from the central view).

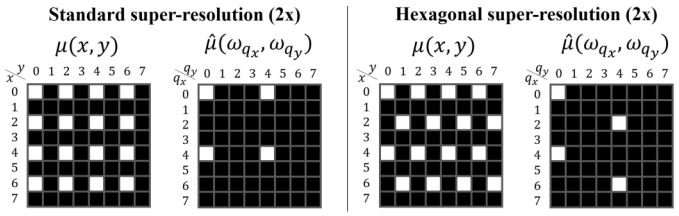

Fig. 8. 2D masks for 2x super-resolution from either a square or an hexagonal sampling. The mask in the Fourier domain $\hat{\mu}$ only has 4 non-zero coefficients for both square and hexagonal samplings.

component (in practice we use the average of the parameters of the 3 components). An example is shown in Fig. 7.

Note that due to a second type of chromatic aberrations referred to as transversal, the color components may appear to be scaled with different factors. This cannot be explicitly compensated in our Fourier domain optimization. However, using a model where the position of each view can also be adjusted for each color component helps reduce moderate transversal chromatic aberrations.

5.3 Completion, Super-Resolution and Demosaicing

Given the calibrated parameters, we can construct the Fourier Disparity Layer model that best fits the known pixels in the extracted views. Due to the irregular sampling of the plenoptic RAW data, we must use the the iterative version of our completion algorithm presented in Section 4.2. However, completing the missing pixels of the extracted views still results in a low spatial resolution. Furthermore, constructing the layers in low resolution has a filtering effect on the spatial aliasing of the reconstructed views, which loses valuable information for performing super-resolution later in the process. Therefore, we perform the super-resolution jointly with the completion. For that purpose, the super-resolution with preconditioning in Eq.(29) is used in step 2 of the completion algorithm. The low resolution views are reconstructed from the high resolution layers (as $\tilde{\mathbf{A}}_\mathbf{r}\tilde{\mathbf{x}}_\mathbf{r}$) in step 3, hence preserving the aliasing. Only at the last iteration, the high resolution views are reconstructed from the layers. Demosaicing is performed simultaneously by using the color regularization term in Eq. (39).

Some plenoptic cameras use an hexagonal array of micro-lenses to optimize the spatial sampling (e.g. Lytro

TABLE 3
PSNR results and average runtimes for 2x super-resolution where the low resolution light field is generated using either the 'box' filter or the gaussian filter (with standard deviation $\sigma = 1$) for downsampling. The PSNR results correspond to the average PSNR of all the views. The runtimes were obtained using a computer with 32GB RAM, Intel Core i7-7700 CPU and Nvidia GeForce GTX 1080 GPU. For each test, the best and the second best results are shown in bold red and blue respectively.

	From 'box' filter downsampling						**From gaussian filter downsampling ($\sigma = 1$)**					
	Bicubic	VDSR [31]	PCA+RR [21]	GB-SQ [19]	Deep LFSSR [27]	FDL SR (ours)	Bicubic	VDSR [31]	PCA+RR [21]	GB-SQ [19]	Deep LFSSR [27]	FDL SR (ours)
Fruits (INRIA)	32.68	33.63	32.50	35.86	**37.98**	**36.26**	30.44	30.77	31.37	**34.04**	32.28	**35.12**
Figurines (INRIA)	36.64	36.99	35.54	**37.21**	**37.98**	36.98	35.17	35.71	35.07	**36.59**	36.06	**37.44**
Friends (EPFL)	36.65	37.61	37.07	37.93	**39.99**	**38.05**	34.66	35.77	35.70	**37.11**	36.18	**38.02**
Vespa (EPFL)	35.45	36.72	35.67	**38.24**	**39.67**	37.59	33.40	34.57	34.42	**36.56**	35.06	**37.20**
Bikes (EPFL)	33.81	35.51	34.80	**36.70**	**39.08**	35.98	31.37	32.40	32.87	**35.30**	33.28	**36.04**
butterfly (HCI old)	37.05	38.26	37.00	39.35	**40.79**	**40.66**	35.31	36.64	36.06	**36.82**	36.66	**39.46**
sideboard (HCI)	27.21	28.14	27.52	**30.86**	**31.50**	30.64	25.56	25.77	26.51	**28.39**	27.06	**29.32**
dino (HCI)	35.57	37.24	35.59	**38.39**	**40.23**	38.35	33.75	34.88	34.63	**35.96**	35.26	**37.49**
cotton (HCI)	43.65	44.30	42.45	46.94	**47.16**	**47.89**	41.85	42.79	41.85	**43.83**	43.21	**45.92**
boxes (HCI)	33.30	34.09	32.56	**36.27**	**37.46**	34.97	31.42	31.20	31.99	**34.20**	33.09	**35.56**
pyramids (HCI)	26.92	26.74	26.94	27.69	**27.84**	**31.29**	26.01	26.10	26.20	25.96	**26.27**	**28.88**
Average PSNR	34.45	35.38	34.33	36.86	**38.15**	**37.15**	32.63	33.33	33.33	**34.98**	34.04	**36.40**
Average Runtime	**1.6s**	17s	95s	~1h30m	48s	**5.8s**	**1.6s**	17s	95s	>10h	51s	**5.9s**

cameras). This results in an hexagonal grid of known pixels in each view instead of the more common square grid. Our super-resolution method can be adapted to this case by using an hexagonal 2D mask μ in the derivations, as illustrated in Fig. 8. Thanks to the sparsity of the mask $\hat{\mu}$ in the Fourier domain (e.g. 4 non-zero coefficients for 2x super-resolution), the hexagonal super-resolution problem simplifies in the same way as presented in Section 4.1. We only need to adapt the indices of the frequency coefficients involved in the same minimization (i.e. used in the construction of each matrix $\tilde{\mathbf{A}}_\mathbf{r}$ and vectors $\tilde{\mathbf{x}}_\mathbf{r}$ and $\tilde{\mathbf{b}}_\mathbf{r}$).

6 Experimental Results

6.1 Light Field Super-Resolution

6.1.1 Experimental Setup

We first evaluate our super-resolution algorithm presented in Section 4.1 in comparison with state-of-the-art light field super-resolution methods: the GB-SQ method [19] using graph-based optimization; the PCA+RR method [21] that learns a mapping between low and high resolution patch volumes in a linear subspace domain; the deep convolutional network recently proposed in [27] (Deep LFSSR) that performs separable convolutions between the spatial and angular dimensions. We also include in our comparisons two reference single image super-resolution methods applied per-view: the bicubic upsampling and the VDSR method [31] based on a very deep convolutional network.

For this experiment, we downsample a light field with a known blurring kernel ϕ. The downsampled version is used as input of the super-resolution algorithm and the original version serves as a ground truth for quality evaluation. For a fair comparison, we do not use the angular blur ψ, since the other methods do not allow depth-dependent deconvolution. However, to evaluate the generalization capability of the different methods we have tested two spatial blur kernels. First, the 'box' filter averages blocks of $S \times S$ pixels (S being the super-resolution factor). It is used in [27], for generating the training low resolution dataset. The second filter tested is the discrete approximation (using a 7x7 kernel) of a gaussian filter with standard deviation $\sigma = 1$. Note that only our method and the GB-SQ method [19] can be parameterized given a known filter ϕ [1]. For the deep learning approaches (VDSR [31] and DeepLFSSR [27]), we have used the available pre-trained model.

The 2x super-resolution results are presented in Table 3 for natural light fields from the datasets of INRIA [32] and EPFL [33] (captured with a Lytro Illum plenoptic camera), as well as synthetic light fields from the HCI dataset [34]. For each tested light field, Table 3 reports the average peak signal-to-noise ratio (PSNR) of all the views. The average runtime of all the light fields is also given (for our FDL super-resolution, the runtime includes the FDL calibration).

6.1.2 Results and Discussion

The results in Table 3 confirm that the methods designed for light fields successfully exploit angular priors and generally obtain better results than single image methods (bicubic, VDSR).

When the low resolution input is generated using the 'box' filter, the best results are obtained with the Deep LFSSR method [27]. However, the performance of this method drops for the gaussian dowsampling (larger kernel than the 'box' filter) and the super-resolved light fields remain blurred (see Fig. 9). This is explained by the fact that the Deep LFSSR method specifically learns a mapping between low resolution light fields generated with a fixed downsampling operator, and their original high resolution versions. The model learned for the 'box' filter downsampling does not generalize well to other filters. Note that this is a limitation of most learning based approaches, which also include the PCA+RR method [21].

The PSNR performances of our FDL method are generally comparable to GB-SQ [19] in the case of the 'box' filter. However, some differences can be observed on the visual results in Fig.9. For a similar PSNR, while the FDL super-resolution may create more artifacts, it better preserves the

1. The author's implementation of GB-SQ was originally designed for the 'box' filter but could be easily adapted for different filters. However, computational complexity significantly increases with larger kernel sizes (see running times in Table 3).

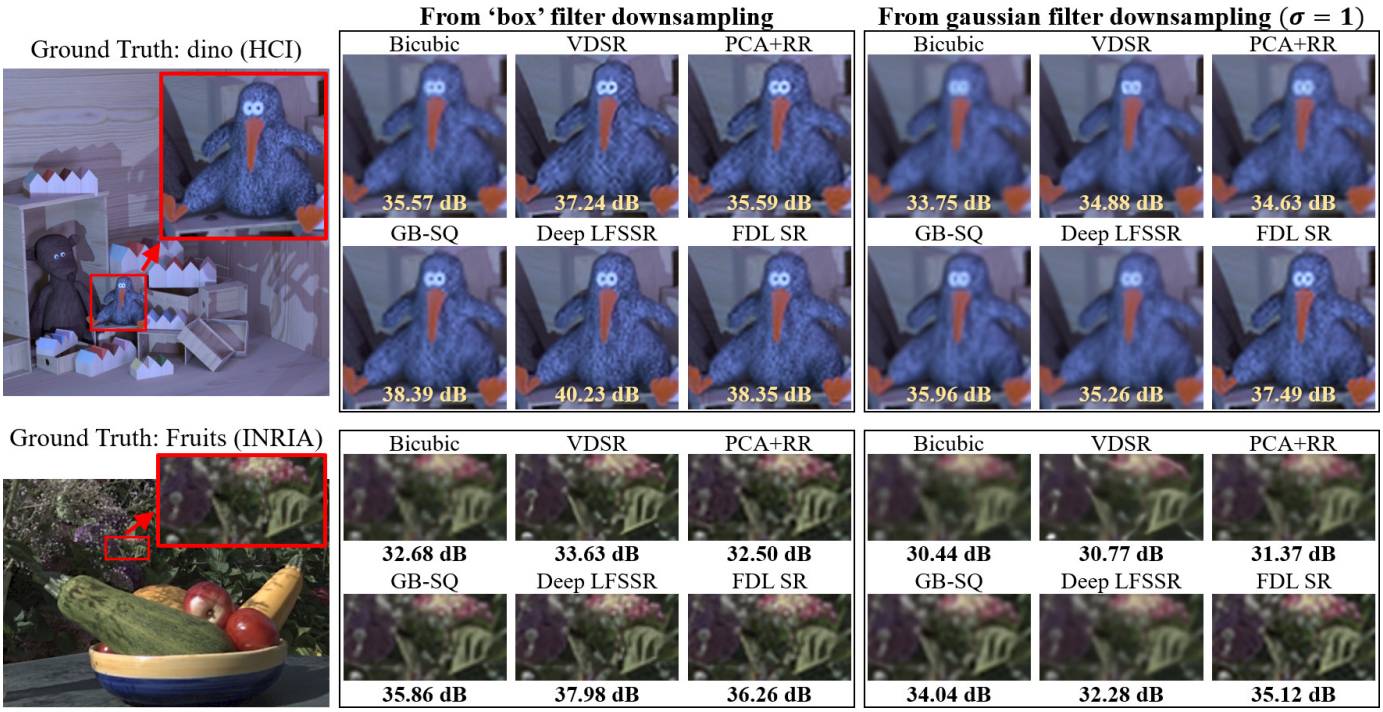

Fig. 9. Visual comparisons for the x2 super-resolution of the fields 'dino' (HCI) and 'Fruits' (INRIA). PSNR values are indicated in each subfigure.

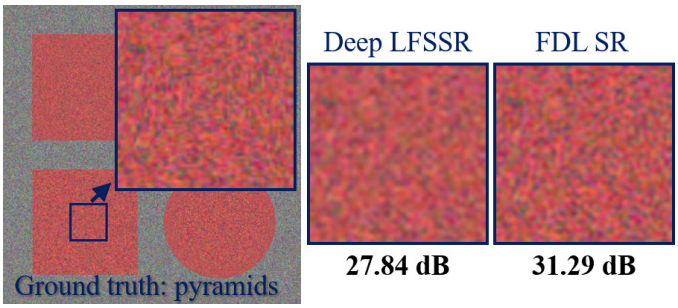

Fig. 10. x2 Super-resolution of the light field 'pyramids' (HCI) in the 'box' filter case for the Deep LFSSR and FDL SR methods.

fine details. This is particularly advantageous when the low resolution input data is more blurred (i.e gaussian filter downsampling). In this case, the FDL super-resolution obtains the best results among all the tested methods.

In order to compare the performances of the angular priors, we have included the light field 'pyramids' which is a lambertian scene containing 3D shapes with very noisy textures. This scene is very challenging for single image methods (only using spatial priors), and can be better tackled using angular priors. For this light field, a large PSNR gain is observed for our method compared to the other approaches (e.g. more than 3dB gain with respect to Deep LFSSR even in the 'box' filter case). The result is shown in Fig. 10. This illustrates the efficiency of the FDL model for exploiting angular consistency and preserving fine details.

Finally, regarding the computational complexity, the FDL super-resolution is faster than the other tested methods at the exception of the bicubic upsampling. Furthermore, it is not significantly affected by the filter used, unlike the GB-SQ method which has a high complexity that is further increased by using the gaussian filter (7x7 kernel instead of 2x2 for the 'box' filter). Note that our method fully takes advantage of GPU computations thanks to the possibility to process different frequencies in parallel. The reported runtimes for the deep learning approaches (i.e. VDSR and Deep LFSSR) were also obtained with GPU implementations.

6.2 RAW Light Field Extraction

Here, we present the result of the view extraction method presented in Section 5. For the experiment, we have used light fields captured with Lytro Illum cameras from the INRIA [32] and EPFL [33] datasets. Our view extraction performs x2 super-resolution (jointly with completion and demosaicing) using the hexagonal pattern of the Lytro Illum camera (see Fig. 8). For the spatial and angular filters ϕ and ψ, we have used gaussian filters with respective standard deviations $\sigma_\phi = 0.33$ and $\sigma_\psi = 1.3$. For the FDL model construction, we use 20 layers and 30 iterations.

For the comparison, we show the results produced by the official Lytro Desktop software (which includes super-resolution), as well as the light fields extracted by a conventional pipeline in [9] and super-resolved with the Deep LFSSR method [27]. Note that [9] is based on the popular light field toolbox in [6], and improves several aspects of the view extraction such as the interpolations, color balance and color consistency between views. However, it remains a sequential approach with successive demosaicing and resampling steps. In order to keep all the details of the extracted views before applying the super-resolution, we did not include the last denoising step suggested in [9].

Since no ground truth is available here, we only present visual results in Fig. 11. Additional results are also presented in the supplementary materials. Finer details are revealed in the light field extracted with our FDL approach. This is made possible by treating simultaneously the demosaicing, resampling, and super-resolution in the same optimization. On the other hand, in the sequential approach (i.e. Sequential Pipeline + DeepLFSSR in Fig. 11), artifacts and detail loss accumulate at each step resulting in a low quality even

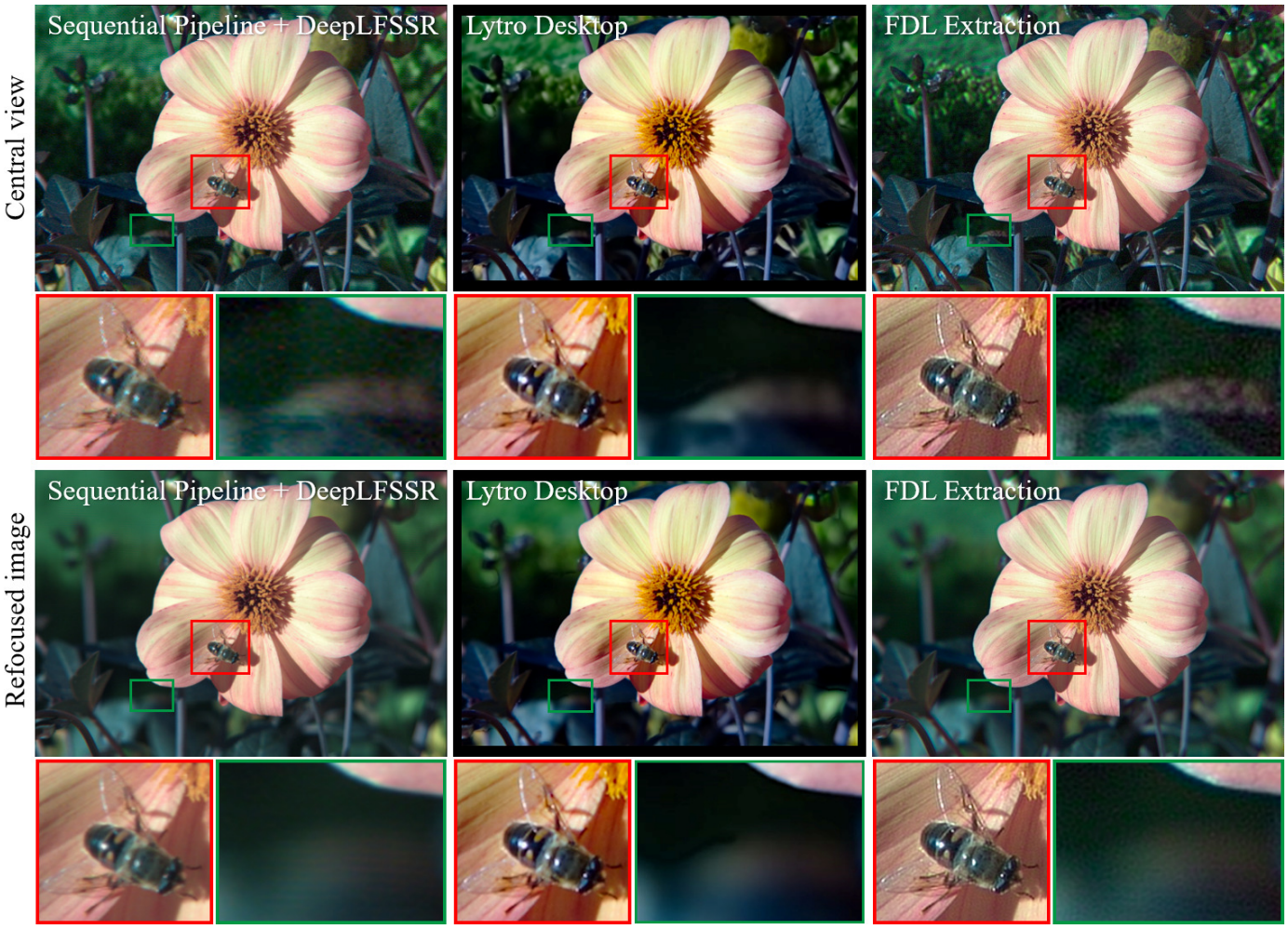

Fig. 11. Light field extraction results from a Lytro Illum RAW image ('Bee2' from the INRIA dataset [32]). Left: conventional pipeline of [9] followed by super-resolution using the Deep LFSSR method [27]. Center: Lytro Desktop software. Right: our method combining FDL completion, super-resolution and demosaicing. Top: central view. Bottom: refocused image from the extracted light field (focused on the bee).

after super-resolution. Furthermore, thanks to the depth-dependent angular filter as well as the sub-pixel angular sampling, we effectively recover the full depth-of-field, while the other methods produce blur in the background and foreground regions far from the focus plane.

7 Conclusion

We have presented a novel approach for high resolution light field recovery based on the Fourier Disparity Layer representation. We have shown that in the Fourier domain, light field super-resolution can be formulated as independent linear optimization problems, each problem only involving a small subset of frequency components. This allows for a fast direct super-resolution. The more general light field completion problem is also addressed with an iterative version of our algorithm that does not require the known pixels to be arranged in a regular grid. We have shown that in the case of demosaicing, where different color components are missing for different pixels, our results are improved by the addition of a regularization term exploiting color component's correlations. For the practical scenario of extracting light fields captured with microlens-based plenoptic cameras, we have proposed a complete optimization framework that jointly performs completion, super-resolution, and demosaicing, while taking into account other optical degradations such as chromatic aberrations and depth-dependent blur. Our approach fully optimizing the whole extraction process can reveal very fine details in real light field captures, allowing us to reach higher image quality than existing RAW light field extraction methods. The evaluation of our super-resolution algorithm alone also shows favorable comparisons with the state-of-the-art, both in terms of accuracy and computational cost.

Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776.

References

[1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-Held Plenoptic Camera," Stanford University, Tech. Rep., Apr. 2005.
[2] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proc. IEEE ICCP*, Apr. 2009.
[3] M. Seifi, N. Sabater, V. Drazic, and P. Perez, "Disparity-guided demosaicking of light field images," in *Proc. IEEE ICIP*, Oct. 2014.
[4] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "An analysis of color demosaicing in plenoptic cameras," in *Proc. IEEE CVPR*, Jun. 2012.

[5] P. David, M. Le Pendu, and C. Guillemot, "White lenslet image guided demosaicing for plenoptic cameras," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2017.

[6] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE CVPR*, 2013.

[7] D. Cho, M. Lee, S. Kim, and Y. W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *Proc. IEEE ICCV*, Dec. 2013.

[8] P. Matysiak, M. Grogan, M. Le Pendu, M. Alain, and A. Smolic, "A pipeline for lenslet light field quality enhancement," in *Proc. IEEE ICIP*, Oct. 2018.

[9] P. Matysiak, M. Grogan, M. Le Pendu, M. Alain, E. Zerman, and A. Smolic, "High quality light field extraction and post-processing for raw plenoptic data," *IEEE Trans. Image Process.*, vol. 29, pp. 4188–4203, 2020.

[10] M. Le Pendu, C. Guillemot, and A. Smolic, "A fourier disparity layer representation for light fields," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5740–5753, Nov. 2019.

[11] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, Mar. 2015.

[12] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 972–986, May 2012.

[13] T. Georgiev, G. Chunev, and A. Lumsdaine, "Superresolution with the focused plenoptic camera," in *Computational Imaging IX*, vol. 7873, 2011, pp. 232–244.

[14] Y. Li and M. Sjöström, "Depth-assisted demosaicing for light field data in layered object space," in *Proc. IEEE ICIP*, Sep. 2019.

[15] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.

[16] R. Farrugia and C. Guillemot, "Light field super-resolution using a low-rank prior and deep convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.

[17] R. A. Farrugia and C. Guillemot, "A simple framework to leverage state-of-the-art single-image super-resolution methods to restore light fields," *Signal Processing: Image Communication*, vol. 80, 2020.

[18] M. Alain and A. Smolic, "Light field super-resolution via LFBM5D sparse coding," in *Proc. IEEE ICIP*, Oct 2018.

[19] M. Rossi and P. Frossard, "Geometry-consistent light field super-resolution via graph-based regularization," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4207–4218, Sep. 2018.

[20] Y. Yoon, H. Jeon, D. Yoo, J. Lee, and I. S. Kweon, "Light-field image super-resolution using convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 848–852, Jun. 2017.

[21] R. A. Farrugia, C. Galea, and C. Guillemot, "Super resolution of light field images using linear subspace projection of patch-volumes," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1058–1071, Oct. 2017.

[22] M. S. K. Gul and B. K. Gunturk, "Spatial and angular resolution enhancement of light fields using convolutional neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2146–2159, May 2018.

[23] Y. Wang, F. Liu, K. Zhang, G. Hou, Z. Sun, and T. Tan, "Lfnet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4274–4286, Sep. 2018.

[24] Y. Yuan, Z. Cao, and L. Su, "Light-field image superresolution using a combined deep cnn based on epi," *IEEE Signal Process. Lett.*, vol. 25, no. 9, pp. 1359–1363, Sep. 2018.

[25] S. Zhang, Y. Lin, and H. Sheng, "Residual networks for light field image super-resolution," in *Proc. IEEE CVPR*, Jun. 2019.

[26] M. Zhu, A. Alperovich, O. Johannsen, A. Sulc, and B. Goldlücke, "An epipolar volume autoencoder with adversarial loss for deep light field super-resolution," in *CVPR Workshops*, 2019.

[27] H. W. F. Yeung, J. Hou, X. Chen, J. Chen, Z. Chen, and Y. Y. Chung, "Light field spatial super-resolution using deep efficient spatial-angular separable convolution," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2319–2330, May 2019.

[28] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96, Aug. 1996, pp. 31–42.

[29] O. Losson, L. Macaire, and Y. Yang, "Chapter 5 - comparison of color demosaicing methods," in *Advances in Imaging and electron Physics*, ser. Advances in Imaging and Electron Physics, P. W. Hawkes, Ed. Elsevier, 2010, vol. 162, pp. 173–265.

[30] H. S. Malvar, L.-W. He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *Proc. IEEE ICASSP*, 2004.

[31] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE CVPR*, Jun. 2016.

[32] "Inria lytro illum dataset," http://clim.inria.fr/IllumDatasetLF/index.html, accessed: 20-11-2019.

[33] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *Proceedings of the International Conference on Quality of Multimedia Experience*, 2016.

[34] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldlücke, "A dataset and evaluation methodology for depth estimation on 4d light fields," in *ACCV*, 2016.

Mikaël Le Pendu received the Engineering degree from the Ecole Nationale Supérieure des Mines de Nantes, Nantes, France, in 2012, and the Ph.D. degree in computer science from the University of Rennes 1, Rennes, France, in 2016. His Ph.D. studies were conducted in conjunction between the Institut National de Recherche en Informatique et en Automatique (INRIA) and Technicolor in Rennes, France, and addressed the compression of High-Dynamic Range video content. After pursuing post-doctoral research at INRIA on light field image processing, he is now a Post-Doctoral Researcher at Trinity College Dublin. His research interests include High Dynamic Range as well as Light Field imaging and covers the full processing chain from capture to compression, including editing tasks.

Aljosa Smolic is the SFI Research Professor of Creative Technologies at Trinity College Dublin (TCD). Before joining TCD, Prof. Smolic was with Disney Research Zurich as Senior Research Scientist and Head of the Advanced Video Technology group, and with the Fraunhofer Heinrich-Hertz-Institut (HHI), Berlin, also heading a research group as Scientific Project Manager. At Disney Research he led over 50 R&D projects in the area of visual computing that have resulted in numerous publications and patents, as well as technology transfers to a range of Disney business units. Prof. Smolic served as Associate Editor of the IEEE Transactions on Image Processing and the Signal Processing: Image Communication journal. He was Guest Editor for the Proceedings of the IEEE, IEEE Transactions on CSVT, IEEE Signal Processing Magazine, and other scientific journals. His research group at TCD, V-SENSE, is on visual computing, combining computer vision, computer graphics and media technology, to extend the dimensions of visual sensation, with specific focus on immersive technologies such as AR, VR, volumetric video, 360/omni-directional video, light-fields, and VFX/animation, with a special focus on deep learning in visual computing.