IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 October 2023, accepted 28 October 2023, date of publication 1 November 2023, date of current version 8 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329328

RESEARCH ARTICLE

Joint Fourier Disparity Layers Unrolling With Learned View Synthesis for Light Field Reconstruction From Few-Shots Focal Stacks

BRANDON LE BON[1], (Member, IEEE), MIKAËL LE PENDU[2], AND CHRISTINE GUILLEMOT[1], (Fellow, IEEE)

[1]INRIA Rennes—Bretagne Atlantique, 35042 Rennes, France
[2]INTERDIGITAL, 35510 Cesson-Sévigné, France

Corresponding author: Brandon Le Bon (brandon.le-bon@inria.fr)

This work was supported by the French ANR Research Agency in the Context of the Artificial Intelligence Project DeepCIM.

ABSTRACT This paper addresses the problem of capturing a light field using a single traditional camera, by solving the inverse problem of dense light field reconstruction from a focal stack containing only very few images captured at different focus distances. An end-to-end joint optimization framework is presented, where a novel unrolled optimization method is jointly optimized with a view synthesis deep neural network. The proposed unrolled optimization method constructs Fourier Disparity Layers (FDL), a compact representation of light fields which samples Lambertian non-occluded scenes in the depth dimension and from which all the light field viewpoints can be computed. Solving the optimization problem in the FDL domain allows us to derive a closed-form expression of the data-fit term of the inverse problem. Furthermore, unrolling the FDL optimization allows to learn a prior directly in the FDL domain. In order to widen the FDL representation to more complex scenes, a Deep Convolutional Neural Network (DCNN) is trained to synthesize novel views from the optimized FDL. We show that this joint optimization framework reduces occlusion issues of the FDL model, and outperforms recent state-of-the-art methods for light field reconstruction from focal stack measurements.

INDEX TERMS Unrolled optimization, view synthesis, joint optimization, Fourier disparity layer, light field reconstruction, focal stack.

I. INTRODUCTION

In a conventional camera, each sensor element sums all the light rays emitted by one point over the lens aperture. In contrast, light field camera architectures aim at capturing the radiance of every light ray, at every position (x, y, z), in every direction (θ, ϕ), for every wavelength λ at any time t, thus enabling functionalities useful for computer vision applications such as post-capture scene refocusing [1], synthetic aperture imaging [2], or depth estimation [3]. An intuitive approach for capturing light fields consists in taking pictures from several viewpoints, either simultaneously thanks to a large camera array [4], or sequentially with a single camera placed on a moving gantry [5]. Alternatively, more lightweight camera designs have been proposed to capture light fields on a single 2D sensor. Plenoptic cameras [6] are based on an array of microlenses placed in front of the photosensor to separate the light rays striking each microlens into a small image on the photosensors pixels, however at the cost of sacrificing the spatial resolution for the angular resolution. More recent designs consider coded masks to modulate 4D light fields into 2D projections captured by 2D digital camera sensors [7], [8], [9]. An alternative, which does not require hardware modifications to conventional cameras, consists of capturing a focal stack, i.e. several images of the scene at different focus distances, in order to reconstruct a light field. However, existing reconstruction methods [10], [11] typically require focal stacks with dense

The associate editor coordinating the review of this manuscript and approving it for publication was Deepak Mishra.

sampling in the focus dimension, so that the details can be retrieved at every depth in the scene. Hence, many shots are needed in the capture process.

The problem of reconstructing a light field from a focal stack with only a few shots can be seen as a form of compressive sensing, hence posed as an ill-posed image inverse problem. A common strategy to deal with ill-posed image inverse problems consists of introducing image priors as regularization terms in optimization methods. The optimization problem is then posed as the minimization of a function composed of two terms: a data-fidelity term and a regularization term. The data-fidelity term measures the fidelity of the solution to the measurements. It is usually expressed as the squared error between the measurements and the estimated solution on which a measurement operator, representing the image formation model, is applied. The regularization term is used to rule out solutions that are unlikely according to our prior knowledge on images.

While traditional approaches consider hand-crafted priors to regularize the optimization problem, such as sparsity [12], smoothness [13] or total variation [11], significant advances have been achieved thanks to the introduction of learned priors. A first category of methods, referred to as ''Plug-and-play'' (PnP) [14], [15], has been introduced where a pre-learned prior is plugged into an iterative optimization algorithm. One advantage of the PnP approach is its genericity in the sense that the learned priors do not need to be re-trained for each new image inverse problem. However, learning priors independently of the targeted inverse problem may not yield the best solution. Unrolled iterative algorithms, introduced in [16], have emerged as a way to learn an optimized task-specific image prior within an iterative optimization algorithm. A learnable network is trained end-to-end within a fixed number of iterations of an iterative optimization algorithm to optimize the learnable prior for a specific image inverse problem. Many iterative algorithms have been unrolled, e.g. the Iterative Shrinkage Thresholding Algorithm (ISTA) [16], the gradient descent [17], the Half-Quadratic Splitting (HQS) [18], the Alternating Direction Method of Multipliers (ADMM) [19]. Since computing the proximal operator of the regularization term is equivalent to applying Gaussian denoising [15], a well-known approach is to unroll a proximal optimization algorithm, e.g. the ADMM, in order to learn a deep denoiser instead of learning the regularization function directly. Learning a Gaussian denoiser has the advantage of tackling a well-studied problem with specific deep network architectures [15], [20]. Thanks to the capacity of deep neural networks to learn complex image priors, several works in the literature achieved state-of-the-art reconstruction performances for a plethora of 2D image inverse problems [17], [21], [22], [23].

Unrolling optimization algorithms in the context of light field reconstruction from a limited number of measurements is, however, a challenging task. Indeed, the measurement operator representing the light field image formation model is usually very large or hard to represent numerically, hence making the data-fidelity optimization problem hard to solve efficiently in an unrolled optimization algorithm. An unrolled HQS optimization approach has been recently introduced for light field reconstruction from coded projections [24], where a novel way to compute the solution of the proximal operator of the data-fidelity term is presented to avoid the computational burden due to the size of the measurement operator.

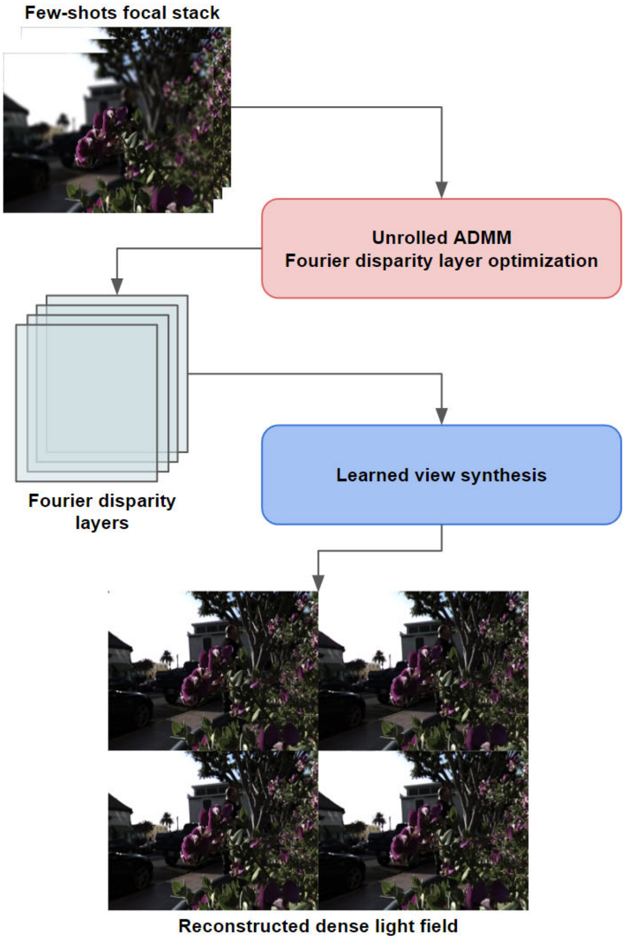


FIGURE 1. **Overview of the proposed joint optimization framework.**

In the context of light field reconstruction from focal stack measurements, optimization algorithms that have been designed in the literature mostly use handcrafted priors [11], [25], [26], [27], [28], [29]. One of them uses a Tikhonov regularization to optimize the Fourier Disparity Layers (FDL) representation of a light field [29], which has been introduced as a compact representation of scenes. It samples the light field in the depth dimension to decompose the scene as a discrete sum of layers in the Fourier domain, hence the name Fourier Disparity Layers. Each layer contains the texture of the scene that corresponds to a specific disparity/depth value, and is computed from input views or focal stack images by solving an optimization problem using a regularized least squares regression approach. One important property of the FDL representation is that the data-fidelity term of the FDL optimization is posed per-frequency component and its proximal operator is a small-scale linear least squares problem. Hence, it can be solved efficiently for each frequency in parallel to reduce the computational burden. However, the FDL optimization requires to design a regularization function directly in the FDL domain which is a challenging task. Furthermore, this representation of light fields assumes non-occluded scenes with Lambertian reflectance. Therefore, artifacts, such as transparency in occluded regions, may appear with the FDL model [29], [30].

Based on these properties, the FDL model is well-suited for unrolled optimization approaches. Indeed, on one hand, the FDL model allows us to efficiently compute a solution to the data-fidelity term. On the other hand, unrolling the FDL optimization allows us to learn the regularization function directly in the FDL domain. Consequently, we proposed, in a previous paper [31], an unrolled ADMM FDL optimization method. In this paper, we enhance this method and present a joint unrolled FDL optimization with a learned view synthesis, as illustrated in Figure 1, to address the problem of light field reconstruction from a small set of focal stack images, which can be captured with a single traditional camera. The main contributions of this work are summarized as follows:

- A novel learned view synthesis process to reconstruct light field views from optimized FDL is presented. A Deep Convolutional Neural Network (DCNN) is added within the FDL view synthesis process and is trained to minimize the errors of the reconstructed views, in particular to cope with the occlusion and reflectance issues of the FDL model, as shown in the ablation study.
- A joint optimization framework is presented, coupling our unrolled ADMM FDL optimization framework in [31] with the proposed learned view synthesis process. Both parts of the network are jointly optimized to ensure the best reconstruction performances. The proposed joint optimization framework significantly outperforms the state-of-the-art methods for light field reconstruction from a set of focal stack images.

II. BACKGROUND AND RELATED WORKS

A. IMAGE INVERSE PROBLEMS

An image inverse problem is the problem of recovering an image $\mathbf{x}$ from a set of incomplete or corrupted measurements $\mathbf{b}$. Formally, an acquisition process can be represented by the following linear system:

$$\mathbf{b} = \mathcal{T}(\mathbf{x}) + \epsilon, \tag{1}$$

where $\mathcal{T}$ is the measurement operator and ϵ is an additive noise. In several applications, the measurement operator $\mathcal{T}$ is ill-conditioned, meaning that the inverse problem of reconstructing the original signal $\mathbf{x}$ from the measurements $\mathbf{b}$ is ill-posed, i.e. an admissible numerical solution is hard to find. A regularized minimization problem is then usually posed, introducing an image prior via a regularization constraint $\mathcal{R}$ along with a data-fidelity term. The optimization problem of recovering an estimate $\hat{\mathbf{x}}$ is thus posed as follows:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathcal{T}(\mathbf{x}) - \mathbf{b}\|_2^2 + \lambda\mathcal{R}(\mathbf{x}), \tag{2}$$

where λ controls the amount of regularization.

B. UNROLLED OPTIMIZATION WITH DEEP PRIORS

Classical approaches to solve the optimization problem posed in Eq. (2) use iterative optimization algorithms, usually introducing hand-crafted [11] or pre-learned [14], [15] image priors. Unrolled optimization methods have enabled major progress in the field of image inverse problems. The idea behind unrolling an iterative optimization algorithm is to learn an image prior within the algorithm so that it performs best for a given iterative optimization algorithm and for a given task. Several iterative optimization algorithms have been unrolled in the literature [16], [17], [18], [19]. In this paper, we focus on the Alternating Direction Method of Multipliers (ADMM) [32], that solves the problem in Eq. (2) by decoupling the data-fidelity term and the regularization term, and with each iteration consisting of the following steps:

$$\hat{\mathbf{x}}^{i+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathcal{T}(\mathbf{x}) - \mathbf{b}\|_2^2 + \frac{\rho}{2}\left\|\mathbf{x} - \mathbf{y}^i + \mathbf{u}^i\right\|_2^2, \tag{3}$$

$$\mathbf{y}^{i+1} = \arg\min_{\mathbf{y}} \frac{\rho}{2}\left\|\mathbf{y} - (\hat{\mathbf{x}}^{i+1} + \mathbf{u}^i)\right\|_2^2 + \lambda \cdot \mathcal{R}(\mathbf{y}), \tag{4}$$

$$\mathbf{u}^{i+1} = \mathbf{u}^i + (\hat{\mathbf{x}}^{i+1} - \mathbf{y}^{i+1}), \tag{5}$$

where ρ is a penalty parameter, $\mathbf{u}$ is called the dual variable which is typically zero-initialized, and $\mathbf{y}$ is an auxiliary variable with $\mathbf{y}^0$ being the initial image estimate. One can note that the sub-problem (4) performs Gaussian denoising of $(\hat{\mathbf{x}}^{i+1} + \mathbf{u}^i)$ assuming a noise variance λ/ρ and under the prior defined by $\mathcal{R}$. Hence, instead of learning $\mathcal{R}$ directly, unrolled ADMM optimization methods typically learn a deep denoiser $\mathcal{D}$ with trainable parameters θ, and replace Eq. (4) with:

$$\mathbf{y}^{i+1} = \mathcal{D}(\hat{\mathbf{x}}^i + \mathbf{u}^i; \theta). \tag{6}$$

The deep denoiser $\mathcal{D}$ can thus be trained within the ADMM optimization algorithm so that unrolling a given number N of ADMM iterations gives the estimate $\hat{\mathbf{x}}^N$ that best reconstructs the ground truth $\mathbf{x}$ for a training image dataset.

C. LIGHT FIELD IMAGING MODEL

Let us consider an input light field, represented by a 4D function $L(x, y, u, v)$ describing the radiance along light rays, with the two-plane parameterization proposed in [5] and [33]. The parameters (u, v) denote the angular (view) coordinates and (x, y) the spatial (pixel) coordinates. In this paper, for notation simplicity and without loss of generality, we consider a 2D light field $L(x, u)$ with one angular dimension and one spatial dimension. Focal stack images taken at different focus distances can be seen as measurements of the light field to be reconstructed. Let a refocused light field L^s be defined as $L^s(x, u) = L(x - us, u)$, with a refocus parameter s. A refocused image $I^s_{u_0}$, at position u_0 on the camera plane, is obtained by integrating the light rays over the angular dimension using the refocused light field and the camera aperture ψ:

$$I^s_{u_0}(x) = \int_{\mathbb{R}} L(x - us, u_0 + u)\psi(u)du. \tag{7}$$

D. LIGHT FIELD RECONSTRUCTION FROM A FOCAL STACK

Optimization methods for light field reconstruction from a set of focal stack images have been introduced, at first, without any image prior. Takahashi et al. [25] proposed an iterative method to construct a light field representation named "tensor-display" from a focal stack. The scene is decomposed into a few light-attenuating layers, from which the light field views can be synthesized. Using the similarity between both light field reconstruction from a focal stack and CT image reconstruction tasks, Liu et al. [26] applied the filtered back-projection and the Landweber iterative methods for light field reconstruction. Yin et al. [27] presented a filter-based iterative method to solve the inverse problem with a linear projection system used to model the focal stack imaging process. Another filter-based iterative method was proposed by Gao et al. [28]. The paper introduces an optimized relaxation strategy and a fast-guided filter in the filter-based Landweber iterative method. Lien et al. [34] proposed a method for light field reconstruction from a focal stack captured in one shot with a stack of transparent graphene photodetectors.

Handcrafted priors have then been introduced in the formulation of the inverse problem of recovering light fields from focal stacks. Gao et al. [11] proposed the ADMM algorithm with a TV-regularization along with a guided filter. A convolution kernel is derived to model the focal stack imaging process. Additionally to sparsity priors, Blocker et al. [35] and Kamal et al. [36] proposed a low-rank prior to respectively model (i) the low angular variation of light fields (ii) the redundancies of high-dimensional visual signal. Le Pendu et al. [29] proposed the Fourier Disparity Layers (FDL) representation of light fields to decompose the scene into a set of additive layers from which any view can be reconstructed. The author used a Tikhonov regularization constraint in the optimization of the FDL.

Deep learning techniques were also recently considered by Huang et al. [37] to reconstruct a light field from a focal stack. They proposed a three sequential convolutional neural networks framework that reconstructs the light field from estimated all-in-focus images, depth maps, and Lambertian light fields. However, the method does not benefit from the advantages of having a data-fidelity block as in optimization algorithms.

In this paper, we propose to combine the FDL model in [29] with an unrolled ADMM optimization method along with a learned view synthesis process in order to introduce learned priors in the context of light field reconstruction from focal stack images.

III. JOINT FOURIER DISPARITY LAYERS UNROLLING WITH LEARNED VIEW SYNTHESIS

In this section, we present our joint optimization framework, illustrated in Figure 2. We first introduce in Section III-A the Fourier Disparity Layers (FDL) by Le Pendu et al. [29] that will be used in our framework. The proposed method is a joint optimization of two different parts introduced in Sections III-B and III-C (i) the parameters θ_1 of a denoiser CNN $\mathcal{D}$ used in an unrolled ADMM FDL optimization, as we proposed in [31] (ii) and the parameters θ_2 of a CNN $\mathcal{S}$ of a novel learned view synthesis process trained to adapt the optimized FDL for each novel view to be reconstructed, in order to cope with the issues of the FDL model. Finally, in Section III-D, we present the joint optimization process.

A. FOURIER DISPARITY LAYERS

Fourier Disparity Layers (FDL) have been introduced in [29] as a compact representation of dense light fields. The FDL model consists of a set of additive layers $\tilde{L}^k$, each associated with a disparity value d^k, inversely proportional to the depth, where each layer mostly contains details in the regions of disparity d^k in the scene. The FDL model is defined such that a sub-aperture view, or viewpoint, at angular coordinate u_0 is reconstructed by shifting each layer L^k by $d_k u_0$, and by summing the shifted layers.

Formally, let a Lambertian non-occluded scene be divided into n spatial regions Ω_k with constant disparity d_k. The Fourier transform $\tilde{L}(\omega_x, \omega_u)$ of a light field $L(x, u)$ can thus be re-written such that the spatial information remains the same for any view:

$$\tilde{L}(\omega_x, \omega_u) = \sum_k \delta(\omega_u - d_k \omega_x)\tilde{L}^k(\omega_x), \tag{8}$$

with $\tilde{L}^k(\omega_x)$, the FDL associated with the disparity d_k, defined by:

$$\tilde{L}^k(\omega_x) = \int_{\Omega_k} e^{-2i\pi x \omega_x} L(x, 0)dx. \tag{9}$$

The relation between the Fourier transform $\tilde{I}^s_{u_0}(\omega_x)$ of a refocus image of a focal stack and the FDL is thus established in [29] as follows:

$$\tilde{I}^s_{u_0}(\omega_x) = \sum_k e^{+2i\pi u_0 d_k \omega_x} \tilde{\psi}(\omega_x(s - d_k)) \cdot \tilde{L}^k(\omega_x). \tag{10}$$

Based on Eq. (10), we can define an optimization algorithm to optimize the FDL from a set of refocused images. Our proposed optimization algorithm will be further detailed in section III-B. It is important to notice that Eqs. (8), (9), and (10) are only verified in the case of non-occluded scenes with Lambertian reflectance. Assuming this model, performing an FDL optimization algorithm will produce light field views with occlusion and reflectance artifacts, e.g. transparency in occluded areas, as illustrated in recent works [29], [30]. We address this problem in section III-C by proposing a neural network based view synthesis process to reconstruct light field views from the optimized FDL.

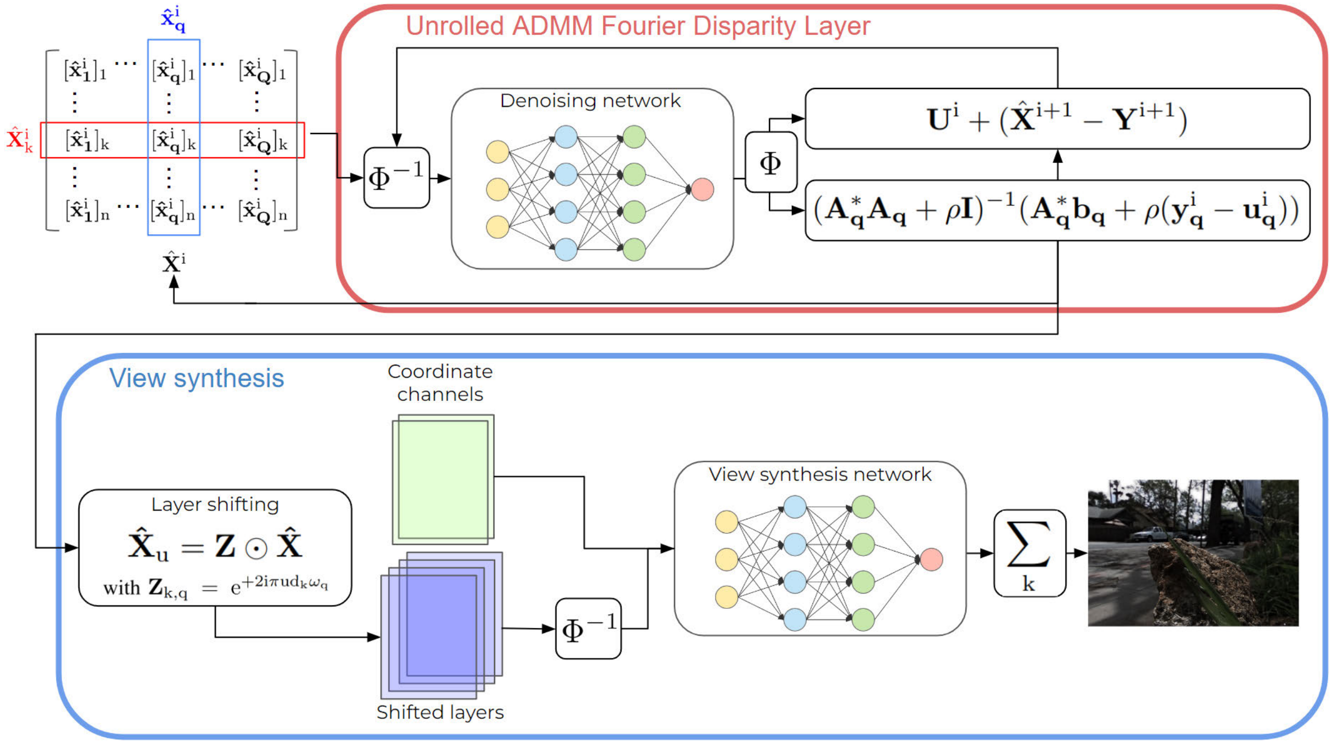

$\hat{x}^i_q$
$[\hat{x}^i_1]_1$ ⋯ $[\hat{x}^i_q]_1$ ⋯ $[\hat{x}^i_Q]_1$
$\hat{x}^i_k$
$[\hat{x}^i_1]_k$ $[\hat{x}^i_q]_k$ $[\hat{x}^i_Q]_k$
$[\hat{x}^i_1]_n$ ⋯ $[\hat{x}^i_q]_n$ ⋯ $[\hat{x}^i_Q]_n$
$\hat{X}^i$
Unrolled ADMM Fourier Disparity Layer
Φ^{-1}
Denoising network
Φ
$\mathbf{U}^i + (\hat{\mathbf{X}}^{i+1} - \mathbf{Y}^{i+1})$
$(\mathbf{A}^*_q\mathbf{A}_q + \rho\mathbf{I})^{-1}(\mathbf{A}^*_q\mathbf{b}_q + \rho(\mathbf{y}^i_q - \mathbf{u}^i_q))$
View synthesis
Coordinate channels
Layer shifting
$\hat{\mathbf{X}}_u = \mathbf{Z} \odot \hat{\mathbf{X}}$
with $\mathbf{Z}_{k,q} = e^{+2i\pi u d_k \omega_q}$
Shifted layers
Φ^{-1}
View synthesis network
$\sum_k$

FIGURE 2. Architecture of the proposed end-to-end framework for light field reconstruction from focal stack measurements. The pipeline is composed of two blocks: (i) an unrolled ADMM FDL optimization (red block) which optimizes a matrix $\hat{\mathbf{X}}^i$, where each row $\hat{\mathbf{X}}^i_k$ corresponds to the vectorized FDL k (ii) a view synthesis (blue block) with a learned network, where the optimized FDL are shifted and concatenated with additional coordinate channels to indicate which view to reconstruct to the network.

B. UNROLLED ADMM FDL OPTIMIZATION

Let us consider an input focal stack containing images I_j. We note m and n respectively the number of measured focal stack images and the number of considered layers in the FDL model. For each spatial frequency component ω_q of index q in the discrete Fourier transform, we note $\mathbf{b_q} \in \mathbb{C}^m$ a vector with $[\mathbf{b_q}]_j = \tilde{I}_j(\omega_q)$, $\mathbf{x_q} \in \mathbb{C}^n$ a vector with $[\mathbf{x_q}]_k = \tilde{L}^k(\omega_q)$, and $\mathbf{A_q} \in \mathbb{C}^{m \times n}$ a matrix defined as follows:

$$[\mathbf{A_q}]_{j,k} = e^{+2i\pi u_j d_k \omega_x}\tilde{\psi}_j(\omega_x(s_j - d_k)). \tag{11}$$

Eq. (10) is thus reformulated as $\mathbf{A_q x_q} = \mathbf{b_q}$. Thus, the construction of the FDL spatial frequencies $\mathbf{x_q}$ from measurements $\mathbf{b_q}$ is posed as a linear least squares optimization problem independently for each frequency component ω_q. The matrices $\mathbf{A_q}$ are usually ill-conditioned, making the latter optimization problem ill-posed. To reduce overfitting on the measurements that may cause severe artifacts in the FDL, the authors in [29] include a Tikhonov regularization term, which results in the following per-frequency minimization problem:

$$\hat{\mathbf{x}}_\mathbf{q} = \arg\min_{\mathbf{x_q}} \left\| \mathbf{A_q x_q} - \mathbf{b_q} \right\|_2^2 + \lambda \left\| \Gamma_\mathbf{q} \mathbf{x_q} \right\|_2^2, \tag{12}$$

with Γ being the Tikhonov matrix. A calibration method is also proposed in [29] to determine the angular coordinate u_0 of each input view and the disparity values d_k of the layers. However, it only applies in the case of sub-aperture images as measurements. In this paper, we consider focal stacks where all the images are taken at the same angular coordinate $u_0 = 0$, and assuming a known focus parameter s and aperture ψ. For the disparity values d_k of the FDL model, we use uniformly sampled values over the disparity range of the scene.

While the author in [29] uses a Tikhonov regularization to encourage smooth variations between the light field views generated by the optimized FDL, designing a more complex prior directly in the FDL domain is a challenging task. To cope with this issue, we propose to unroll the FDL optimization, following the ADMM unrolling framework, in order to automatically learn a deep prior in the FDL domain. In order to account for complex image statistics on the FDL model, we consider a regularization of the full layers, rather than a per-frequency regularization as in Eq. (12). Furthermore, since most neural networks operate on images in the pixel domain, we regularize the images obtained by the inverse Fourier transform of the FDL layers.

Let us define the matrix $\mathbf{X} = [\mathbf{x_1}|\ldots|\mathbf{x_Q}]$ representing the full FDL as a concatenation of the column vectors $\mathbf{x_q}$ for all the frequency components ω_q with $q \in [1..Q]$. The regularized FDL reconstruction problem is then formulated as:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \left(\lambda \cdot \mathcal{R}(\mathbf{X}\Phi^{-1}) + \sum_q \left\| \mathbf{A_q x_q} - \mathbf{b_q} \right\|_2^2 \right), \tag{13}$$

where Φ^{-1} is the inverse 2D Fourier transform, applied to each FDL layer (i.e. rows of $\mathbf{X}$) to regularize the images in the pixel domain. The steps of the unrolled ADMM iteration in Eqs. (3), (6), (5), can then be written:

$$\hat{\mathbf{x}}_{\mathbf{q}}^{i+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\left\|\mathbf{A}_{\mathbf{q}}\mathbf{x}-\mathbf{b}_{\mathbf{q}}\right\|_2^2+\frac{\rho}{2}\left\|\mathbf{x}-\mathbf{y}_{\mathbf{q}}^{i}+\mathbf{u}_{\mathbf{q}}^{i}\right\|_2^2, \quad (14)$$

$$\mathbf{Y}^{i+1} = \mathcal{D}((\hat{\mathbf{X}}^{i+1}+\mathbf{U}^{i})\Phi^{-1};\theta_1)\Phi, \quad (15)$$

$$\mathbf{U}^{i+1} = \mathbf{U}^{i}+(\hat{\mathbf{X}}^{i+1}-\mathbf{Y}^{i+1}), \quad (16)$$

where we note $\hat{\mathbf{X}}^i = [\hat{\mathbf{x}}_{\mathbf{1}}^i|\ldots|\hat{\mathbf{x}}_{\mathbf{Q}}^i]$ and $\hat{\mathbf{Y}}^i = [\hat{\mathbf{y}}_{\mathbf{1}}^i|\ldots|\hat{\mathbf{y}}_{\mathbf{Q}}^i]$. For the regularization, one can see in Eq. (15) that denoising can be applied in the pixel domain by performing the inverse 2D Fourier transform of the denoiser's input layers $(\hat{\mathbf{X}}^{i+1}+\mathbf{U}^i)$, and reapplying the 2D Fourier transform on the denoised output. Instead of using a pre-learned denoiser as in the Plug-and-Play approach [15], [38], the denoiser $\mathcal{D}$ is here trained end-to-end within the unrolled algorithm to better train it for the task of FDL denoising. On the other hand, the data-fidelity subproblem in Eq. (14) can still be solved independently per-frequency component, and has a well-known closed form solution:

$$\hat{\mathbf{x}}_{\mathbf{q}} = (\mathbf{A}_{\mathbf{q}}^*\mathbf{A}_{\mathbf{q}}+\rho\mathbf{I})^{-1}(\mathbf{A}_{\mathbf{q}}^*\mathbf{b}_{\mathbf{q}}+\rho(\mathbf{y}_{\mathbf{q}}^{i}-\mathbf{u}_{\mathbf{q}}^{i})), \quad (17)$$

where $\mathbf{I}$ is the identity matrix and * is the Hermitian transpose operator. Note that for each frequency component of index q, the matrix inversion $(\mathbf{A}_{\mathbf{q}}^*\mathbf{A}_{\mathbf{q}}+\rho\mathbf{I})^{-1}$ in Eq. (17) can be performed efficiently thanks to the small dimensions of the matrix $\mathbf{A}_{\mathbf{q}}$ ($\mathbf{A}_{\mathbf{q}}^*\mathbf{A}_{\mathbf{q}} \in \mathbb{C}^{n\times n}$, with n the number of layers). The per-frequency computation of the proximal operator allowed by the FDL model thus significantly reduces the computational burden of computing the estimate $\hat{\mathbf{X}}$.

C. VIEW SYNTHESIS FROM OPTIMIZED FDL

As derived in [29], the Fourier transform $\tilde{L}_u(\omega_x)$ of a view $L(x,u)$ can be reconstructed by applying a shift-and-sum on the optimized k FDL $\tilde{L}^k(\omega_x)$ as follows:

$$\tilde{L}_u(\omega_x) = \sum_k e^{+2i\pi u d_k \omega_x}\tilde{L}^k(\omega_x). \quad (18)$$

However, it is well-known that artifacts will occur in specific areas with this technique, e.g. in occluded regions [29], [30], as mentioned in Section III-A. Since these artifacts are different for each reconstructed light field view, we need to slightly adjust the optimized k FDL $\tilde{L}^k(\omega_x)$ for each novel view. Since the ground truth views $L_{gt}(x,u)$ are known during the training phase, we propose to train the parameters θ_2 of a CNN $\mathcal{S}$ to modify the k optimized FDL $\tilde{L}^k(\omega_x)$ for each view to be reconstructed, such that the reconstructed views well-estimate their corresponding ground truth views. As described in Eq. (9), the FDL contain only the spatial information of the light field within each depth plane. Therefore, we also need to add angular information to the input of the network $\mathcal{S}$ in order to specify which view to reconstruct. We propose to shift the optimized FDL accordingly to the angular coordinates of the view to be reconstructed as in Eq. (18). However, instead of directly summing the shifted layers, we first concatenate them along with two additional channels $\mathbf{C}$, each containing an angular coordinate of the view. The resulting tensor is fed into a view synthesis CNN $\mathcal{S}$ which computes the modified shifted layers. These modified layers are then summed to reconstruct the view, as in Eq. (18).

Formally, let u be the coordinates of the view to be reconstructed, $\hat{\mathbf{X}} \in \mathbb{C}^{n\times Q}$ be the matrix representing the concatenation of the optimized FDL as in Eq. (13), and $\mathbf{Z} \in \mathbb{C}^{n\times Q}$ be a matrix with $\mathbf{Z}_{k,q} = e^{+2i\pi u d_k \omega_q}$. The matrix $\hat{\mathbf{X}}_u \in \mathbb{C}^{n\times Q}$, being the concatenation of the shifted FDL associated to the angular coordinates u, is thus computed as follows:

$$\hat{\mathbf{X}}_u = \mathbf{Z}\odot\hat{\mathbf{X}}, \quad (19)$$

with $\odot$ being the Hadamard product. With $\mathbf{C}_u$ being a channel filled with the value of the angular coordinate u of the view to be reconstructed, the parameters θ_2 of the network $\mathcal{S}$ are thus optimized as follows:

$$\theta_2 = \arg\min_{\theta_2}\left\|\sum_k\left[\mathcal{S}(\hat{\mathbf{X}}_u\Phi^{-1},\mathbf{C}_u;\theta_2)\right]-L_{gt}(x,u)\right\|_2^2, \quad (20)$$

where we compute the inverse Fourier transform of the shifted layers $\hat{\mathbf{X}}_u\Phi^{-1}$ so that the network $\mathcal{S}$ in the view synthesis process operates in the pixel domain, similarly to the denoising network in Eq. (15).

In practice, we observed that pre-training the framework using only the shifted FDL as the input of the network, and then fine-tuning by adding the coordinate channels to the input offers the best performances. One can notice that several approaches can be used to model the input of the network. We further discuss our choice in comparison with other approaches in Section IV-E2.

D. JOINT OPTIMIZATION

The proposed framework is composed of 2 successive optimizations: the unrolled ADMM FDL optimization, with a network $\mathcal{D}$ parameterized with θ_1, described in Section III-B, and the learned view synthesis process, with a network $\mathcal{S}$ parameterized with θ_2, described in Section III-C. Instead of training both networks independently, we propose a joint optimization in an end-to-end framework. A joint optimization of θ_1, θ_2 ensures that both networks are optimized such that the synthesized views well-estimate their corresponding ground truths. Let $\mathcal{F}$ be a function parameterized with θ_1, which computes the application of the whole forward pass of the unrolled ADMM FDL optimization algorithm. The optimization problem of the entire end-to-end framework is:

$$\begin{aligned}\theta_1,\theta_2 &= \arg\min_{\theta_1,\theta_2}\left\|\sum_k\left[\mathcal{S}(\hat{\mathbf{X}}_u\Phi^{-1},\mathbf{C}_u;\theta_2)\right]-L_{gt}(x,u)\right\|_2^2,\\ \text{with}\quad &\hat{\mathbf{X}}_u = \mathbf{Z}\odot\hat{\mathbf{X}},\\ \text{and}\quad &\hat{\mathbf{X}} = \mathcal{F}(\mathbf{b};\theta_1),\end{aligned} \quad (21)$$

where **b** is a vector containing the measured focal stack images. The joint optimization algorithm is described in Algorithm 1. For more implementation details, our pytorch implementation is available at: https://github.com/BrandonLeBon/Joint_Optimization_FDL.

Algorithm 1 : Proposed Joint Optimization

```
initialize $\theta_1$,$\theta_2$
for each training iteration do
    $\hat{\mathbf{X}}^0$, $\mathbf{Y}^0$, $\mathbf{U}^0 \leftarrow 0$
    $\mathbf{b} \leftarrow$ input measurements
    $L\_gt \leftarrow$ groundtruth views
    $L\_recons \leftarrow 0$

    for each unrolled iteration $i$ do
        for each frequency component $\mathbf{q}$ do
            $\hat{\mathbf{x}}_{\mathbf{q}} \leftarrow (\mathbf{A}_{\mathbf{q}}^*\mathbf{A}_{\mathbf{q}} + \rho\mathbf{I})^{-1}(\mathbf{A}_{\mathbf{q}}^*\mathbf{b}_{\mathbf{q}} + \rho(\mathbf{y}_{\mathbf{q}}^i - \mathbf{u}_{\mathbf{q}}^i))$
        end for
        $\mathbf{Y}^{i+1} \leftarrow \mathcal{D}((\hat{\mathbf{X}}^{i+1} + \mathbf{U}^i)\Phi^{-1}; \theta_1)\Phi$
        $\mathbf{U}^{i+1} \leftarrow \mathbf{U}^i + (\hat{\mathbf{X}}^{i+1} - \mathbf{Y}^{i+1})$
    end for

    for each view coordinate $u$ do
        $\hat{\mathbf{X}}_u^I = \mathbf{Z} \odot \hat{\mathbf{X}}^I$
        $L\_recons_u \leftarrow \sum_k \left[\mathcal{S}(\hat{\mathbf{X}}_u^I\Phi^{-1}, \mathbf{C}_u; \theta_2)\right]$
    end for
    $loss = \|L\_recons - L\_gt\|_2^2$
    $\theta_1 = \theta_1 - \lambda\nabla_{\theta_1}(loss)$
    $\theta_2 = \theta_2 - \lambda\nabla_{\theta_2}(loss)$
end for
```

IV. EXPERIMENTS

We assess our framework for light field reconstruction from focal stacks containing very few shots, i.e. with 2 and 3 shots. We compare the proposed method against the most recent and efficient state-of-the-art methods for this task: the Fourier Disparity Layers by Le Pendu et al. [29], the TV regularized sparse light field reconstruction model based on guided-filtering recently proposed by Gao et al. [11], the light field reconstruction and depth estimation using convolutional neural networks proposed by Huang et al. [37], and the Unrolled ADMM Fourier Disparity Layer Optimization by Le Bon et al. [31]. For fair comparisons, the methods of Huang et al. [37] and of Le Bon et al. [31] have been re-trained using the datasets listed in Section IV-A. Additionally, an ablation study is proposed in Section IV-E to study the importance of (i) using jointly the unrolled ADMM FDL optimization and the learned view synthesis network compared to our previous work [31] (ii) using the shifted version of the FDL as well as the angular coordinate channels as network input in the view synthesis process.

A. DATASETS

Two-thirds of both the Stanford Lytro light field archive dataset [40] and the Kalantari dataset [39] were used as training datasets. Reconstruction performances are then evaluated with the remaining third of both datasets along with the Linköping Light Field dataset [41]. The input measurements consist of focal stacks with 2 or 3 images (i.e. shots) synthesized from ground truth views with the shift-and-add method [6] and with focus parameters s covering the disparity range of the scene. As ground truth, a dense light field with a 7 × 7 angular resolution is considered.

B. ARCHITECTURE AND TRAINING SETTINGS

We used the DRUNet denoising architecture as in [15] for both the denoiser $\mathcal{D}$ in Eq. (15) and the view synthesis network $\mathcal{S}$ in Eq. (20). A total of 30 layers in the FDL model and 12 unrolled iterations have been used. Both networks $\mathcal{D}$ and $\mathcal{S}$ use as input the concatenation of all the layers, in order to treat them jointly. For the input of $\mathcal{S}$, the layers are additionally concatenated with the coordinate channels as described in Section III-C. The penalty term ρ in Eq. (14) is trained along with the weights of the two networks. During training, we used a patch size of 64 × 64 with an additional padding of size 8. Networks are trained for 1000 epochs with a learning rate of 10^{-5} and a batch size of 1. The networks have been retrained specifically for each number of measurements. The loss function $\mathcal{L}$ used was the squared ℓ_2-norm between the ground truth light field sub-aperture views and the corresponding synthesized views as defined in Eq. (21)

C. RECONSTRUCTION PERFORMANCES

To evaluate the reconstruction performances of the different methods, we measured the quality of the reconstructed light field views using the PSNR and the SSIM metrics, traditionally used by the the image processing community. Table 1 gives the average PSNR and SSIM values over the three considered testing datasets for light field reconstruction from 2 and 3 focal stack images as measurements. It shows that the proposed approach significantly outperforms all the state-of-the-art methods on every dataset, with an average gain of 1 dB compared to the best approaches. Additionally, Fig. 3 shows a reconstructed central view for each evaluated method. As illustrated in the figure, the proposed joint optimization method better reconstructs finer details compared to other approaches.

D. ALGORITHM COMPLEXITY

In this section, we evaluate the complexity of our proposed joint optimization algorithm compared to other state-of-the-art iterative methods. Since the implementation of the method by Gao et al. [11] in CPU only, we present results obtained on both CPU and GPU for fair comparisons. In Table 3, we computed the average computation time for the different iterative reconstruction algorithms, and the average

TABLE 1. Comparisons with efficient state-of-the-art methods: average PSNR and SSIM for light field reconstruction.

Datasets	Kalantari [39]		Stanford [40]		Linköping [41]	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Number of shots	2					
Huang et al. [37]	31.44 dB	0.880	31.02 dB	0.875	24.18 dB	0.780
Le Pendu et al. [29]	33.01 dB	0.924	34.76 dB	0.933	26.62 dB	0.861
Gao et al. [11]	35.62 dB	0.936	35.12 dB	0.935	27.19 dB	0.853
Le Bon et al. [31]	39.82 dB	0.968	37.32 dB	0.955	29.22 dB	0.902
Joint optimization FDL	40.93 dB	0.974	38.35 dB	0.961	29.96 dB	0.917
Number of shots	3					
Huang et al. [37]	31.53 dB	0.895	30.59 dB	0.883	23.62 dB	0.788
Le Pendu et al. [29]	35.47 dB	0.947	36.83 dB	0.953	29.15 dB	0.900
Gao et al. [11]	37.21 dB	0.950	36.38 dB	0.947	28.10 dB	0.872
Le Bon et al. [31]	40.79 dB	0.974	38.48 dB	0.964	30.75 dB	0.920
Joint optimization FDL	41.83 dB	0.978	39.39 dB	0.969	31.87 dB	0.933

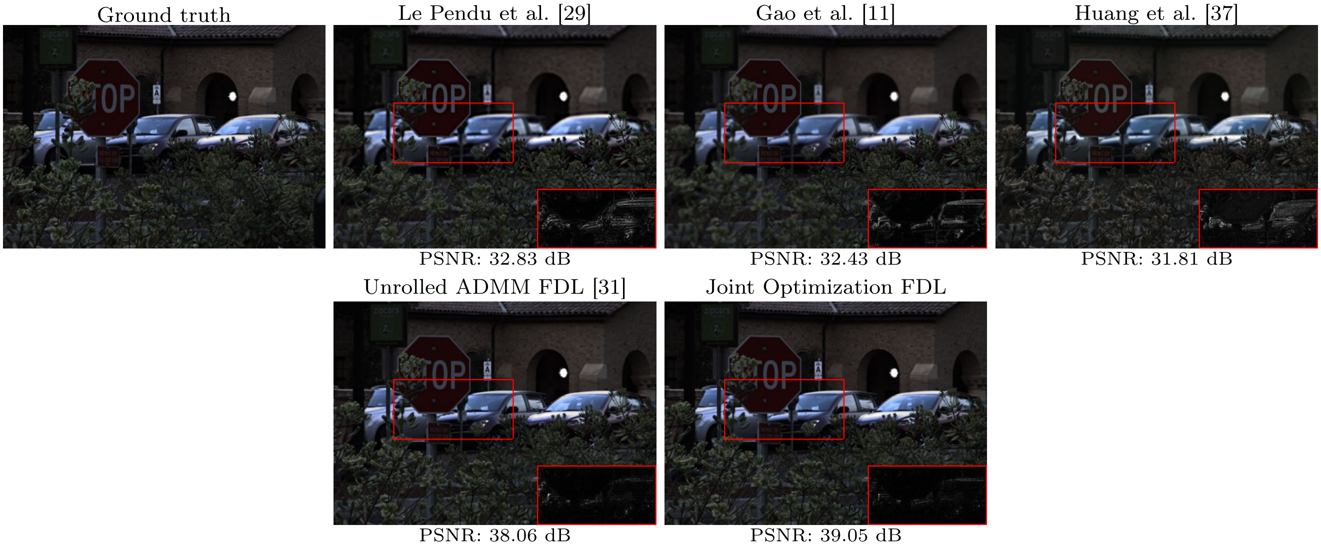


FIGURE 3. Reconstructed central views for the light field *occlusions_26_eslf* from the Stanford dataset [42] using 2-shots, with the different evaluated methods. A portion of the error map is highlighted.

TABLE 2. Ablation study: average PSNR for light field reconstruction.

Datasets	Kalantari [39]		Stanford [40]		Linköping [41]	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Number of shots	2					
FDL + view synthesis	38.75 dB	0.966	36.56 dB	0.954	27.91 dB	0.890
Unrolled ADMM FDL	39.82 dB	0.968	37.32 dB	0.955	29.22 dB	0.902
Joint optimization FDL	40.93 dB	0.974	38.35 dB	0.961	29.96 dB	0.917
Number of shots	3					
FDL + view synthesis	39.77 dB	0.971	38.08 dB	0.963	30.12 dB	0.910
Unrolled ADMM FDL	40.79 dB	0.974	38.48 dB	0.964	30.75 dB	0.920
Joint optimization FDL	41.83 dB	0.978	39.39 dB	0.969	31.87 dB	0.933

computation time for the synthesis of a single view with the FDL-based methods.

On one hand, the obtained computation time shows that the unrolled ADMM FDL optimization in [31] and in our proposed joint optimization method increases the computation time compared to the original FDL reconstruction algorithm presented in [29]. This difference in computation time is mostly due to the computation of the closed-form solution in Eq. (12) and to the application of the denoiser $\mathcal{D}$ in Eq. (15) for several iterations. However, the overall iterative reconstruction algorithm in the FDL domain stays faster to compute than the iterative reconstruction algorithm by Gao et al. [11].

On the other hand, the learned view synthesis presented in Section III-C increases the computation time of computing a single view from the optimized FDL. Indeed, while each view is computed by a simple shift-and-sum applied on the optimized FDL with the methods in [29] and [31],

TABLE 3. Algorithm complexity: average computation time (in seconds) (i) for the iterative reconstruction algorithms (ii) for the rendering of a single view.

	Reconstruction algorithm		View synthesis	
	CPU	GPU	CPU	GPU
Gao et al. [11]	443.27 s	-	-	-
Le Pendu et al. [29]	5.56 s	0.22 s	0.02 s	0.002 s
Le Bon et al. [31]	159.49 s	4.49 s	0.02 s	0.002 s
Joint optimization FDL	159.49 s	4.49 s	5.51 s	0.220 s

the synthesis network $\mathcal{S}$ in (20) is applied for each view to be reconstructed in our proposed method. Therefore, the computation time of rendering a dense light field is a limitation of the proposed method. However, it is important to notice that the view synthesis process can be parallelized to synthesize several views simultaneously, which permits to overcome this computational issue.

E. ABBLATION STUDY: THE LEARNED VIEW SYNTHESIS

In this section, we first study the importance of both the unrolled ADMM FDL optimization and the view synthesis network in the proposed end-to-end framework. We then propose to evaluate different approaches for the input of the network used in the view synthesis process.

1) END-TO-END FRAMEWORK

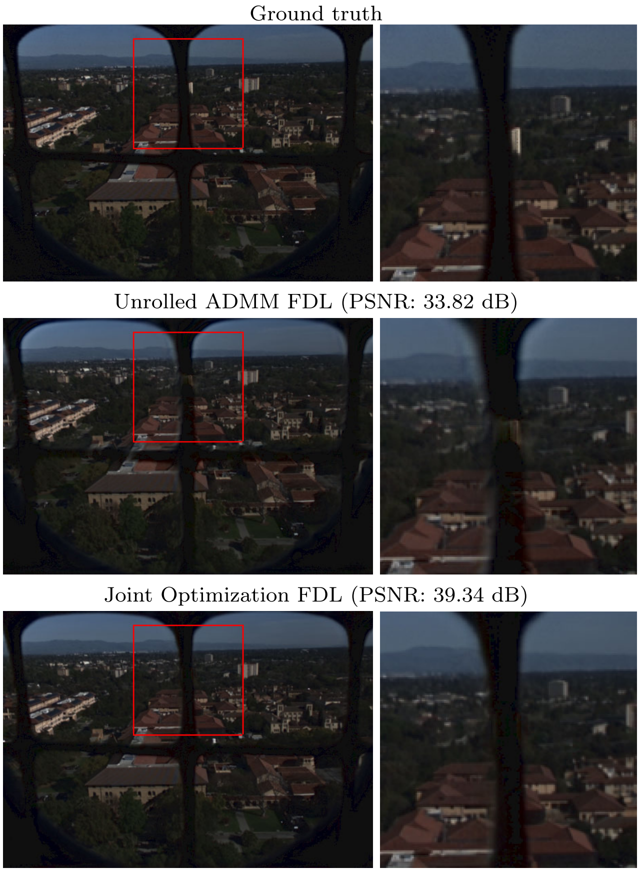
First of all, we propose to compare the light field reconstruction performances for different frameworks that consider different parts of the proposed joint optimization:

- FDL + view synthesis: this framework uses the FDL optimization proposed in [29], without any learned prior. The learned view synthesis process is trained to reconstruct views from the estimated FDL.
- Unrolled ADMM FDL: the unrolled ADMM FDL optimization presented in [31] without learning the view synthesis process.
- Joint optimization FDL: the proposed joint optimization FDL that considers both the unrolled ADMM FDL optimization and the learned view synthesis parts.

The reconstruction performances are listed in Table 2. As shown in the table, having both the unrolled optimization and the view synthesis network offers the best performances by a large margin.

To further study this improvement over our previous work [31], we propose to empirically verify that the learned view synthesis process is able to reduce the occlusion artifacts not well-handled by the FDL model, as explained theoretically in Sections III-A and III-C. Since the FDL are optimized from focal stack measurements captured at angular coordinates $u_0 = 0$, the optimized FDL are then well-defined to reconstruct the central view for any type of scene, while artifacts are expected on the other views in certain areas, e.g. transparency in occluded regions [29], [30]. Therefore, we expect the joint optimization method to reduce these artifacts in order to improve the PSNR of the reconstructed views that are far from the central view.

Fig. 4 illustrates a transparency artifact occurring in an occluded region with the unrolled ADMM FDL method. We can visually see that the proposed joint optimization method significantly reduces this artifact. In Fig. 5, we computed the average PSNR gain over the Kalantari testing dataset [39] with the end-to-end approach over the unrolled ADMM FDL optimization for several views with different angular coordinates. As shown in Fig. 5, the proposed framework always improves the reconstruction quality compared to the unrolled ADMM FDL method, especially for the views that are distant from the central view with an average gain of over 1 dB.


Ground truth

Unrolled ADMM FDL (PSNR: 33.82 dB)

Joint Optimization FDL (PSNR: 39.34 dB)

FIGURE 4. Example of an occluded region in the light field *occlusion_36_eslf*. The middle row illustrates the transparency artifacts with the FDL model in occluded regions: a building is visible through the grid of a window. These artifacts are reduced in the last row thanks to the learned view synthesis block of the proposed joint optimization.

2) NETWORK INPUT

In this section, we propose a study of different approaches for the network input in the view synthesis process. To be able to reconstruct any view from the optimized FDL, the network needs an input which contains all the spatial information carried by the optimized FDL, but also angular information so that its output are specific and optimal for each view. A first approach is to concatenate the optimized FDL with additional

TABLE 4. Ablation study on view synthesis network input: average PSNR for light field reconstruction.

Coordinate channels	FDL shifted	Kalantari [39]		Stanford [40]		Linköping [41]	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
yes	no	40.77 dB	0.974	38.19 dB	0.960	29.67 dB	0.914
no	yes	40.82 dB	0.974	38.26 dB	0.960	29.93 dB	0.916
yes	yes	40.93 dB	0.974	38.35 dB	0.961	29.96 dB	0.917

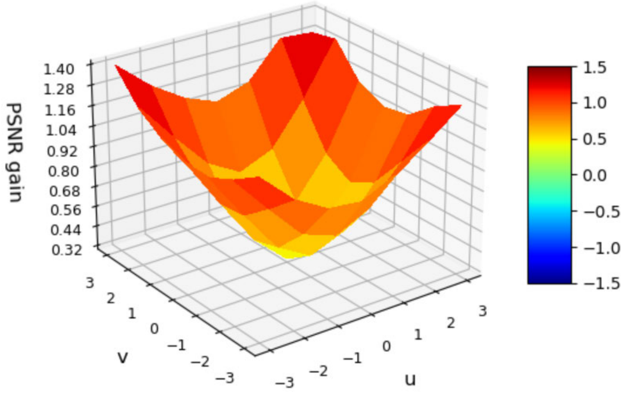


FIGURE 5. Mean PSNR gain over the test set of the Kalantari dataset [39] for different light field view coordinates with the proposed joint optimization framework compared to the unrolled ADMM FDL optimization using 3-shots focal stacks.

channels that contains the value of the angular coordinates of the view to be reconstructed. Another approach is to directly incorporate the angular information in the optimized FDL, i.e. by shifting the optimized FDL accordingly to the angular coordinates of the view to be reconstructed, following the view synthesis process of the FDL model in Eq. (18).

In order to select the best approach, we propose to compare the reconstruction performances with different network input configurations, using either additional coordinate channels or the shifted version of the FDL, or both at the same time. In our experiments, when using both approaches at the same time, we obtained better results by first training the joint optimization method by considering only the shifted FDL without any additional coordinate channels as network input, then fine-tune this pre-trained model by adding the coordinate channels in the network input. We listed the obtained results in Table 4 for 2-shots focal stacks. As shown in the table, the joint optimization method is able to efficiently reconstruct the light fields with both approaches. According to these results, both approaches are also complementary, giving the best results when using both at the same time.

V. CONCLUSION

In this paper, we have presented a method to reconstruct a light field from a set of focal stack images captured with a single traditional camera. A joint unrolled ADMM FDL optimization with a learned view synthesis network is presented to extend the Fourier Disparity Layer (FDL) representation of scenes to occluded and non-Lambertian scenes. The Alternating Direction Method of Multipliers (ADMM) optimization method is unrolled using a deep convolutional denoiser of FDL, where a closed-form solution of the proximal operator of the data-fit term is derived. Additionally, a deep network is trained to adapt the optimized FDL for each view to be reconstructed, in order to minimize the artifacts created with the generation of the views from the FDL model. Thanks to the capacity of deep networks to represent complex priors, the proposed approach significantly outperforms state-of-the-art methods for light field reconstruction from focal stacks with very few shots, with an average gain of about 1 dB of PSNR, on each considered dataset, in comparison with the best approaches.

REFERENCES

[1] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 1–15, 2015.

[2] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic aperture confocal imaging," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 825–834, Aug. 2004.

[3] J. Shi, X. Jiang, and C. Guillemot, "A framework for learning depth from a flexible subset of dense and sparse light field views," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5867–5880, Dec. 2019.

[4] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005.

[5] M. Levoy and P. Hanrahan, "Light field rendering," in *Seminal Graphics Papers: Pushing the Boundaries*, vol. 2. New York, NY, USA: ACM, 2023, pp. 441–452.

[6] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep.*, vol. 2, no. 11, 2005.

[7] S. D. Babacan, R. Ansorge, M. Luessi, R. Molina, and A. K. Katsaggelos, "Compressive sensing of light fields," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2337–2340.

[8] E. Miandji, J. Unger, and C. Guillemot, "Multi-shot single sensor light field camera using a color coded mask," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 226–230.

[9] H.-N. Nguyen, E. Miandji, and C. Guillemot, "Multi-mask camera model for compressed acquisition of light fields," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 191–208, 2021.

[10] J. R. Alonso, A. Fernández, and J. A. Ferrari, "Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack," *Appl. Opt.*, vol. 55, no. 9, pp. 2380–2386, Mar. 2016.

[11] S. Gao, G. Qu, M. Sjöström, and Y. Liu, "A TV regularisation sparse light field reconstruction model based on guided-filtering," *Signal Process., Image Commun.*, vol. 109, Nov. 2022, Art. no. 116852.

[12] A. Levin and Y. Weiss, "User assisted separation of reflections from a single image using a sparsity prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1647–1654, Sep. 2007.

[13] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, "SoftCuts: A soft edge smoothness prior for color image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 969–981, May 2009.

[14] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, Dec. 2016.

[15] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.

[16] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 399–406.

[17] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, "Unrolled optimization with deep priors," 2017, *arXiv:1705.08041*.

[18] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2774–2781.

[19] J. Sun, H. Li, and Z. Xu, "Deep ADMM-net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[20] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[21] D. Gilton, G. Ongie, and R. Willett, "Neumann networks for linear inverse problems in imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 328–343, 2020.

[22] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1123–1133, 2021.

[23] X. Tao, H. Zhou, and Y. Chen, "Image restoration based on end-to-end unrolled network," *Photonics*, vol. 8, p. 376, Jun. 2021.

[24] G. Le Guludec and C. Guillemot, "Deep unrolling for light field compressed acquisition using coded masks," *IEEE Access*, vol. 10, pp. 42933–42948, 2022.

[25] K. Takahashi, Y. Kobayashi, and T. Fujii, "From focal stack to tensor light-field display," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4571–4584, Sep. 2018.

[26] C. Liu, J. Qiu, and M. Jiang, "Light field reconstruction from projection modeling of focal stack," *Opt. Exp.*, vol. 25, no. 10, pp. 11377–11388, 2017.

[27] X. Yin, G. Wang, W. Li, and Q. Liao, "Iteratively reconstructing 4D light fields from focal stacks," *Appl. Opt.*, vol. 55, no. 30, pp. 8457–8463, 2016.

[28] S. Gao and G. Qu, "Filter-based Landweber iterative method for reconstructing the light field," *IEEE Access*, vol. 8, pp. 138340–138349, 2020.

[29] 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.

[30] T. Herfet, K. Chelli, and M. Le Pendu, "Light field representation: The dimensions in light fields," in *Immersive Video Technologies*. Amsterdam, The Netherlands: Elsevier, 2023, pp. 173–199.

[31] B. L. Bon, M. Le Pendu, and C. Guillemot, "Unrolled Fourier disparity layer optimization for scene reconstruction from few-shots focal stacks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[33] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Seminal Graphics Papers: Pushing the Boundaries*. New York, NY, USA: ACM, 2023, vol. 2, pp. 453–464.

[34] M.-B. Lien, C.-H. Liu, I. Y. Chun, S. Ravishankar, H. Nien, M. Zhou, J. A. Fessler, Z. Zhong, and T. B. Norris, "Ranging and light field imaging with transparent photodetectors," *Nature Photon.*, vol. 14, no. 3, pp. 143–148, Mar. 2020.

[35] C. J. Blocker, Y. Chun, and J. A. Fessler, "Low-rank plus sparse tensor models for light-field reconstruction from focal stack data," in *Proc. IEEE 13th Image, Video, Multidimensional Signal Process. Workshop (IVMSP)*, Jul. 2018, pp. 1–5.

[36] M. H. Kamal, B. Heshmat, R. Raskar, P. Vandergheynst, and G. Wetzstein, "Tensor low-rank and sparse light field photography," *Comput. Vis. Image Understand.*, vol. 145, pp. 172–181, Apr. 2016.

[37] Z. Huang, J. A. Fessler, T. B. Norris, and I. Y. Chun, "Light-field reconstruction and depth estimation from focal stack images using convolutional neural networks," in *Proc. IEEE 45th Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 8648–8652.

[38] S. Zheng, Y. Liu, Z. Meng, M. Qiao, Z. Tong, X. Yang, S. Han, and X. Yuan, "Deep plug-and-play priors for spectral snapshot compressive imaging," *Photon. Res.*, vol. 9, no. 2, p. B18, 2021.

[39] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–10, Nov. 2016.

[40] R. Shah, G. Wetzstein, A. S. Raj, and M. Lowney, "Stanford lytro light field archive," Stanford Comput. Imag. Lab., Stanford Univ., 2016. [Online]. Available: http://lightfields.stanford.edu/LF2016.html

[41] E. Miandji, H.-N. Nguyen, S. Hajisharif, J. Unger, and C. Guillemot, "Compressive HDR light field imaging using a single multi-ISO sensor," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1369–1384, 2021.

[42] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Proc. 13th Asian Conf. Comput. Vis.*, 2016, pp. 19–34.

BRANDON LE BON (Member, IEEE) received the Engineering degree in computer science and digital imaging from École Supérieur d'Ingénieur de Rennes (ESIR), in 2020. He is currently pursuing the Ph.D. degree with INRIA Rennes. His research interests include 2D image and light field processing and methods to solve image inverse problems using deep learning techniques.

MIKAËL LE PENDU received the Engineering degree from 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, Rennes, and addressed the compression of high-dynamic range video content. Then, he has pursued several postdoctoral positions with INRIA Rennes and Trinity College Dublin, where he studied light field image processing and deep learning techniques for solving inverse problems in image processing. Since 2022, he has been a Researcher with InterDigital Rennes, working on 2D video coding. His research interests include 2D video coding, deep learning, and high dynamic range and light field imaging, covering the full processing chain from capture to compression, including editing tasks.

CHRISTINE GUILLEMOT (Fellow, IEEE) received the Ph.D. degree from École Nationale Supérieure des Télécommunications (ENST), Paris, and the Habilitation degree in research direction from the University of Rennes. From 1985 to 1997, she was with France Télécom, where she was involved in various projects in the area of image and video coding for TV, HDTV, and multimedia. From 1990 to 1991, she was a Visiting Scientist with Bellcore, NJ, USA. She is currently the Director of Research with Institut National de Recherche en Informatique et en Automatique (INRIA) and the Head of a research team dealing with image and video modeling, processing, coding, and communication. Her research interests include signal and image processing, and in particular, 2D and 3D image and video processing for various problems, such as compression, and inverse problems, such as restoration, super-resolution, and inpainting. She has served as a Senior Member of the Editorial Board for IEEE Journal of Selected Topics in Signal Processing, from 2013 to 2015. She has served as an Associate Editor for IEEE Transactions on Image Processing, from 2000 to 2003, and from 2014 to 2016, IEEE Transactions on Circuits and Systems for Video Technology, from 2004 to 2006, and IEEE Transactions on Signal Processing, from 2007 to 2009. She was a Senior Area Editor of IEEE Transactions on Image Processing (2016–2020).