Compressively Sampled Light Field Reconstruction Using Orthogonal Frequency Selection and Refinement

Fatma Hawary, Guillaume Boisson, Christine Guillemot, Philippe Guillotel

Compressively Sampled Light Field Reconstruction Using Orthogonal Frequency Selection and Refinement

Fatma Hawary, Guillaume Boisson, Christine Guillemot and Philippe Guillotel

***Abstract*—This paper considers the compressive sensing framework as a way of overcoming the spatio-angular trade-off inherent to light field acquisition devices. We present a novel method to reconstruct a full 4D light field from a sparse set of data samples or measurements. The approach relies on the assumption that sparse models in the 4D Fourier domain can efficiently represent light fields. The proposed algorithm reconstructs light fields by selecting the frequencies of the Fourier basis functions that best approximate the available samples in 4D hyper-blocks. The performance of the reconstruction algorithm is further improved by enforcing orthogonality of the approximation residue at each iteration, *i.e.* for each selected basis function. Since sparsity is better preserved in the continuous Fourier domain, we propose to refine the selected frequencies by searching for neighboring non-integer frequency values. Experiments show that the proposed algorithm yields performance improvements of more than 1dB compared to state-of-the-art compressive light field reconstruction methods. The frequency refinement step also significantly enhances the visual quality of reconstruction results of our method by a 1.8dB average.**

***Index Terms*—Light fields, computational photography, sparse reconstruction, compressive sensing, Fourier transform, continuous spectrum.**

I. Introduction

LIGHT field imaging has acquired a significant interest over the last two decades, both in research and industry. Providing a rich representation of the captured scene, light fields bring a variety of novel post-capture applications and enable immersive experiences. Several camera setups have been proposed for light field acquisition. Plenoptic cameras use an array of micro-lenses placed in front of the photosensor that provides an additional angular information at the expense of a decreased spatial resolution [1, 2]. Plenoptic consumer cameras, such as Lytro's, have become widely available, then SLR cameras and smartphones started to feature dual-pixel or quad-pixel sensors. Such light fields do not provide much angular variation because of their limited disparities, but paved the way for more ambitious applications. On the other hand, light fields can also be captured by arrays of cameras [3] usually dedicated to professional applications. Camera arrays offer both higher spatial resolution and wider parallax than plenoptic cameras. This enables accurate depth estimation, which is required for virtual reality applications and cinema production. Note that each new generation of smartphones is equipped with an increasing number of cameras, which foreshadows light field imaging in the future for mobile devices [4]. Another approach for capturing light fields consists in moving a single camera (*e.g.* the Stanford Gantry[1]), in front of a static scene. That kind of setup excludes the acquisition of video light fields, but provides data with high spatial and angular resolutions. Besides, at the research level, some alternative solutions have been proposed for light field acquisition with improved resolution [5–7], or for a more flexible acquisition [8, 9].

While the industry calls for increasing image resolutions, the task of acquiring high-quality 4D light field content remains challenging, due to the complexity and size of optics, photosensors and, ultimately, because of the bottleneck of data storage. Indeed, capturing light field videos requires sophisticated systems and engineering proficiency to operate the incoming data stream. When distributed storage is not an option, *e.g.* for real-time pre-visualization, light field video acquisition setups have no choice but sacrificing the resolution in one or several dimensions: spatial, angular, or temporal. Let us take the example of a grid of 16 4K-cameras acquiring videos at a frame-rate of 30 fps. This corresponds to 4 Gigabytes per second of data to write on the disk, which exceeds the actual SSD (Solid-State Drive) throughput capacities.

In this context, we consider a light field acquisition system that would store only a few measurements of the captured light field content. More precisely, the acquisition system consists of a sensor grid (a grid of cameras), and the captured data is transferred to a computer where a software program extracts data samples to be stored, in order to meet the disk throughput requirements. Each view of the light field is sampled following random sampling patterns which are independent from one view to another. By the sampling operation we retain and store only a sparse set of light field pixels. The sampling mask is assumed to be known by the reconstruction algorithm. Our experimental results show that the proposed random spatial sampling yields better performances compared with an angular sampling which would consist in keeping only a subset of entire views, as e.g. in [10]. In the validation tests, we assume that the light field images have been demosaicked, to make the comparison to state-of-the-art methods possible. Note that the sampling and reconstruction are applied independently within the three color channels (Red, Green and Blue).

This paper describes a light field reconstruction method from a sparse set of randomly selected samples, with no prior sampling pattern specifications. The method, named Orthogonal Frequency Selection (OFS), performs the reconstruction per 4D hyper-block, where a model is iteratively generated

This work has been in part supported by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM).

[1] http://lightfield.stanford.edu/lfs.html

to best fit the available data in the hyper-block and its 4D surroundings. The proposed method extends the 2D image Frequency Selective Reconstruction (FSR) approach described in [11] to 4D light fields while further improving it by introducing an orthogonality constraint on the residue. As in our previous work [12], the proposed OFS method exploits the assumption that the light field data is sparse in the Fourier domain [13], meaning that the Fourier transform of the light field can be expressed as a linear combination of a small number of Fourier basis functions. The reconstruction algorithm therefore searches for these bases (*i.e.* their frequencies) which best represent the 4D Fourier spectrum of the sampled light field.

Besides, since sparsity is better verified in the continuous Fourier domain [14], the method is further extended to reconstruct the light field at non-integer angular frequency positions. This method, called OFS+refinement, allows us to better approximate the Fourier spectrum of the signal, and to overcome the problem of having a small number of samples in the angular direction. Furthermore, analytical forms can be derived in the Fourier domain, and the expansion coefficient computation can be directly done using the Fourier transforms of the signal. Our solutions can be applied to any captured light field, independently from the acquisition system. They are also free of any prior knowledge of scene geometry and do not require any pre-processing step such as depth estimation.

Experimental results with several light fields show that the proposed OFS and OFS+refinement methods yield a high reconstruction quality from a small number of input samples (*e.g.* at sampling rates going down to 4%). Results with synthetic and real light fields show that it outperforms reference methods that either similarly pose the problem in a compressive sensing framework as [15] where the authors use a union of trained dictionaries, or exploit sparsity in the 4D Fourier domain as [14]. Given that storing a subset of views rather than randomly sampling all the views could also be considered to address the data rate issue, we also compare the proposed OFS algorithm with the view synthesis method of [10] which uses deep neural networks.

This paper is organized as follows. Section II gives an overview of the related work. Section III gives a detailed description of the proposed reconstruction method for 4D light fields. The non-integer frequency refinement is also described in this section. The experimental results with the proposed algorithms are given in Section IV. Section V concludes the whole paper.

II. Related work

In this section, we present an overview of the topics related to our work. We first give a brief review of state-of-the-art light field imaging solutions. Then, we present existing compressive sensing systems for higher-resolution light field acquisition. We finally introduce sparse light field reconstruction methods in the Fourier domain.

A. Coded aperture imaging

Numerous efforts have already been presented to improve the spatio-angular resolution trade-off of acquired light fields. Liang *et al.* [16] proposed a programmable aperture approach which exploits the fast multiple-exposure feature of digital sensors to sequentially capture multiple subsets of light rays, but at the cost of a longer exposure time. In [17], two attenuation masks are used, one placed at the aperture and the other one in front of the 2D photosensor. Zhang *et al.* [18] presented a phase-based approach to reconstruct a 4D light field using a micro-baseline stereo pair. Yagi *et al.* [19] proposed an aperture pattern and reconstruction algorithm derived from principal component analysis (PCA). In this method, the compressive pattern chosen is dependent of the original signal, which presents a limitation since it may not be known. Another approach was proposed later in [20] where the basis vectors are derived from non-negative matrix factorization (NMF). A new deep learning-based reconstruction method is later proposed using the same coded aperture acquisition process [21].

B. Light field view synthesis

Some recent methods aimed to synthesize a full light field from a subset of views. Wanner *et al.* [22] introduced a variational light field angular super-resolution framework by utilizing the estimated depth map from the input views to warp them to novel views. Several works also proposed light field synthesis using deep learning and Convolutional Neural Networks (CNNs). In [10], a full light field is generated from its four corner views, using two sequential CNNs, one for disparity estimation and then another for color prediction on the top of warped images. With a similar pipeline, Srinivasan *et al.* [23] addressed the problem of extrapolating a light field from a single central view, with good visual results. Note that both works only considered light fields captured with plenoptic cameras that present limited parallax. Larger baselines make view synthesis more complex, with wider occlusions and more non-Lambertian effects such as specularities. Wu *et al.* [24] address the problem of light field angular super-resolution by first spatially down-sampling the views in order to balance spatial and angular information. The views are then interpolated using a bi-cubic filter, and extracted Epipolar Plane Images (EPI) are then fed to a CNN architecture to restore angular details on EPI. A non-blind deblur step is then applied to recover spatial details. The framework was then extended into several applications [25] including depth enhancement and interpolation for unstructured input.

C. Compressive Sensing for light fields

Reconstructing a light field from a set of measurements is a compressive sensing (CS) problem that is solved assuming sparsity priors on the captured data. The CS theory [26] relies on the assumption that the signal is sparse (or compressible) in some transform domains like wavelets, Discrete Cosine Transform (DCT), or even dictionaries learned from large datasets. The original data is restored by solving a basis pursuit denoising (BPDN) problem [27] given an over-complete dictionary as in [28, 29].

Let $x \in \mathbb{R}^N$ be a signal (here a light field) assumed to be sparse in the transform domain defined by the basis Ψ. We say that x is k-sparse in the transform domain defined by Ψ

if one can find a sparse vector α such that $x = \Psi\alpha$, where $||\alpha||_0 \leq k$.

The acquisition of the vector $y \in \mathbb{R}^M$ of measurements on the input signal can be written as $y = \Phi x \in \mathbb{R}^M$, where M is the number of randomly chosen vectors $\phi \in \mathbb{R}^N$ and $\Phi = [\phi_1, \phi_2, ..., \phi_M]^T$ is named the measurement matrix.

Reconstructing the sparse signal x from the measurements y involves solving the following optimization problem:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} ||y - \Phi\Psi\alpha|| \ + \ \lambda \cdot ||\alpha||_1$$

and then computing the estimate $\hat{x} = \Psi\hat{\alpha}$.

Several light field compressive acquisition schemes have been recently proposed based on the use of coded masks, in order to overcome the spatio-angular resolution trade-off of plenoptic cameras. In all the proposed schemes, the sensing matrix is assumed to be materialized by a coded physical mask. The authors in [30] proposed two architectures which compress sampled data on angular or spatial dimension respectively using coded masks. One mask is placed at the aperture for compressive acquisition in the angular dimension while the second mask is placed in front of a lenslet array for compressive acquisition in the spatial dimension. Wang *et al.* [31] propose a camera architecture using a random coded mask placed at the aperture and a random convolution CMOS optical sensor that can compress on the image plane without additional optical elements. Compared to architectures using two coded masks, the proposed solution permits a better light efficiency, while combining double measurements.

A monochrome coded mask is also used in [32] to capture random linear combinations of angular samples, and the light field is rebuilt via a hierarchical Bayesian framework. The authors in [28] propose a camera architecture that records optically coded projections on a single image sensor, while the authors in [15] and [29] use respectively a random binary mask or a moving colour coded mask affixed to the sensor to extract incoherent measurements. In both cases, the light field is then reconstructed using a compressive sensing framework, assuming that the light field is sparse in a domain defined by an overcomplete dictionary [28, 29] or a set of 2D separable dictionaries [15].

The problem of light field reconstruction from a sparse set of measurements can also be efficiently solved using deep learning architectures. The authors in [33–35] assume a pre-defined mask pattern and propose convolutional neural network architectures to reconstruct the light field from the measurements given the coded mask. The authors in [34] trained a two-branch neural network to reconstruct compressed light field, in which one branch is a fully connected network which limits the patch size. From the compressed image, the authors in [33] extract the central view and a disparity map, which are used to reconstruct the final light field by warping the central view. For that purpose, three different CNNs are used. Note that the network architectures in both [33–35] are adjusted for only one mask pattern which is assumed to be monochrome in [33] and [34], and colored in [35]. Therefore, they are not invariant to different locations on the sensor as each patch on the sensor is generated by a different compressing matrix. The authors in [21] pose the coded aperture acquisition and light field reconstruction as an auto-encoder and optimize the mask pattern together with the parameters of the reconstruction algorithm in the end-to-end learning of the auto-encoder. This approach therefore assumes a learned and hence pre-defined mask specification.

D. Sparse Fourier reconstruction

Different reconstruction methods have been proposed based on the assumption that light fields are sparse in the 4D Fourier domain [14,36]. In [36], the authors propose a linear approach to synthesize novel views from a focus stack of images. Assuming that the light field data resides on a 3D manifold in the 4D Fourier domain, the method retrieves the light field views by deconvolution of the focal stack images. The method in [14] based on a Sparse Fast Fourier Transform (SFFT) [37,38] exploits sparsity in the angular dimensions of the 4D Fourier domain to recover the light field from a subset of views. The reconstruction algorithm searches for the frequency values of the Fourier basis functions and the corresponding coefficients. An optimization step then approximates the spectrum in the continuous Fourier domain by refining the initial frequency positions with a small non integer step. Note that the input views are selected according to a pre-defined pattern (1D viewpoint trajectories) that ensures a good initialization of the selection of the appropriate Fourier basis functions in the sparse approximation. However, all the above methods assume specific sampling patterns for the positions of the input views, which may limit their use in practical applications. Instead, Vagharshakyan *et al.* [39,40] pose the angular super-resolution problem as an inpainting problem on the Epipolar Plane Images (EPI), and use a discrete shearlet transform to increase the light field angular resolution.

III. Light Field Reconstruction Method

A. Problem statement

Let $\mathcal{L}(x, y, u, v)$ denote a given 4D light field, that we assume to be sparse in the 4D Fourier domain. Hence, the light field $\mathcal{L}$ can be represented by a sparse vector α as

$$\mathcal{L} = \Psi\alpha, \tag{1}$$

where Ψ is the matrix containing the Fourier basis functions and α being the sparse vector of expansion coefficients. Let $\mathcal{L}_s(x, y, u, v)$ be the 4D randomly-sampled version of $\mathcal{L}$, obtained as

$$\mathcal{L}_s = \Phi\mathcal{L} = \Phi\Psi\alpha \tag{2}$$

where Φ represents the sub-sampling matrix containing 0 and 1 values to define the available data sample positions. $\mathcal{L}_s(x, y, u, v)$ constitutes the set of measurements that will be recorded on the disk, and from which we search to reconstruct the light field $\mathcal{L}$.

The goal of the reconstruction algorithm is to search for the sparse vector $\hat{\alpha}$ that will allow us to reconstruct a light field $\hat{\mathcal{L}}_r$ as

$$\hat{\mathcal{L}}_r = \Psi\hat{\alpha} \tag{3}$$

TABLE I: Notations

$\mathbf{j}$	The imaginary unit: $\mathbf{j}^2 = -1$
$z^* = \Re(z) - \mathbf{j}.\Im(z)$	Complex conjugate of z
i	Iteration index
$\Omega = [\![1;K]\!] \times [\![1;L]\!] \times [\![1;M]\!] \times [\![1;N]\!]$	Local light field domain
$P = \lvert\Omega\rvert = K \cdot L \cdot M \cdot N$	Number of samples in Ω
$A \subset \Omega$	Subset of known samples
$B \subset \Omega$	Subset of unknown samples
$C \subset \Omega$	Subset of reconstructed samples
$\mathbf{p} = (k,l,m,n) \in \Omega$	A pixel's position within Ω
$f : \Omega \to \mathbb{R}$	Local light field
$g^{(i)} : \Omega \to \mathbb{R}$	Approximation model at iteration i
$w : \Omega \to \mathbb{R}$	Weighting function
$r^{(i)} : \Omega \to \mathbb{R}$	Weighted residue at iteration i
$\boldsymbol{\vartheta} = (\mu,\nu,\zeta,\xi)$	A frequency in the 4D spectrum
$\varphi_{\boldsymbol{\vartheta}} : \Omega \to \mathbb{R}$	A 4D Fourier basis function
$\Theta^{(i)} = \{\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_i\}$	Frequency subset at iteration i
$X_{\boldsymbol{\vartheta}} = \sum_{\mathbf{p}} x[\mathbf{p}]\varphi^*_{\boldsymbol{\vartheta}}[\mathbf{p}] \in \mathbb{C}$	Capitals denote Fourier transforms

that will be as close as possible to the captured light field $\mathcal{L}(x,y,u,v)$. The coefficients in $\hat{\alpha}$ correspond to the contributions of the frequencies that our reconstruction method iteratively selects. Figure 1 gives an overview of the reconstruction objective. Note that the sampled views look dark since all the missing pixels are set to 0.

B. Sparse model for light field reconstruction

We conduct the reconstruction per 4D hyper-block of the light field. An example of 4D hyper-block is shown in Fig. 2 (set of 2D blocks surrounded in red). Together with its local surrounding area (of a given width in the angular and spatial dimensions), the red hyper-block forms a 4D domain Ω spanning over $M \times N$ views with co-located $K \times L$ patches. We define $f[k,l,m,n]$ as the signal spanning in the 4D domain Ω, (k,l) and (m,n) denoting the spatial and angular coordinates of the pixels within the hyper-block. We aim to find an approximation model $g[k,l,m,n]$ from both original and reconstructed samples within the domain Ω.

Let A, B and C respectively denote the subsets of known, unknown, and previously reconstructed samples in the 4D hyper-block and in its surrounding. At each iteration of the proposed method, the set of input measurements is therefore $\Omega = A \cup B \cup C$. As neighboring hyper-blocks may already have been processed before, the corresponding reconstructed samples are contained in the area C to be used in the reconstruction of the currently considered hyper-block. For convenience, we summarize the notations used throughout this paper in TABLE I.

A weighting function w is defined as

$$w[\mathbf{p}] = w[k,l,m,n] = \begin{cases} \rho_s^{\sqrt{\overline{k}^2+\overline{l}^2}} \rho_a^{\sqrt{\overline{m}^2+\overline{n}^2}} & \text{for } \mathbf{p} \in A \\ 0 & \text{for } \mathbf{p} \in B \\ \sigma.\rho_s^{\sqrt{\overline{k}^2+\overline{l}^2}} \rho_a^{\sqrt{\overline{m}^2+\overline{n}^2}} & \text{for } \mathbf{p} \in C \end{cases} \quad (4)$$

where $\overline{k} = k - \frac{K+1}{2}$, $\overline{l} = l - \frac{L+1}{2}$, $\overline{m} = m - \frac{M+1}{2}$, $\overline{n} = n - \frac{N+1}{2}$, and $0 < \rho_s, \rho_a, \sigma < 1$. The goal of the function w is to give more or less weight to the input measurements, depending on their position in Ω and on whether they correspond to original or previously reconstructed samples. More precisely, the factors ρ_s and ρ_a respectively determine the weighting decay in the spatial dimension (within a view) and in the angular (cross-view) dimension. In other words, the weight of a sample decreases as its distance to the hyper-block center increases. Finally, the parameter σ is used to differentiate the contribution of previously reconstructed data from the available data.

Let f be the local light field defined on Ω. Assuming the light field to be sparse in the 4D Fourier domain, the algorithm estimates a sparse approximation model g of the signal f

$$g[\mathbf{p}] = \sum_{\boldsymbol{\vartheta} \in \Theta} c_{\boldsymbol{\vartheta}} \cdot \varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] \quad (5)$$

as a weighted combination of 4D Fourier basis functions

$$\varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] = \varphi_{\mu\nu\zeta\xi}[k,l,m,n] = e^{2\pi\mathbf{j}\left(\frac{k\mu}{\mathbf{K}} + \frac{l\nu}{\mathbf{L}} + \frac{m\zeta}{\mathbf{M}} + \frac{n\xi}{\mathbf{N}}\right)}, \quad (6)$$

with $c_{\boldsymbol{\vartheta}}$ being the contribution of the Fourier basis function $\varphi_{\boldsymbol{\vartheta}}$. For simplicity, we will use in the following the symbol $\boldsymbol{\vartheta}$ to refer to a 4D frequency position.

C. Iterative Reconstruction by Orthogonal Frequency Selection (OFS)

The different steps of the method are depicted in Fig. 3. The algorithm iteratively selects the Fourier basis functions that would best represent (minimizing an approximation error) the available samples within the hyper-block. Since the reconstruction of a hyper-block greatly depends on the amount of information available in its neighborhood, the hyper-block processing order is defined according to the number of known samples in Ω.

In practice, we assign to each hyper-block a surrounding density measure that is updated (increased) each time one of its neighboring hyper-blocks is reconstructed. The surrounding density measure $sd(b)$ of a hyper-block b takes into account the densities d of its neighbors $\{b_x \in \mathcal{N}(b)\}$ and is computed as

$$sd(b) = d(b) \; + \sum_{b_x \in \mathcal{N}(b)} d(b_x) \quad (7)$$

This ensures a higher approximation quality, since we ensure a higher density of known samples at each hyper-block reconstruction.

1) Frequency selection: Let $r^{(i)}$ be the weighted residue of the approximation model with respect to the signal at iteration i:

$$r^{(i)} = (f - g^{(i)}) \cdot w \quad (8)$$

At each iteration i, the algorithm selects the basis function $\boldsymbol{\vartheta}_i$ maximizing the projection of the residue $r^{(i-1)}$, *i.e.*

$$\begin{aligned} \boldsymbol{\vartheta}_i &= \operatorname*{argmax}_{\boldsymbol{\vartheta}} \left| \langle \varphi_{\boldsymbol{\vartheta}}, r^{(i-1)} \rangle \right| \\ &= \operatorname*{argmax}_{\boldsymbol{\vartheta}} \left| \sum_{\mathbf{p} \in \Omega} \left(r^{(i-1)}[\mathbf{p}] \cdot \varphi^*_{\boldsymbol{\vartheta}}[\mathbf{p}] \right) \right| \end{aligned} \quad (9)$$

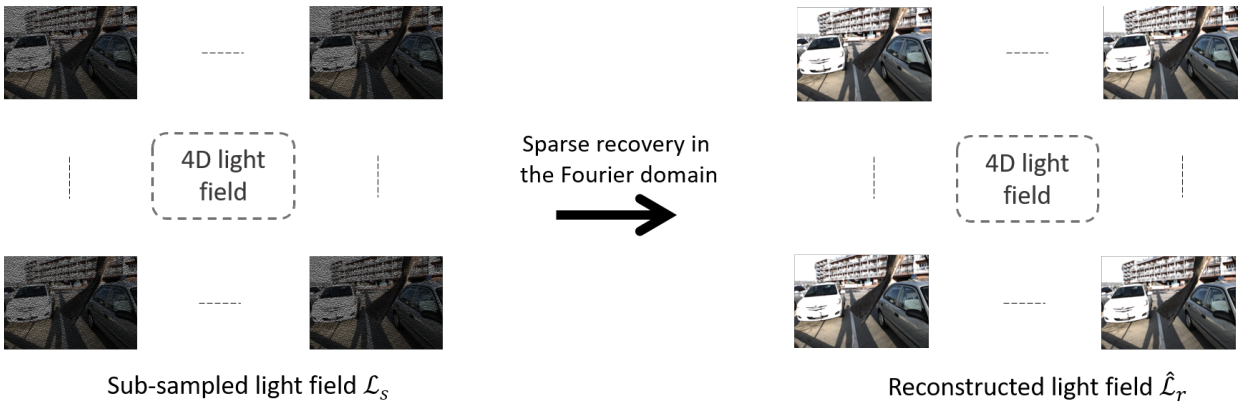


Fig. 1: Overview of the reconstruction method input and objective.

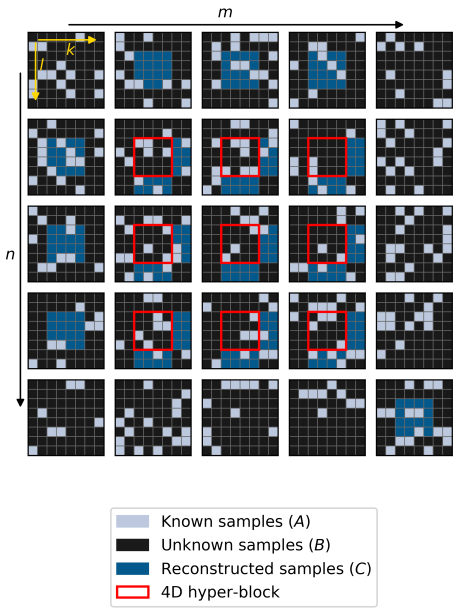


Fig. 2: 4D hyper-block (outlined in red) and its local spatio-angular neighborhood ($K = L = 8$ and $M = N = 5$): the angularly neighboring hyper-blocks located at top+left of the hyper-block as well as the spatially-neighboring hyper-blocks located at right+bottom are previously reconstructed.

2) Update with no orthogonality constraint: The corresponding expansion coefficient $c^{(i)}_{\boldsymbol{\vartheta}_i}$ is computed by minimizing the weighted residual energy $E^{(i)}_w$

$$E^{(i)}_w = \sum_{\mathbf{p}\in\Omega} \left| f[\mathbf{p}] - g^{(i-1)}[\mathbf{p}] - c^{(i)}_{\boldsymbol{\vartheta}_i} \cdot \varphi_{\boldsymbol{\vartheta}_i}[\mathbf{p}] \right|^2 \cdot w[\mathbf{p}]. \quad (10)$$

Looking for zeros of the derivative of $E^{(i)}_w$ with respect to $c^{(i)}_{\boldsymbol{\vartheta}_i}$ yields

$$c^{(i)}_{\boldsymbol{\vartheta}_i} \cdot \sum_{\mathbf{p}\in\Omega} \left(w \cdot \varphi_{\boldsymbol{\vartheta}_i} \cdot \varphi^*_{\boldsymbol{\vartheta}_i}\right)[\mathbf{p}] = \sum_{\mathbf{p}\in\Omega} \left(r^{(i-1)} \cdot \varphi^*_{\boldsymbol{\vartheta}_i}\right)[\mathbf{p}], \quad (11)$$

and therefore:

$$c^{(i)}_{\boldsymbol{\vartheta}_i} = \frac{\sum_{\mathbf{p}\in\Omega} r^{(i-1)}[\mathbf{p}] \cdot \varphi^*_{\boldsymbol{\vartheta}_i}[\mathbf{p}]}{\sum_{\mathbf{p}\in\Omega} w[\mathbf{p}]}. \quad (12)$$

At that point, a straightforward update step of both the approximation model and the weighted residue would consist in:

$$\begin{cases} g^{(i)} = g^{(i-1)} + c^{(i)}_{\boldsymbol{\vartheta}_i} \cdot \varphi_{\boldsymbol{\vartheta}_i}, \\ r^{(i)} = r^{(i-1)} - c^{(i)}_{\boldsymbol{\vartheta}_i} \cdot \varphi_{\boldsymbol{\vartheta}_i} \cdot w. \end{cases} \quad (13)$$

However, while the Fourier basis functions are orthogonal on Ω, a weighted basis function may not be orthogonal to another basis function:

$$\forall\ \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}'\ ,\ \langle w \cdot \varphi_{\boldsymbol{\vartheta}}, \varphi_{\boldsymbol{\vartheta}'} \rangle \neq 0 \quad (14)$$

Therefore, if the update step (13) cancels the energy of $r^{(i-1)}$ along $\boldsymbol{\vartheta}_i$ in $r^{(i)}$, it also alters its spectrum for every $\boldsymbol{\vartheta}_{j<i}$, *i.e.*

$$\begin{cases} \langle r^{(i)}, \varphi_{\boldsymbol{\vartheta}_i} \rangle = 0 \\ \forall\ j < i\ :\ \boldsymbol{\vartheta}_j \neq \boldsymbol{\vartheta}_i\ ,\ \langle r^{(i)}, \varphi_{\boldsymbol{\vartheta}_j} \rangle \neq 0 \end{cases} \quad (15)$$

In other words, the weighted residue is not orthogonal to the subspace spanned by the already selected functions. As a consequence, a basis function can possibly be selected several times. To attenuate this orthogonality deficiency, the FSR method, proposed in Seiler *et al.* [11] for 2D image reconstruction, rely on a compensation factor between 0 and 1 in the computation of $c_{\boldsymbol{\vartheta}_i}$. Still, their approach requires many iterations to reach a fair approximation quality, for example, 100 iterations for 4×4 blocks and 32×32 neighborhood [11]. Moreover, the value of the compensation factor is derived from a trained parameter. As such, a suitable dataset needs to be selected for training, which may limit the method performances when applied on other types of data.

3) Orthogonality and Hermitian symmetry: To enforce orthogonality, at each iteration i, we compute an update $\Delta c^{(i)}_{\boldsymbol{\vartheta}}$ of the coefficient $c^{(i)}_{\boldsymbol{\vartheta}}$ associated to all the basis functions selected so far. More precisely, the model and the approximation residue at each iteration i are computed as

$$\begin{cases} g^{(i)} = g^{(i-1)} + \sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}} \Delta c^{(i)}_{\boldsymbol{\vartheta}} \cdot \varphi_{\boldsymbol{\vartheta}}, \\ r^{(i)} = r^{(i-1)} - \sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}} \Delta c^{(i)}_{\boldsymbol{\vartheta}} \cdot \varphi_{\boldsymbol{\vartheta}} \cdot w. \end{cases} \quad (16)$$

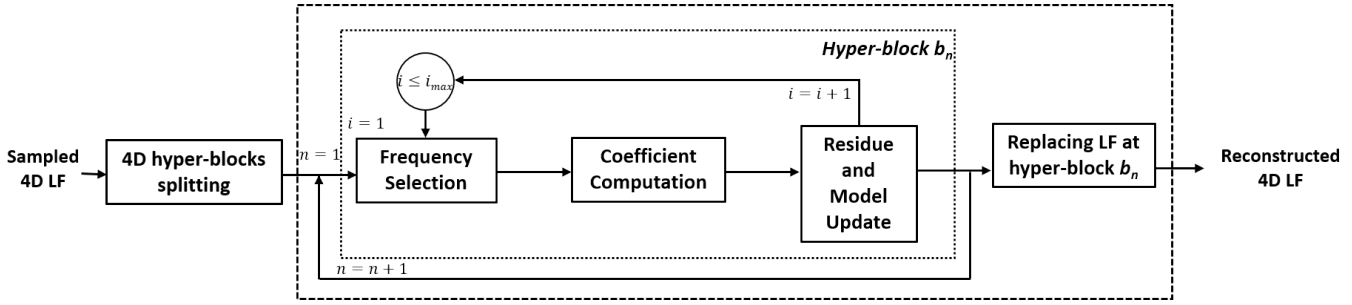


Fig. 3: Representative scheme of our proposed compressive light field reconstruction method.

The basis functions are selected as before (in (9)). However, we now search for updates $\Delta c_{\boldsymbol{\vartheta}}$ of coefficients that will minimize

$$E_w^{(i)} = \sum_{\mathbf{p}\in\Omega} \left| f[\mathbf{p}] - g^{(i-1)}[\mathbf{p}] - \sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}} \Delta c_{\boldsymbol{\vartheta}}^{(i)} \cdot \varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] \right|^2 \cdot w[\mathbf{p}] \tag{17}$$

Also, as real-valued signals are Hermitian, their Fourier spectra exhibit complex conjugate symmetries. So, to ensure that the approximation g yields a real-valued signal, we modify (5) into:

$$g[\mathbf{p}] = \frac{1}{2} \sum_{\boldsymbol{\vartheta}\in\Theta} \left(c_{\boldsymbol{\vartheta}} \cdot \varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] + c_{\boldsymbol{\vartheta}}^* \cdot \varphi_{\boldsymbol{\vartheta}}^*[\mathbf{p}] \right) \tag{18}$$

This means adding to the model the conjugate complex of a frequency each time it is selected, as

$$E_w^{(i)} = \sum_{\mathbf{p}\in\Omega} \left| \left(f - g^{(i-1)} - \frac{1}{2} \sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}} \left(\Delta c_{\boldsymbol{\vartheta}}^{(i)} \cdot \varphi_{\boldsymbol{\vartheta}} + \Delta c_{\boldsymbol{\vartheta}}^{(i)*} \cdot \varphi_{\boldsymbol{\vartheta}}^* \right) \right)[\mathbf{p}] \right|^2 \cdot w[\mathbf{p}] \tag{19}$$

The new minimization criterion with respect to each $\Delta c_{\boldsymbol{\vartheta}}^{(i)}$ yields a system of equations whose solution ensures that the residue is orthogonal to each basis function selected so far:

$$\begin{aligned} &\sum_{\mathbf{p}\in\Omega} \left(w[\mathbf{p}] \varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] \sum_{\boldsymbol{\vartheta'}\in\Theta^{(i)}} \frac{1}{2} \left[\Delta c_{\boldsymbol{\vartheta'}}^{(i)} \varphi_{\boldsymbol{\vartheta'}}[\mathbf{p}] + \Delta c_{\boldsymbol{\vartheta'}}^{(i)*} \varphi_{\boldsymbol{\vartheta'}}^*[\mathbf{p}] \right] \right) \\ &= \sum_{\mathbf{p}\in\Omega} r^{(i-1)}[\mathbf{p}] \varphi_{\boldsymbol{\vartheta}}[\mathbf{p}], \quad \forall \boldsymbol{\vartheta} \in \Theta^{(i)} \end{aligned} \tag{20}$$

4) Approximation model and residue update: Once a new basis function is selected, an update step introduces the contribution of the selected basis function to the approximation model, and updates the contributions of the already selected functions, by computing the corresponding coefficients as:

$$g^{(i)} = g^{(i-1)} + \frac{1}{2} \sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}} \left(\Delta c_{\boldsymbol{\vartheta}}^{(i)} \cdot \varphi_{\boldsymbol{\vartheta}} + \Delta c_{\boldsymbol{\vartheta}}^{(i)*} \cdot \varphi_{\boldsymbol{\vartheta}}^* \right) \tag{21}$$

The residue is similarly updated as

$$r^{(i)} = r^{(i-1)} - \frac{1}{2} \sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}} \left(\Delta c_{\boldsymbol{\vartheta}}^{(i)} \cdot \varphi_{\boldsymbol{\vartheta}} + \Delta c_{\boldsymbol{\vartheta}}^{(i)*} \cdot \varphi_{\boldsymbol{\vartheta}}^* \right) \cdot w \tag{22}$$

The algorithm then proceeds until a predefined number of iterations is reached. The orthogonality property leads to a more accurate reconstruction with a reduced number of iterations.

D. Analytical solution in the Fourier domain

So far, the reconstruction method has been explained in the spatial domain. Using DFT functions as basis functions allows us to express all the equations in the frequency domain, leading to a more efficient implementations. Only one local Discrete Fourier Transform (DFT) at the beginning and one inverse transform at the end of the algorithm are necessary, all intermediate steps being expressed in the Fourier domain.

The frequency selection step (9) can be expressed in the Fourier domain as

$$\boldsymbol{\vartheta}_i = \underset{\boldsymbol{\vartheta}}{\operatorname{argmax}} \left| R_{\boldsymbol{\vartheta}}^{(i-1)} \right|, \tag{23}$$

where R denotes the Fourier transform of the weighted residue r. For the expansion coefficients computation, the system of equations (20) becomes, for every $\boldsymbol{\vartheta} \in \Theta^{(i)}$,

$$\sum_{\boldsymbol{\vartheta'}\in\Theta^{(i)}} \frac{1}{2} \left(\Delta c_{\boldsymbol{\vartheta'}} \cdot W_{\boldsymbol{\vartheta'}+\boldsymbol{\vartheta}}^* + \Delta c_{\boldsymbol{\vartheta'}}^* \cdot W_{\boldsymbol{\vartheta'}-\boldsymbol{\vartheta}} \right) = R_{\boldsymbol{\vartheta}}^{(i-1)^*}, \tag{24}$$

where W denotes the Fourier transform of the weighted function w.

Estimating the updates of all the expansion coefficients $\{\Delta c_{\boldsymbol{\vartheta}_j}^{(i)}\}_{j=1..i}$, amounts to solving the following equation:

$$\boldsymbol{\Delta}\mathbf{c}^{(i)} = 2 \, \mathbf{W}^{(i)^{-1}} \cdot \mathbf{R}^{(i-1)} \tag{25}$$

where we define the $(2i-1)$ vectors $\boldsymbol{\Delta}\mathbf{c}^{(i)}$ and $\mathbf{R}^{(i-1)}$, and the matrix $\mathbf{W}^{(i)}$ of size $(2i-1) \times (2i-1)$ as follows:

$$\boldsymbol{\Delta}\mathbf{c}^{(i)} = \begin{pmatrix} \Re(\Delta c_{\boldsymbol{\vartheta}_1}^{(i)}) \\ \vdots \\ \Re(\Delta c_{\boldsymbol{\vartheta}_i}^{(i)}) \\ \Im(\Delta c_{\boldsymbol{\vartheta}_2}^{(i)}) \\ \vdots \\ \Im(\Delta c_{\boldsymbol{\vartheta}_i}^{(i)}) \end{pmatrix} \qquad \mathbf{R}^{(i-1)} = \begin{pmatrix} \Re(R_{\boldsymbol{\vartheta}_1}^{(i-1)}) \\ \vdots \\ \Re(R_{\boldsymbol{\vartheta}_i}^{(i-1)}) \\ \Im(R_{\boldsymbol{\vartheta}_2}^{(i-1)}) \\ \vdots \\ \Im(R_{\boldsymbol{\vartheta}_i}^{(i-1)}) \end{pmatrix} \tag{26}$$

$$\mathbf{W}^{(i)} = \begin{pmatrix} \mathbf{W}_{11}^{(i)} & \mathbf{W}_{12}^{(i)} \\ \mathbf{W}_{21}^{(i)} & \mathbf{W}_{22}^{(i)} \end{pmatrix}$$

with:

$$\begin{aligned} \mathbf{W}_{11}^{(i)} &= \left[\Re(W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y} + W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y}) \right]_{(x,y)\in[\![1;i]\!]\times[\![1;i]\!]} \\ \mathbf{W}_{12}^{(i)} &= \left[\Im(W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y} - W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y}) \right]_{(x,y)\in[\![1;i]\!]\times[\![2;i]\!]} \\ \mathbf{W}_{21}^{(i)} &= \left[\Im(W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y} + W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y}) \right]_{(x,y)\in[\![2;i]\!]\times[\![1;i]\!]} \\ \mathbf{W}_{22}^{(i)} &= \left[\Re(W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y} - W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y}) \right]_{(x,y)\in[\![2;i]\!]\times[\![2;i]\!]} \end{aligned}$$

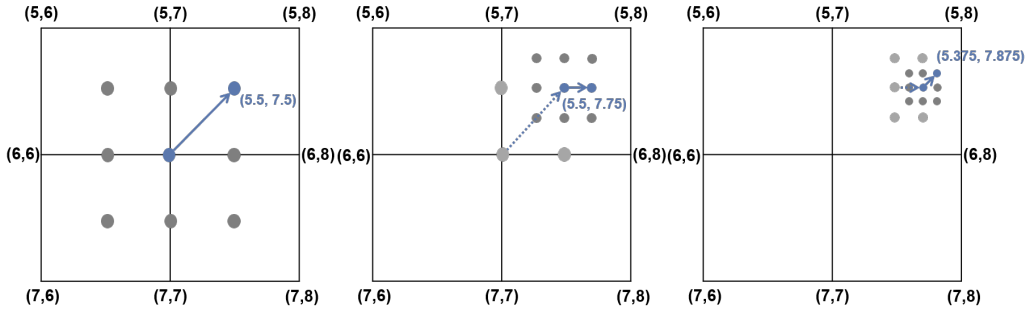


Fig. 4: Refinement example. Shifting the integer frequency by a small step to one of the eight directions at each iteration (refinement level).

Consistently with these definitions, we conventionally force the first selected frequency $\boldsymbol{\vartheta}_1$ to be the zero frequency. Since the signals considered are real-valued, so are their average value on the hyper-block. The corresponding expansion coefficients have therefore no imaginary part, and the corresponding null rows/columns are removed from the matrix definitions. Hence the $(2i-1)$ size instead of expected $2i$.

Eventually, the parametric model and the residue are updated in the Fourier domain for every $\boldsymbol{\vartheta} \in \Theta^{(i)}$ as

$$\begin{cases} G_{\boldsymbol{\vartheta}}^{(i)} = G_{\boldsymbol{\vartheta}}^{(i-1)} + \frac{1}{2}\ P \cdot \Delta c_{\boldsymbol{\vartheta}}^{(i)} \\ G_{-\boldsymbol{\vartheta}}^{(i)} = G_{-\boldsymbol{\vartheta}}^{(i-1)} + \frac{1}{2}\ P \cdot \Delta c_{\boldsymbol{\vartheta}}^{*(i)} \end{cases} \tag{27}$$

where P represents the total number of samples in the considered hyper-block. The residue is then updated as

$$R_{\boldsymbol{\vartheta}}^{(i)} = R_{\boldsymbol{\vartheta}}^{(i-1)} - \sum_{\boldsymbol{\vartheta'} \in \Theta^{(i)}} \frac{1}{2}\left(\Delta c_{\boldsymbol{\vartheta'}}^{(i)} \cdot W_{\boldsymbol{\vartheta'}-\boldsymbol{\vartheta}}^{*} + \Delta c_{\boldsymbol{\vartheta'}}^{(i)*} \cdot W_{\boldsymbol{\vartheta'}+\boldsymbol{\vartheta}}\right) \tag{28}$$

E. Frequency refinement to non-integer values

So far, the OFS approximation model is generated by including the discrete Fourier functions that best represent the available data samples. However, the actual spectrum of a signal is not necessarily aligned with the discrete sampling grid. This is why window effects can be observed in the spectrum of a light field's 4D DFT. Besides, since the angular resolution of a light field (number of views or sub-aperture images) is usually lower than its spatial resolution (the number of pixels per view or per sub-aperture image), the windowing effect is stronger in the angular dimensions, as observed in [14]. We propose to refine the discrete angular frequencies to non-integer values, in order to better approximate the continuous Fourier spectrum of the light field. The refinement can just as well be performed in the spatial dimensions, but with a lesser impact since light fields usually present higher spatial than angular sampling.

Unlike the approach of Shi *et al.* [14], where the recovery of continuous frequencies is conducted in the full light field spectrum, as a global post-processing, the OFS refinement is performed locally within the hyper-block for each selected basis function. Each time a new frequency is selected via residual energy minimization, we shift its position in the 4D spectrum by a small fractional step δ to all the eight angular directions, as shown in Fig. 4. The residual energy is calculated in all the eight corresponding frequency positions. The position that maximizes the residue decrease is selected as the final frequency to include to the model. An overview of the refinement is detailed in Algorithm 1.

Algorithm 1 Angular Frequency refinement

$\Delta = \{(-1,-1),(-1,0),(-1,1),(0,-1),(0,1),(1,-1),(1,0),(1,1)\}$
$\delta = 1/2$
define n_r
Select a frequency position $\boldsymbol{\vartheta} = (\mu, \nu, \zeta, \xi)$
$R_{max} = |R_{\boldsymbol{\vartheta}}|$
while $\delta \geq \frac{1}{2^{n_r}}$ **do**
  for $(d\zeta, d\xi) \in \Delta$ **do**
    $\boldsymbol{\vartheta'} = \boldsymbol{\vartheta} + (0, 0, \delta.d\zeta, \delta.d\xi)$
    Calculate $R_{\boldsymbol{\vartheta'}}$
  $\boldsymbol{\vartheta^*} = \text{argmax}_{\boldsymbol{\vartheta'}} |R_{\boldsymbol{\vartheta'}}|$
  if $|R_{\boldsymbol{\vartheta^*}}| > R_{max}$ **then**
    $\boldsymbol{\vartheta} = \boldsymbol{\vartheta^*}$
    $R_{max} = |R_{\boldsymbol{\vartheta^*}}|$
  $\delta = \delta/2$

The reconstruction steps that follow the selection are the same as in the regular case (with integer frequencies), except that computations can no longer be performed in the DFT domain, but in the pixel domain.

There are several advantages in refining frequencies as soon as they are selected, rather than as a global hyper-block post-processing. On-the-fly refinement helps converging swiftly to the actual light field spectrum, instead of dealing with several discrete frequencies that correspond to the same spectrum peak. Thus the number of iterations is reduced, and the overall reconstruction quality is improved.

F. Differences with an OMP-based approach

Although the iterative selection of basis functions in the proposed OFS method finds some similarities with the Matching Pursuit (MP) algorithm [41], and its orthogonal variant (OMP), there exist several aspects that significantly differentiates it from the MP method.

First, the proposed OFS method defines an hyper-block processing order to take into account the number of known or previously reconstructed samples in a given neighborhood. A weighting function is in addition used to give more weight in the approximation to known versus previously reconstructed samples, as well as to samples in the considered window that are closer to the center of the hyper-block to be reconstructed.

Besides, a refinement step is introduced to better approximate the continuous Fourier spectrum of the light field. Furthermore, analytical expressions in the Fourier domain allow us to significantly reduce the computational time compared with a computation in the pixel domain.

IV. Experiments

The proposed OFS and OFS+refinement with non-integer frequencies algorithms, as well as FSR [11] extended to 4D here, are compared against three different types of method. We compare the proposed approach with:

1) the compressive sensing method based on a union of trained dictionaries of [15], with overlapping and non overlapping patches.
2) the light field reconstruction method from a set of views, exploiting sparsity of light fields in the 4D Fourier domain [14].
3) the deep learning-based view synthesis approach of [10]. This comparison is motivated by the fact that the problem of view synthesis can also be regarded as a problem of reconstruction from sparse measurements (the subset of input views). In addition, storing a subset of views instead of random samples can also be considered a solution to the high data rate problem posed by video light fields.

We consider both synthetic and real light field datasets. The synthetic light field considered is the 5×5 light field *Dragon*[2]. The real light fields used in the tests are plenoptic light fields from [10] and light fields acquired from a camera moving on a gantry from the Stanford dataset[3].

A. Parameter settings

In order to keep the computational load manageable, the transform model size is set to $32\times32\times5\times5$, which means a block size of $4\times4\times3\times3$, and spatial and angular border widths of 14 and 1 respectively. We set the decay factors ρ_s (in the spatial dimensions) and ρ_a (in the angular dimensions) of the weighting function w as 0.7 and 0.5 respectively. The weight to differentiate reconstructed versus original samples is set to $\sigma = 0.5$. Since the maximum number of frequencies a hyperblock can contain is equal to its size (here 144), and that, at each iteration, the OFS method permits to add 2 frequencies (the selected one and its conjugate), we set the maximum number of iterations to 72. To make a fair comparison, we set this parameter to 200 for the non-orthogonal version of the method which can select several times the same basis functions. Finally, for the refinement step, we decided to limit the maximum level of refinement to $n_r = 3$.

B. Results

1) Comparison with a compressive sensing method: Fig. 5 compares the PSNR obtained with the proposed FSR, OFS and OFS+refinement methods with the ones of [15], for different sampling rates. The input data for both methods is the same randomly sampled light field with various sampling rates. As shown in the graph, our OFS method gives a much higher PSNR than the method in [15], even for a very low sampling rate. The average PSNR gain is 4.5dB. The non-integer refinement adds an average PSNR gain of 1.2dB to the OFS method. As compared to the non-orthogonal method FSR, OFS achieves a PSNR gain of 0.72dB with a lower number of iterations: 72 iterations *vs.* 200 for FSR.

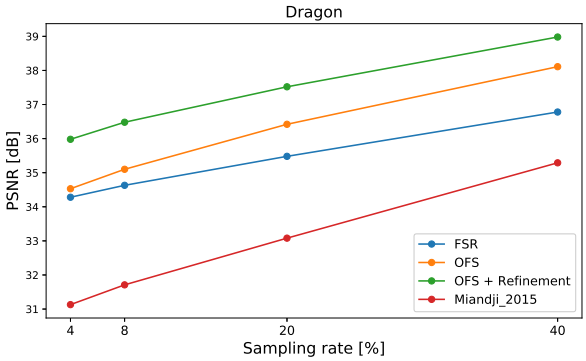


Fig. 5: PSNR comparison of 4D-FSR, 4D-OFS and 4D-OFS with refinement, with a state-of-the-art method: Miandji *et al.* [15] for different sampling rates.

Fig. 7 shows the reconstructed central views and the differences with the ground truth of *Dragon* at a sampling rate of 4%. One can see that a better visual quality is achieved with the OFS method over the dictionary-learning method of [15].

2) Comparison with a method exploiting sparsity in the Fourier domain: We now compare with the Sparse Fourier Transform (SFFT) method [14], using light fields (*Bunny*, *Crystal*, *Amethyst* and *Lego Knights*) from the Stanford Archive[4]. We reconstruct each light field from a sampled set of 81 views. The method [14] has been tested with a number of 45 input views chosen using their box-and-X pattern, which corresponds to $45/81 = 0.55$ of the full light field. We use this same ratio for our input data sampling. TABLE II gives the PSNR and SSIM values of the light fields reconstructed with both methods. The PSNR and SSIM values are averaged over the 9×9 reconstructed views. One can see that our method achieves a much better reconstruction quality with a PSNR average gain of 7.37dB.

Fig. 8 shows reconstruction examples (view $(5,2)$ of *Crystal* and *Amethyst*). The difference images show a strong unstructured noise in the SFFT method result. This noise cannot be related to the original data since we do not find it in our reconstruction result, but is more likely to be inherent to the initialization step in [14] where a rough estimation of the frequency positions is made using a voting strategy from the available views spectra. In the highly non-Lambertian scene *Amethyst*, our method succeeds in reconstructing the specular and reflective features, with very small difference to the original data, while the SFFT based method [14] fails in reconstructing regions with reflections and refraction.

Our code was designed for flexible experimentation and is currently not optimized. Using an 8-core machine, typical

[2]http://web.media.mit.edu/~gordonw/SyntheticLightFields
[3]http://lightfield.stanford.edu/lfs.html

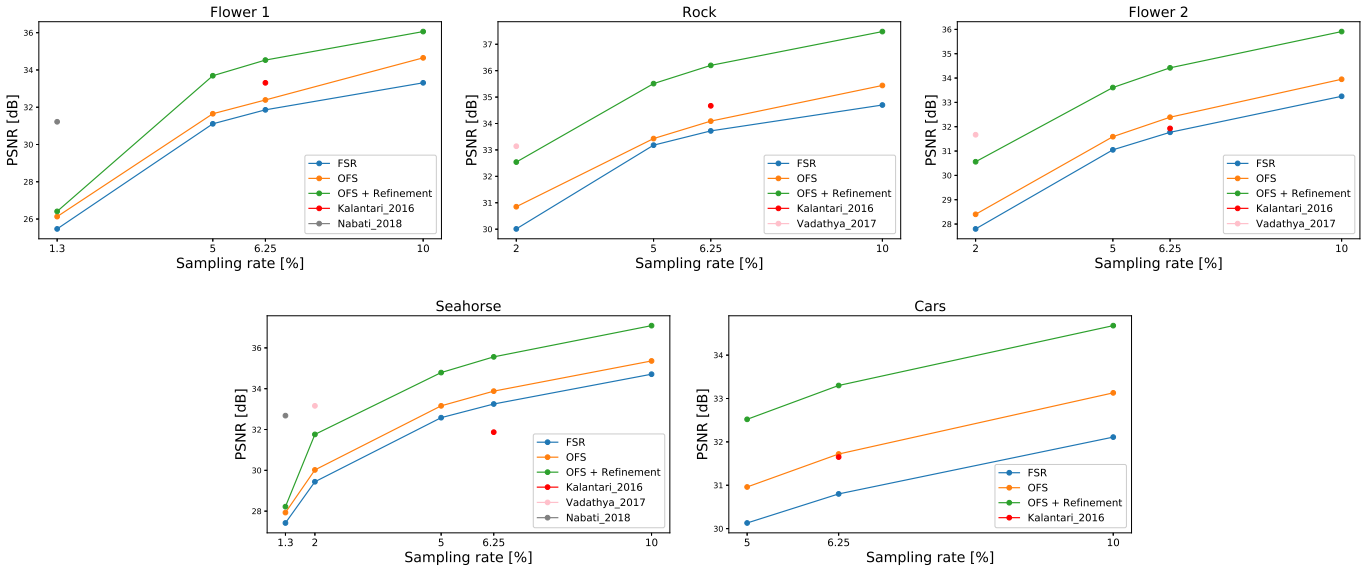
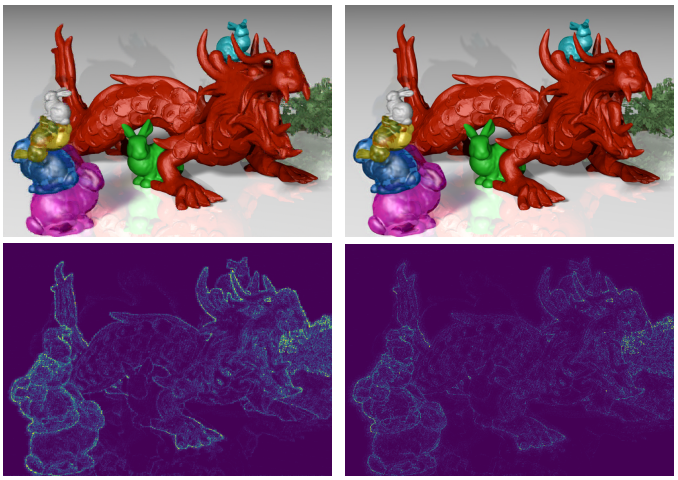


Fig. 6: PSNR of light fields reconstructed with different methods: Kalantari *et al.* [10], Vadathya *et al.* [33], Nabati *et al.* [35], FSR [11] in 4D, and ours (OFS and OFS+refinement).



Miandji *et al.* [15] Ours

Fig. 7: Reconstruction quality comparison on the light field *Dragon* at 4% sampling rate. Top: reconstructed images. Bottom: difference from the ground truth (magnified by 5).

runtimes range from 3 to 6 hours for the OFS method, depending on the light field resolution, plus 3 to 4 additional hours if the non-discrete frequency refinement is on. This computational time is to be compared with the 6-8 hours that the SFFT-based method requires to reconstruct one light field on a 16-core machine.

3) Comparison with deep learning based methods: Finally, we compare our method to three deep learning based reconstruction methods: Kalantari *et al.* [10] use 4 corner viewpoints to reconstruct the full 8×8 views, Vadathya *et al.* [33] reconstruct the light field at 7×7 angular resolution from a coded image, and finally Nabati *et al.* [35] use a coded color mask to even lower the sampling rate to $1/(25 \times 3) = 1.3\%$ to reconstruct 5×5 viewpoints.

We apply the same compression ratios to our method for comparison purposes, using different light fields of [10]: *Flower 1*, *Rock*, *Flower 2*, *Seahorse* and *Cars*. The results are summarized in Fig. 6. The PSNR results for [33] and [35] were extracted from the corresponding articles respectively, more results could be presented for the method in [10] for which the software was available. As shown in the graphs of the figure, our method outperforms the approach in [10] by at least 1.22dB, and achieves high PSNR values of more than 32dB at only 5% of input data. The reconstruction quality is lower nevertheless if we take a very low number of samples, of 2% or lower, where the compressive sensing theory is limited to achieve a correct restoration of the missed data. Fig. 9 shows a visual comparison of our reconstruction results to the ones from [10] on the view $(4, 4)$ of each 8×8 tested light field. The difference images demonstrate some diffuse small errors in our results with a minimum value of SSIM equal to 0.969 while the results in [10] exhibit important errors on edges, probably related to disparity estimation errors.

Please note that the light fields used in the tests exhibit different disparity ranges. While the disparity of the real Lytro light fields is small, around 1 or 2 pixels, the Stanford light fields have disparities between 3 and 6 pixels. When the disparity is large, the sparsity of the 4D light field signal within the 4D hyper-block, which means that more high frequencies are needed to represent the signal in the Fourier domain. In this context, our method can still give a perform good quality reconstruction, provided that the number of frequencies to be selected is increased. The depth-based method [1] (that the reviewer gives reference to) is known to perform well for small disparities of light fields, since the first CNN used for depth estimation relies on features (the mean and variance images as input) that can only describe well the small continuous disparity of dense light fields, typically datasets captured with

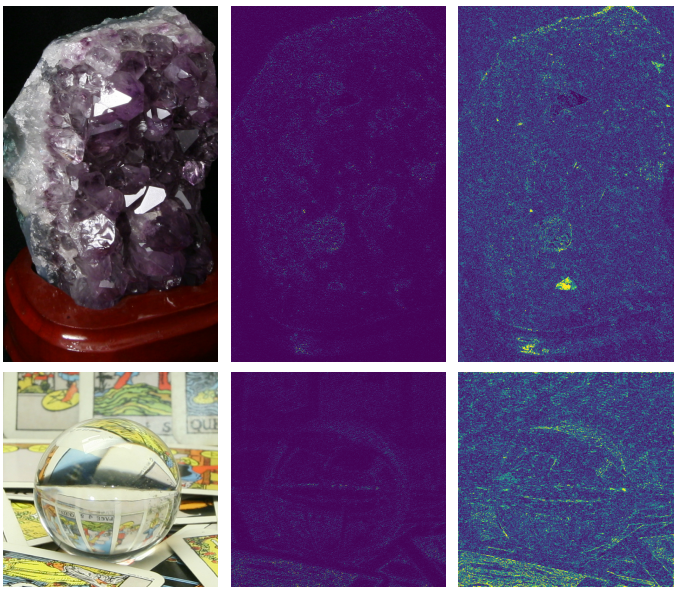
[4]http://lightfield.stanford.edu/lfs.html

TABLE II: Reconstruction quality comparison of real light fields from Stanford Gantry datasets.

	Bunny		*Crystal*		*Amethyst*		*Lego Knights*		*Lego Truck*	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Shi *et al.* [14]	40.44	0.9797	32.96	0.9633	35.33	0.9379	31.85	0.8277	39.18	0.941
Ours	**47.49**	**0.9969**	**41.04**	**0.9956**	**41.91**	**0.9906**	**39.65**	**0.9763**	**43.75**	**0.9860**

a Lytro camera. Besides, in the case of wide disparities, the warping step would introduce more errors in larger occluded regions where a simple interpolation is applied. We can already see these errors in the difference images of the results in Fig. 9. With a much larger disparity, one could imagine larger areas of similar errors, due to more occlusions. Note that the tested light fields exhibit small disparities, which helps providing an accurate depth estimation. However, considering large disparities, depth-based reconstruction algorithms would suffer from higher limitations in the depth estimation step.

While the Lytro light fields exhibit small disparities, the ones of the Stanford dataset captured by a camera moving on a gantry have larger parallax (up to 9 pixels). When the disparity is large, 4D hyper-block will be less sparse, hence more high frequencies will be needed to represent the signal in the Fourier domain.



Ground truth Ours Shi *et al.* [14]

Fig. 8: Reconstruction quality comparison with Shi *et al.* [14]. Top: Amethyst. Bottom: Crystal (difference images are magnified by 10).

V. Conclusion

In this paper, we introduced a new iterative block-wise algorithm to compressively reconstruct light field images. We tackle the challenge of capturing high-resolution images of the scene, by storing compressive data and reconstructing the full resolution images using an approximation of the available samples in the Fourier domain. The approximation model is generated by sparsely selecting the Fourier basis functions that best fit the sampled data, while ensuring the orthogonality of the residue to the subspace spanned by the already selected basis functions. The angular frequency positions are furthermore refined to non-integer values in order to preserve the sparsity that may be limited by the small angular sampling. Experimental results show that high-quality reconstruction is achieved by our approximation method, and demonstrate the advantage of the improved version, which increases the quality of the reconstruction while using a reduced number of iterations. Moreover, refining the model to non-integer frequency positions permits a better approximation. Comparisons with state-of-the-art methods demonstrate that our approach is competitive both in terms of PSNR and SSIM, even for low sampling rates. More importantly, our solution does not require multiple shots, or any prior knowledge of the original data. Therefore, the proposed method can be extended to applications that require real-time shooting, such as very high-resolution light field videos.

References

[1] R. Ng, "Digital light field photography," *Stanford University*, pp. 1–203, 2006.

[2] T. Georgiev, G. Chunev, and A. Lumsdaine, "Super-resolution with the focused plenoptic camera," *Proc. of SPIE*, vol. 7873, pp. 78 730X–78 730X–13, 2011.

[3] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. on Graphics (TOG)*, vol. 24, no. 3, 2005.

[4] K. Venkataraman, D. Lelescu, J. Duparre, A. McMahon, G. Molina, P. Chatterjee, and R. Mullis, "PiCam: An Ultra-Thin High Performance Monolithic Camera Array," *ACM Transactions on Graphics, Proceedings of SIGGRAPH Asia*, vol. 32, 2014.

[5] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing," *ACM Transactions on Graphics*, vol. 26, 2007.

[6] M. Hirsch, S. Sivaramakrishnan, S. Jayasuriya, A. Wang, A. Molnar, R. Raskar, and G. Wetzstein, "A switchable light field camera architecture with angle sensitive pixels and dictionary-based sparse coding," *IEEE International Conference on Computational Photography (ICCP)*, 2014.

[7] J. Iseringhausen, B. Goldlücke, N. Pesheva, S. Iliev, A. Wender, M. Fuchs, and M. B. Hullin, "4D Imaging through Spray-On Optics," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 35:1–35:11, 2017.

[8] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured lumigraph rendering," *ACM SIGGRAPH*, 2001.

[9] A. Davis, M. Levoy, and F. Durand, "Unstructured light fields," *EUROGRAPHICS*, vol. 31, no. 2, 2012.

[10] N. K. Kalantari, T. C. Wang, and R. Ramamoorthi, "Learning-Based View Synthesis for Light Field Cameras," *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, vol. 35, no. 6, 2016.

[11] J. Seiler, M. Jonscher, M. Schoberl, and A. Kaup, "Resampling Images to a Regular Grid From a Non-Regular Subset of Pixel Positions Using Frequency Selective Reconstruction," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4540–4555, 2015.

[12] F. Hawary, G. Boisson, C. Guillemot, and P. Guillotel, "Compressive 4D Light Field Reconstruction Using Orthogonal Frequency Selection," *IEEE ICIP*, pp. 3863–3867, 2018.

[13] F. Hawary, C. Guillemot, D. Thoreau, and G. Boisson, "Scalable Light Field Compression Scheme Using Sparse Reconstruction and Restoration," *IEEE ICIP*, pp. 3250–3254, 2017.

[14] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, "Light field reconstruction using sparsity in the continuous Fourier domain," *ACM Transactions on Graphics*, vol. 34, no. 1, pp. 1–13, 2014.

[15] E. Miandji, J. Kronander, and J. Unger, "Compressive Image Reconstruction in Reduced Union of Subspaces," *Computer Graphics Forum*, vol. 34, no. 2, pp. 33–44, 2015.

[16] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," in *Proc. of ACM SIGGRAPH*, vol. 27, 2008, pp. 1–10.

[17] Z. Xu and E. Y. Lam, "A high-resolution light field camera with dual-mask design," *In Proc. SPIE*, vol. 8500, p. 85000U, 2012.

[18] Q. D. Zhoutong Zhang, Yebin Lin, "Light Field from Micro-baseline Image Pair," *CVPR*, 2015.

[19] Y. Yagi, K. Takahashi, T. Fujii, T. Sonoda, and H. Nagahara, "PCA-coded aperture for light field photography," *IEEE ICIP*, no. 2, pp. 3031–3035, 2017.

[20] ——, "Designing Coded Aperture Camera Based on PCA and NMF for Light Field Acquisition," *IEICE Transactions on Information and Systems*, vol. 101, pp. 2190–2200, 2018.

[21] Y. Inagaki, Y. Kobayashi, K. Takahashi, T. Fujii, and H. Nagahara, "Learning to Capture Light Fields through a Coded Aperture Camera," *The European Conference on Computer Vision (ECCV)*, pp. 418–434, 2018.

[22] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014.

[23] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng, "Learning to Synthesize a 4D RGBD Light Field from a Single Image," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2262–2270, 2017.

[24] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light Field Reconstruction Using Deep Convolutional Network on EPI," *CVPR*, pp. 1638–1646, 2017.

[25] G. Wu, Y. Liu, L. Fang, Q. Dai, and T. Chai, "Light field reconstruction using convolutional network on epi and extended applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1681–1694, 2019.

[26] E. J. Candès and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, p. 21–30, 2008.

[27] E. J. Candès, N. Braun, and M. B. Wakin, "Sparse Signal and Image Recovery from Compressive Samples," in *Proc. of the Int. Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 976–979.

[28] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 32, no. 4, pp. 1–11, 2013.

[29] E. Miandji, J. Unger, and C. Guillemot, "Multi-shot single sensor light field camera using a color coded mask," *EUSIPCO*, pp. 1–5, 2018.

[30] A. Ashok and M. A. Neifeld, "Compressive Light Field Imaging," *In Proc. SPIE*, vol. 7690, p. 76900Q, 2010.

[31] Y. P. Wang, L. C. Wang, D. H. Kong, and B. C. Yin, "High-Resolution Light Field Capture with Coded Aperture," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5609–5618, 2015.

[32] S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, "Compressive Light Field Sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4746–4757, 2012.

[33] A. K. Vadathya, S. Cholleti, G. Ramajayam, V. Kanchana, and K. Mitra, "Learning Light Field Reconstruction from a Single Coded Image," *ACPR*, 2017.

[34] M. Gupta, A. Jauhari, K. Kulkarni, S. Jayasuriya, A. Molnar, and P. Turaga, "Compressive Light Field Reconstructions Using Deep Learning," *CVPRW*, pp. 1277–1286, 2017.

[35] O. Nabati, D. Mendlovic, and R. Giryes, "Fast and Accurate Reconstruction of Compressed Color Light Field," *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–11, 2018.

[36] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 1831–1838.

[37] H. Hassanieh, P. Indyk, D. Katabi, and E. Price, "Simple and practical algorithm for sparse fft," *In Proc. of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, p. 1183–1194, 2012.

[38] B. Ghazi, H. Hassanieh, P. Indyk, D. Katabi, E. Price, and L. Shi, "Sample-optimal average-case sparse Fourier Transform in two dimensions," *51st Annual Allerton Conference on Communication, Control, and Computing*, pp. 1258–1265, 2013.

[39] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using Shearlet transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 133–147, 2018.

[40] ——, "Accelerated Shearlet-Domain Light Field Reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1082–1091, 2017.

[41] S. Mallat and Z. Zhang, "Matching Pursuits with Time-frequency Dictionaries," *Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

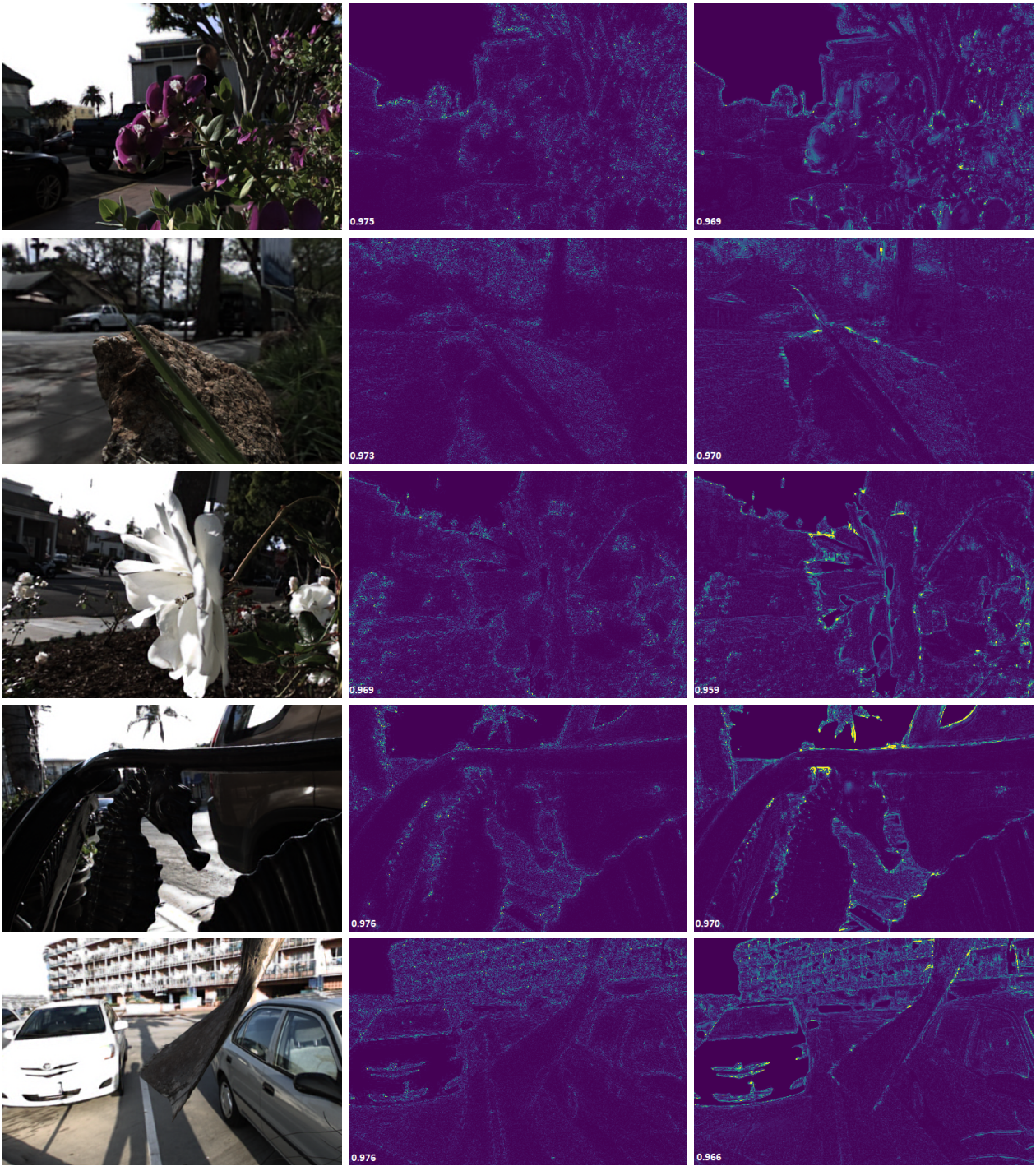


Fig. 9: Reconstruction quality obtained with different methods. Light fields (Top to bottom) *Flower 1*, *Rock*, *Flower 2*, *Seahorse* and *Cars*. (Left to right) our reconstruction image, difference of our result to ground truth, difference of result from [10] to ground truth (difference images are magnified by 5).