



HAL
open science

Color and angular reconstruction of light fields from incomplete-color coded projections

Hoai-Nam Nguyen, Christine Guillemot

► **To cite this version:**

Hoai-Nam Nguyen, Christine Guillemot. Color and angular reconstruction of light fields from incomplete-color coded projections. IEEE International Conference on Acoustics, Speech, and Signal Processing, May 2020, Barcelone, Spain. hal-02479455

HAL Id: hal-02479455

<https://hal.science/hal-02479455v1>

Submitted on 14 Feb 2020

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

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

COLOR AND ANGULAR RECONSTRUCTION OF LIGHT FIELDS FROM INCOMPLETE-COLOR CODED PROJECTIONS

Hoai-Nam Nguyen and Christine Guillemot

Inria Rennes – Bretagne Atlantique
35042 Rennes Cedex

ABSTRACT

We present a simple variational approach for reconstructing color light fields (LFs) in the compressed sensing (CS) framework with very low sampling ratio, using both coded masks and color filter arrays (CFAs). A coded mask is placed in front of the camera sensor to optically modulate incoming rays, while a CFA is assumed to be implemented at the sensor level to compress color information. Hence, the LF coded projections, operated by a combination of the coded mask and the CFA, measure incomplete color samples with a three-times-lower sampling ratio than reference methods that assume full color (channel-by-channel) acquisition. We then derive adaptive algorithms to directly reconstruct the light field from raw sensor measurements by minimizing a convex energy composed of two terms. The first one is the data fidelity term which takes into account the use of CFAs in the imaging model, and the second one is a regularization term which favors the sparse representation of light fields in a specific transform domain. Experimental results show that the proposed approach produces a better reconstruction both in terms of visual quality and quantitative performance when compared to reference reconstruction methods that implicitly assume prior color interpolation of coded projections.

Index Terms— light field, compressed sensing, demosaicing, variational algorithm

1. INTRODUCTION

A light field can be seen as a sampling of the plenoptic function which describes the radiance of light rays emitted by the scene along any direction, at any times and for every wavelength [1]. LFs are commonly represented as 4D functions with spatial and angular coordinates [2, 3], and can be regarded as collections of 2D images taken from different viewpoints. A number of acquisition devices have been developed to capture 4D LF. Existing devices range from bulky systems (such as single cameras mounted on moving gantries [2] or camera arrays [4, 5]) to handheld camera architectures -e.g. plenoptic 1.0 and 2.0 cameras using lenslet arrays to spatially multiplex angular (viewpoint) information [6, 7, 8]. It is a very effective way to acquire multiple views; however, it also induces some trade-off between the spatial and angular resolution of the captured LFs and reduces the spatial resolution by order of magnitude compared to a raw sensor image.

To overcome this limitation, mask-based camera models have emerged as alternative solutions to lenslet-based models. In [9], Veeraraghavan *et al.* use an attenuation mask with sum-of-sinusoid (SoS) patterns located between the camera aperture and the sensor to encode the angular spectrum of the 4D LFs into the spatial spectrum of 2D sensor images. The LF reconstruction is then performed in frequency domain using the Fourier slice theorem. A similar SoS-based approach is proposed in [10], in which the authors introduce a dual-mask design: one mask is mounted on the aperture, while the second one is placed before the sensor to improve the spatial resolution of the reconstructed LFs when compared to [9].

Over the past years, random coded masks have been considered in CS-based designs [11]. One of the first CS camera models for LF acquisition is reported in [12] which uses a monochrome mask placed at the aperture for optical modulation purposes. The acquisition scheme is also equipped with a reconstruction algorithm based on the total variation (TV) of angular images. Similarly, the authors of [13] proposed to place a (monochrome) mask at a small distance from the sensor to get better incoherent measurements, and used sparse representations on overcomplete dictionaries to reconstruct the original LFs. This dictionary-based approach is further extended in [14] using random color masks instead of monochrome ones. Recently, several learning-based LF reconstruction methods have been introduced [15, 16] for both monochrome and color masks. They provide in general better results with lower processing time compared to existing dictionary-based approaches [13, 14], at the cost of long training time and expensive computational resources.

In this work, we focus on the reconstruction of LFs from 2D coded projections which are induced by random masks, while accounting for the presence of a CFA (e.g. the well-known RGB Bayer pattern) within the 2D sensor. This kind of sensors is indeed commonly implemented in standard consumer cameras as those used in hardware experiments in [12, 13], implying that captured images are not full color but only incomplete color samples. The authors of [12]-[15] however reconstruct original LFs from full color images. This is possible only if the captured color-sampled image has first been demosaiced, e.g. using gradient corrected color interpolation [17], or if each color channel is acquired separately. While the first method may have severe impacts on the reconstruction quality due to color interpolation errors, the second one requires camera modifications (the sensor and/or the optical setup in this case) which are sometimes not feasible.

To tackle this problem, we adapt the imaging equation in

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

[12]-[16] by taking into account the use of CFAs in the image formation model. The sensing matrix can therefore be seen as the product of two matrices: one matrix represents the coded mask while the second matrix represents the implemented CFA. We show that this requires very little changes of the reconstruction algorithms since only the sensing operator is modified when compared to a full color acquisition scheme. Moreover, we introduce a new LF prior by extending the famous TV regularization [18] for 4D color data, and propose a proximal-splitting-based algorithm for LF reconstruction using this TV-based prior. The proposed algorithm hence performs at the same time full color restoration of the light ray measurements (i.e. some form of demosaicing) and LF viewpoint reconstruction.

2. LIGHT FIELD ACQUISITION

We adopt the two-plane parameterization [2, 3] for the in-camera light field $\mathcal{L}(x, u, \lambda)$ which describes the intensity in the wavelength λ along a ray passing through the points $\mathbf{x} = (x, y)$ and $\mathbf{u} = (u, v)$ located on the sensor and aperture planes respectively. Let D denote the distance between these two planes. For simplicity, we also assume that \mathcal{L} is zero outside the physical bounds of the sensor $\Omega \subset \mathbb{R}^2$ and the aperture $\Theta \subset \mathbb{R}^2$. If we insert a mask $M(\xi, \lambda)$ at a small distance d from the sensor and when the camera architecture allows color-by-color acquisition, the imaging equation for a pixel $\mathbf{x} \in \Omega$, in a wavelength λ , can be written as [13, 14]:

$$\mathcal{I}_\lambda(\mathbf{x}) = \int_{\Theta} M(\xi_{\mathbf{x}, \mathbf{u}, d}, \lambda) \mathcal{L}(\mathbf{x}, \mathbf{u}, \lambda) d\mathbf{u}, \quad (1)$$

where $\xi_{\mathbf{x}, \mathbf{u}, d} = \mathbf{u} + \frac{D-d}{D}(\mathbf{x} - \mathbf{u})$. Each image \mathcal{I}_λ can be seen as a compressed version of the set of angular images $\{\mathcal{L}(\cdot, \mathbf{u}, \lambda)\}_{\mathbf{u} \in \Theta}$. In discrete form, assuming that $n = n_x n_y$ is the number of sensor pixels and $\nu = \nu_u \nu_v$ is the number of views, we denote by $\mathcal{L}_j^c \in \mathbb{R}^n$ (with $c \in \mathcal{C} = \{R, G, B\}$) the color channel corresponding to λ_c of the j -th angular image (with $j = l_1 + \nu_u(l_2 - 1)$, for $1 \leq l_1 \leq \nu_u$ and $1 \leq l_2 \leq \nu_v$). The column vector $\mathcal{L} \in \mathbb{R}^{3n\nu}$ which represents the discrete version of \mathcal{L} can be arranged in the following way:

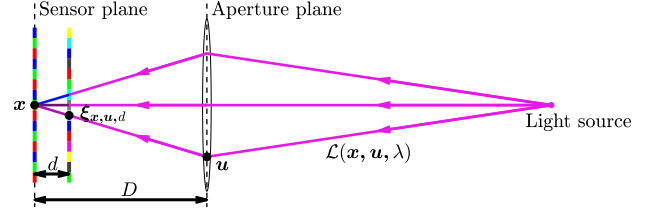
$$\mathcal{L} = [(\mathcal{L}^R)^\top \quad (\mathcal{L}^G)^\top \quad (\mathcal{L}^B)^\top]^\top, \quad (2)$$

where $(\cdot)^\top$ denotes the matrix transpose operator and each $\mathcal{L}^c = [(\mathcal{L}_1^c)^\top \quad (\mathcal{L}_2^c)^\top \quad \dots \quad (\mathcal{L}_\nu^c)^\top]^\top \in \mathbb{R}^{n\nu}$ is a color channel of \mathcal{L} corresponding to the wavelength λ_c . Using this vectorization, the discrete form of the color-by-color acquisition scheme (Eq. 1) can be expressed as a matrix-vector multiplication as follows:

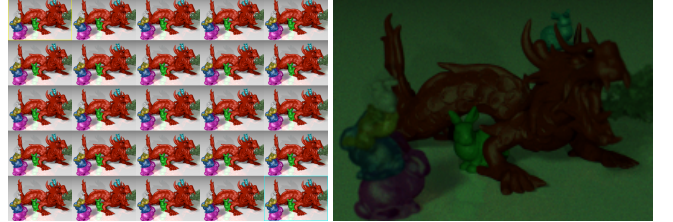
$$\underbrace{\begin{bmatrix} \mathcal{I}^R \\ \mathcal{I}^G \\ \mathcal{I}^B \end{bmatrix}}_{\mathcal{I}} = \underbrace{\begin{bmatrix} M^R & 0 & 0 \\ 0 & M^G & 0 \\ 0 & 0 & M^B \end{bmatrix}}_M \underbrace{\begin{bmatrix} \mathcal{L}^R \\ \mathcal{L}^G \\ \mathcal{L}^B \end{bmatrix}}_{\mathcal{L}}, \quad (3)$$

where each sub-matrix $M^c = [M_1^c \quad M_2^c \quad \dots \quad M_\nu^c]$ and each $M_j^c \in \mathbb{R}^{n \times n}$ is a sparse matrix containing the coefficients of the coded mask M on its diagonal.

To improve the compression rate, Nabati *et al.* proposed to compress color information by using monochrome sensors



(a) Acquisition scheme using coded mask and CFA



(b) Original LF

(c) Incomplete-color coded projection



(d) Full RGB coded projection

(e) Color interpolation of (c) using [17]

Fig. 1: Mask-based compressed acquisition of LFs. (c) Incomplete color coded projection (compression rate: $1/(3 \times 5 \times 5)$). (d) Full RGB (color-by-color) acquisition (compression rate: $1/(5 \times 5)$). (e) Color interpolation (demosaicing) of incomplete-color coded projection looks slightly blurred compared to the true (full RGB) version.

(without any CFA) as [16]:

$$\bar{I} = \sum_{c \in \mathcal{C}} \mathcal{I}^c = \underbrace{\begin{bmatrix} M^R & M^G & M^B \end{bmatrix}}_M \mathcal{L}, \quad (4)$$

which consists in squeezing M by removing all its zero sub-matrices, implying that the captured image \bar{I} is simply the sum of the three channels of \mathcal{I} . For the LF reconstruction task, the authors of [16] have considered a fully connected network (FCN) which does not only estimate the viewpoint parallax but also re-color the scene from gray-scale acquisitions. Despite promising results, the re-colorization may fail in some cases and only work for random RGBW patterns as used in [16].

Let us now consider the acquisition with color sensors (i.e. sensors with overlaid CFA patterns) instead of monochrome sensors. Accordingly, the corresponding imaging equation is expressed as:

$$I = \sum_{c \in \mathcal{C}} F^c \mathcal{I}^c = \underbrace{\begin{bmatrix} F^R M^R & F^G M^G & F^B M^B \end{bmatrix}}_{\Phi} \mathcal{L}, \quad (5)$$

where each $F^c \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal contains values in $[0, 1]$ representing the CFA pattern. Note that one can also factorize the resulting sensing matrix Φ as:

$$\Phi = F M = \begin{bmatrix} F^R \\ F^G \\ F^B \end{bmatrix}^\top \begin{bmatrix} M^R & 0 & 0 \\ 0 & M^G & 0 \\ 0 & 0 & M^B \end{bmatrix}. \quad (6)$$

We remark that I is nothing else than a sampled-color version of \mathcal{I} . Also, it is easy to check that the images I and \bar{I} have the same dimensions, and thus the two acquisition schemes (Eq. 4) and (Eq. 5) achieve the same compression rate. But unlike (Eq. 4) which does not preserve color information, the proposed acquisition scheme (Eq. 5) allows to retrieve the sampled color component of each pixel of I using the predefined CFA pattern. In our opinion, this allows to design simpler algorithms for color LF reconstruction. On the down-side, the use of CFAs also reduces the amount of incoming lights received by the sensor due to color filtering effect, implying lower light-transmission rate. However, this approach represents a cheap and efficient way to capture color LFs while having three-time-higher compression rate compared with full RGB (color-by-color) acquisitions.

3. RECONSTRUCTION ALGORITHMS

3.1. Dictionary-based approach

The CS theory [11] relies on the assumption that the signal is sparse (or *compressible*) in some transform domains like wavelets, DCT, or even dictionaries learned from large datasets. In the context of compressive LF acquisition using coded masks, the original LF can be restored by solving a basis pursuit denoising [19] (BPDN) problem given an overcomplete dictionary as in [13, 14]. Let us denote by $\mathcal{D} = [\mathcal{D}^1 \ \mathcal{D}^2 \ \dots \ \mathcal{D}^{n_d}] \in \mathbb{R}^{3n\nu \times n_d}$ the given dictionary (where each $\mathcal{D}^k \in \mathbb{R}^{3n\nu}$ is a color 4D LF atom), and by $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_{n_d}]^T$ a sparse vector such that $\mathcal{D}\alpha$ approximates \mathcal{L} ($\mathcal{D}\alpha \simeq \mathcal{L}$). The LF reconstruction problem amounts therefore to minimize the following convex energy:

$$E_{\mathcal{D}}(\alpha) = \frac{1}{2} \|\Phi\mathcal{D}\alpha - I\|_2^2 + \eta \|\alpha\|_1, \quad (7)$$

which can be efficiently solved by many existing algorithms (e.g. OMP [20], LARS [21], ISTA [22] etc.) with a small modification compared to [13, 14] (by considering $\Phi = FM$ instead of $\Phi = M$). In practice, due to the large size of LFs, dictionary-based reconstruction is performed on small patches obtained by dividing LFs over the spatial domain while including all angular and color dimensions. Reconstructed patches are then aggregated to compute the full-size LFs by averaging pixels on overlapped regions.

3.2. Proposed TV-based algorithm

TV is commonly used to regularize epipolar plane images (EPIs) for LF denoising, super-resolution or depth estimation [23, 24]. To our knowledge, there is no published work that considers the 4D spatio-angular gradient of color LF data as an extension of TV. Let us define the spatio-angular gradient operator applied to the LF \mathcal{L} at the (discrete) spatio-angular coordinates $\mathbf{p} \in \mathcal{P} = \llbracket 1, n_x \rrbracket \times \llbracket 1, n_y \rrbracket \times \llbracket 1, \nu_u \rrbracket \times \llbracket 1, \nu_v \rrbracket$ as the following Jacobian matrix:

$$(\nabla\mathcal{L})_{\mathbf{p}} = \begin{bmatrix} (\partial_x \mathcal{L}^R)_{\mathbf{p}} & (\partial_x \mathcal{L}^G)_{\mathbf{p}} & (\partial_x \mathcal{L}^B)_{\mathbf{p}} \\ (\partial_y \mathcal{L}^R)_{\mathbf{p}} & (\partial_y \mathcal{L}^G)_{\mathbf{p}} & (\partial_y \mathcal{L}^B)_{\mathbf{p}} \\ (\partial_u \mathcal{L}^R)_{\mathbf{p}} & (\partial_u \mathcal{L}^G)_{\mathbf{p}} & (\partial_u \mathcal{L}^B)_{\mathbf{p}} \\ (\partial_v \mathcal{L}^R)_{\mathbf{p}} & (\partial_v \mathcal{L}^G)_{\mathbf{p}} & (\partial_v \mathcal{L}^B)_{\mathbf{p}} \end{bmatrix}, \quad (8)$$

where ∂_{dir} denotes the directional derivative with respect to $\text{dir} \in \mathcal{O} = \{x, y, u, v\}$. For simplicity, we use forward differ-

ence to implement ∂_{dir} , but other implementations of ∂_{dir} can be envisaged. The TV semi-norm of \mathcal{L} is then defined as the mixed L_1 -Frobenius norm as follows:

$$\text{TV}(\mathcal{L}) = \|\nabla\mathcal{L}\|_{1,\text{F}} := \sum_{\mathbf{p} \in \mathcal{P}} \sqrt{\sum_{\mathbf{c} \in \mathcal{C}} \sum_{\text{dir} \in \mathcal{O}} (\partial_{\text{dir}} \mathcal{L}^{\mathbf{c}})_{\mathbf{p}}^2}. \quad (9)$$

Considering this regularizer, we seek a solution of the following minimization problem:

$$\min_{\mathbf{L} \in \mathbb{R}^{3n\nu}} \frac{1}{2} \|\Phi\mathbf{L} - I\|_2^2 + \mu \text{TV}(\mathbf{L}). \quad (10)$$

A possible choice to solve (Eq. 10) is the proximal gradient method (also known as forward-backward splitting method [25, 22]) that requires computing prox_{TV} (the proximity operator of TV). It is however not easy to implement prox_{TV} in the case of high dimensional signals like color 4D LFs. Here, we consider instead the full-splitting approach [26] which allows to design an iterative algorithm using only ‘‘simple’’ operations as follows:

Choose the parameters $\gamma, \tau > 0$ and the initial estimates $\mathbf{L}^{(0)} \in \mathbb{R}^{3n\nu}$, $\mathbf{K}^{(0)} \in \mathbb{R}^{3n\nu \times 4 \times 3}$. Then iterate, for $\ell \geq 0$

$$\tilde{\mathbf{L}}^{(\ell+1)} = \Phi^*(\Phi\mathbf{L}^{(\ell)} - I) + \nabla^* \mathbf{K}^{(\ell)}, \quad (11)$$

$$\mathbf{L}^{(\ell+1)} = \mathbf{L}^{(\ell)} - \gamma \tilde{\mathbf{L}}^{(\ell+1)}, \quad (12)$$

$$\tilde{\mathbf{K}}^{(\ell+1)} = \nabla(2\mathbf{L}^{(\ell+1)} - \mathbf{L}^{(\ell)}), \quad (13)$$

$$\mathbf{K}^{(\ell+1)} = \text{prox}_{\tau(\mu\|\cdot\|_{1,\text{F}})^*}(\mathbf{K}^{(\ell)} + \tau \tilde{\mathbf{K}}^{(\ell+1)}), \quad (14)$$

where γ, τ are proximal parameters, $\mathbf{K}^{(\ell)}$ is auxiliary variable, Φ^* (resp. ∇^*) denotes the adjoint operator of Φ (resp. ∇) and $\text{prox}_{\tau(\mu\|\cdot\|_{1,\text{F}})^*}$ is defined as:

$$\left(\text{prox}_{\tau(\mu\|\cdot\|_{1,\text{F}})^*}[\mathbf{K}]\right)_{\mathbf{p}}^{\text{c,dir}} = \frac{\mathbf{K}_{\mathbf{p}}^{\text{c,dir}}}{\max\left(1, \frac{\|\mathbf{K}_{\mathbf{p}}\|_{\text{F}}}{\mu}\right)}, \quad (15)$$

for every $\mathbf{K}_{\mathbf{p}} = \{\mathbf{K}_{\mathbf{p}}^{\text{c,dir}}\}_{\mathbf{c} \in \mathcal{C}, \text{dir} \in \mathcal{O}} \in \mathbb{R}^{4 \times 3}$. In practice, we usually use random initialization for $\mathbf{L}^{(0)}$ and set $\mathbf{K}^{(0)} = 0$. According to [26], the sequence $\mathbf{L}^{(\ell)}$ converges to the solution of (Eq. 10) if $\gamma(\frac{1}{2}\|\Phi^*\Phi\| + \tau\|\nabla^*\nabla\|) < 1$. In contrast to dictionary-based approaches, the proposed algorithm is able to perform the reconstruction of full-size LFs while avoiding to divide them into small patches. It allows to obtain reconstructed LFs with homogeneous regions (i.e. without block artifacts that may happen when aggregating small patches as in [14]).

4. EXPERIMENTAL RESULTS

We evaluate the proposed reconstruction algorithm on synthetic LFs from the MIT Media Lab archive [13]. These LFs contain 5×5 views rendered from different scenes with various aperture sizes (from 20 to 50 mm) and each angular image has 840×593 pixels. The acquisition is performed using random color masks of [14] where the mask pixel values are drawn following a uniform distribution on $[0, 1]$. We assume that the CFA is the Bayer RGB pattern and no noise is added. An example of acquisitions is illustrated in Fig. 1c showing a coded projection composed of red, green and blue pixels. Its demosaiced version is shown in Fig. 1e which is visually similar to the captured image obtained by a color-by-color acquisition scheme (see Fig. 1d). However, it tends to over-

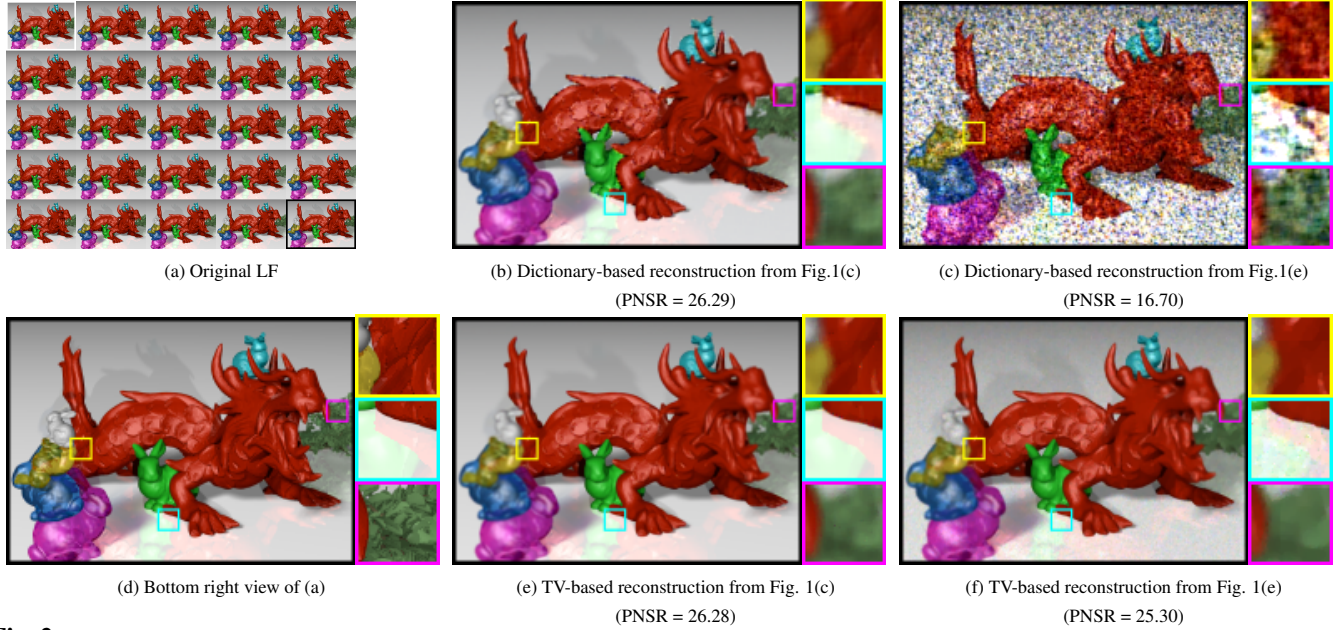


Fig. 2: Reconstruction results. TV-based reconstruction produces better visual quality than dictionary-based approach for both acquisition scenarios, i.e. incomplete-color coded projection (sub-figures (b) and (e), scenario (ii) in Table 1) and using pre-demaicing (sub-figures (c) and (f), scenario (iii) in Table 1). Reconstruction obtained from demosaiced coded projection are noisy mainly due to color interpolation errors.

Light field	TV			Dictionary
	(i)	(ii)	(iii)	(i)
dragon	26.83	26.28	25.30	26.32
dice	22.70	22.53	21.89	22.76
fish	21.95	21.65	21.27	21.83
messerschmitt	27.08	26.82	26.33	26.73

Table 1: PSNR obtained with the proposed TV-based algorithm on three acquisition scenarios: (i) color-by-color acquisition (Eq. 3); (ii) sampled-color acquisition (Eq. 5); (iii) sampled-color acquisition with pre-demaicing [17], and with dictionary-based methods [13, 14] on full RGB (color-by-color) acquisition.

smooth color transition and thus fails to recover high frequency color information.

Here, we focus on single-shot reconstruction (i.e. from one coded projection). Reconstruction algorithms are applied on three acquisition scenarios: (i) color-by-color acquisition (Eq. 3); (ii) sampled-color acquisition (Eq. 5); and (iii) sampled-color acquisition followed by demosaicing [17]. Note that the first scenario (considered in [12]-[15]) is not appropriate for standard consumer cameras due to major modifications of the camera architecture (e.g. using prisms to separate light colors and using monochrome sensor to capture each color component) and its compression ratio is three-time-higher than the second one. The third scenario, that assumes a first demosaicing of captured incomplete-color samples, introduces interpolation errors as depicted in Fig. 1e and can bias the reconstruction results. We compare our algorithm with state-of-the-art (SOTA) dictionary-based method [13, 14]. For the latter, we use an overcomplete dictionary with atoms of size $9 \times 9 \times 5 \times 5 \times 3 = 6075$ pixels (4D color patch) as in [14]. The dictionary is trained using the K-SVD toolbox [27]. The sparse code is then computed using LARS algorithm [21] and the LF is reconstructed by merging all patches with averaged overlapping pixels.

Fig. 2 shows reconstruction results obtained with the TV-based and dictionary-based algorithms. We can see that they

manage to recover the LF parallax. However, the dictionary-based approach often suffers from noisy reconstruction of patches containing object boundaries (see Fig. 2b and 2c), while the proposed TV-based method tends to produce results, with sharper edges and smooth homogeneous regions (see Fig. 2e and 2f), which are visually more pleasant. For the two approaches, direct reconstruction from raw sensor measurements (i.e. (ii) incomplete-color coded projections without prior color demosaicing) yields better results (visual quality and PSNR) when compared to reconstruction results from demosaiced data (i.e. acquisition scenario (iii) pre-demaicing the captured color samples before LF reconstruction). Table 1 reports the best PSNR values obtained by the proposed algorithm with three acquisition scenarios compared to the results of SOTA method [13, 14] on color-by-color acquisition (i). Despite higher compression rate, the duo of sampled-color acquisition scheme (ii) and TV-based reconstruction produces PSNR scores which are comparable, or even better than scenario (i) combined with SOTA dictionary-based reconstruction. In terms of processing time, the reconstruction of a $9 \times 9 \times 5 \times 5 \times 3$ patch for 10000 iterations takes only 4.5 seconds in average with our algorithm, while the used implementation of SOTA algorithms needs more than 7.5 seconds for the same setting.

5. CONCLUSION

We have presented a TV-based algorithm for LF reconstruction from compressed acquisition measurements using coded masks and CFAs. Our approach produces competitive results, both in terms of visual quality and PSNR performance, with higher compression rate and lower processing time, when compared to state-of-the-art dictionary-based approaches for mask-based LF cameras. Moreover, the proposed TV regularizer can be utilized in LF processing applications such as denoising, inpainting or super-resolution.

References

- [1] E. H. Adelson and J. R. Bergen, “The plenoptic function and the elements of early vision,” *Comput. Models Visual Process.*, MIT Press, pp. 3–20, 1991.
- [2] M. Levoy and P. Hanrahan, “Light field rendering,” in *23rd Annual Conf. Comput. Graphics Interactive Tech. (SIGGRAPH '96)*, 1996, pp. 31–42, ACM.
- [3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *23rd Annual Conf. Comput. Graphics Interactive Tech. (SIGGRAPH '96)*, 1996, pp. 43–54.
- [4] J.C. Yang, M. Everett, C. Buehler, and L. McMillan, “A real-time distributed light field camera,” in *Eurographics Symp. Rendering (EGSR)*, P. Debevec and S. Gibson, Eds., 2002, pp. 77–86.
- [5] 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. Graphics (TOG)*, vol. 24, no. 3, pp. 765–776, 2005.
- [6] E. H. Adelson and J. Y. A. Wang, “Single lens stereo with a plenoptic camera,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, 1992.
- [7] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a handheld plenoptic camera,” *Tech. Report CTSR 2005-02, Stanford University*, 2005.
- [8] A. Lumsdaine and T. Georgiev, “The focused plenoptic camera,” in *IEEE Int. Conf. Comput. Photography (ICCP)*, 2009, pp. 1–8.
- [9] 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 Trans. Graphics (TOG)*, vol. 26, no. 3, 2007.
- [10] Z. Xu and E.Y. Lam, “A high-resolution lightfield camera with dual-mask design,” in *SPIE Conf. Image Reconstruction from Incomplete Data VII, SPIE Opt. Eng. + Appl.*, 2012, vol. 8500.
- [11] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
- [12] S. D. Babacan, R. Ansorge, M. Luessi, R. Molina, and A. K. Katsaggelos, “Compressive sensing of light fields,” in *16th IEEE Int. Conf. Image Process. (ICIP)*, 2009, pp. 2337–2340.
- [13] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, “Compressive light field photography using overcomplete dictionaries and optimized projections,” *ACM Trans. Graphics (Proc. of SIGGRAPH)*, vol. 32, no. 4, pp. 46:1–46:12, 2013.
- [14] E. Miandji, J. Unger, and C. Guillemot, “Multi-shot single sensor light field camera using a color coded mask,” in *European Signal Process. Conf. (EUSIPCO)*, 2018, pp. 226–230.
- [15] M. Gupta, A. Jauhari, K. Kulkarni, S. Jayasuriya, A. Molnar, and P. Turaga, “Compressive light field reconstructions using deep learning,” in *IEEE Conf. Comput. Vision Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1277–1286.
- [16] O. Nabati, D. Mendlovic, and R. Giryes, “Fast and accurate reconstruction of compressed color light field,” in *IEEE Int. Conf. Comput. Photography (ICCP)*, 2018, pp. 1–11.
- [17] H. S. Malvar, L.-W. He, and R. Cutler, “High-quality linear interpolation for demosaicing of bayer-patterned color images,” in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2004, vol. 3, pp. iii–485.
- [18] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.
- [19] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Scientific Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [20] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *27th Asilomar Conf. Signals Syst. Comput.*, 1993, vol. 1, pp. 40–44.
- [21] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, 2004.
- [22] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imag. Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [23] B. Goldluecke and S. Wanner, “The variational structure of disparity and regularization of 4d light fields,” in *IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, 2013, pp. 1003–1010.
- [24] N. B. Monteiro, J. P. Barreto, and J. Gaspar, “Dense lightfield disparity estimation using total variation regularization,” in *Image Anal. Recognition*, 2016, pp. 462–469, Springer International Publishing.
- [25] P. Combettes and V. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [26] L. Condat, “A generic proximal algorithm for convex optimization—Application to total variation minimization,” *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 985–989, 2014.
- [27] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.