Low-Rank plus Sparse Tensor Models for Light-field Reconstruction from Focal Stack Data

Cameron J. Blocker†, Il Yong Chun†, Jeffrey A. Fessler
Department of Electrical Engineering and Computer Science, The University of Michigan
Ann Arbor, MI 48019-2122 USA
Email: cblocker@umich.edu, iychun@umich.edu, fessler@umich.edu

***Abstract*—Hand-held light-field cameras have enabled new photographic features such as refocusing and perspective shifts in post-processing. These cameras have traditionally sampled the 4D light-field directly by multiplexing angular measurements with spatial measurements on a single photosensor. This requires an undesirable trade-off between spatial and angular resolution. Focal stack cameras record varying projections of the light-field by capturing a set of photographs at different sensor positions. Prior techniques for reconstructing the 4D light-field from focal stack measurements have required depth estimation or ignored the existence of occlusions in the scene.**

We present low-rank plus sparse models for the light-field and apply them to the problem of reconstructing from focal stack measurements. We explore regularizers based on low-rank tensor decompositions to better exploit the dimensionality of the data. We optimize our model with a block proximal gradient method using a majorizer that provides a convergence guarantee. Numerical experiments show a several dB improvement in PSNR over traditional reconstruction methods and improved accuracy of depth estimation from light-fields reconstructed by the proposed methods.

I. Introduction

The light-field models the intensities of all rays in free space, from a geometric optics perspective. In particular we can use a light-field to model all of the rays that enter a camera's aperture and terminate at the camera's photosensor, using a two-plane parameterization. Thus every ray in the camera is uniquely determined by the (u, v) coordinate where it passes through the aperture and the (x, y) coordinate where it meets the sensor plane. A light-field, once acquired, can then be used to simulate different focal settings by a simple rebinning of rays to the spatial locations they would have terminated.

Hand-held light-field cameras, such as those made by Lytro and Ratrix, acquire the 4D light-field by multiplexing angular coordinates with spatial coordinates using a microlens array. This configuration reduces the measured spatial resolution by a factor of the angular resolution, leading to an undesirable trade-off. Furthermore, visual inspection of 4D light-fields suggests the data is highly redundant: holding different angular coordinates fixed results in sub-aperture images of the scene with only slight variations in perspective.

†*These two authors contributed equally to this work.*
This work is supported in part by the Keck Foundation.

In [1], Ng showed that photographs taken at different focus settings are 2D slices of the 4D light-field in Fourier space, analogous to the Fourier slice-projection theorem. All such 2D slices make up a 3D manifold, termed the Focal Manifold. Levin and Durand used these results as a prior to reconstruct light-field data from a focal stack by restricting the reconstruction to the 3D Focal Manifold in 4D Fourier space [2]. They also showed that most of the energy of a light-field lies on the Focal Manifold, due to typical scenes being very Lambertian. As such, their reconstruction technique was prone to more error in non-Lambertian areas, such as at occlusion boundaries. Other techniques for reconstructing a light-field require a non-linear depth estimation process [3].

Low-rank plus sparse models have been applied to reconstructing image frames across time [4]. Typically the data is flattened into a 2D matrix before enforcing a low-rank matrix structure; tensors, a generalization of matrices to higher dimensions, allow us to enforce a low-rank structure in higher dimensions and avoid removing the inherent structure that exists between different dimensions. Adding a sparse component to the model can help it be robust to signal outliers and details that would otherwise require a high number of rank 1 components to represent.

Kamal *et al.* applied a low-rank plus sparse (L+S) tensor model to reconstruct 5D (spatial-angular-temporal data) light-fields from a coded-aperture acquisition [5], where the L+S model was applied to local patches of the light-field. Instead of using patches, this paper uses a regularizer that models the light-field's global structure as a low-rank plus sparse tensor $(\mathcal{L} + \mathcal{S})$ for the purpose of reconstructing from a focal stack of images. In addition, we provide an optimization algorithm with a convergence guarantee for this non-convex problem.

II. Model: Light-Field Imaging with Focal Stacks

Given a 4D light-field, $L_F(x, y, u, v)$ parameterized by an aperture and sensor planes separated by F, we can compute the digital photograph taken by integrating the rays across the aperture, D for each $\Delta_{\mathrm{x}} \times \Delta_{\mathrm{y}}$ pixel

$$I_F[m, n] = \iint \left(\iint_{(u,v)\in D} L_F(x, y, u, v) \, \mathrm{d}u \, \mathrm{d}v \right) \mathrm{rect}(x/\Delta_{\mathrm{x}} - m) \, \mathrm{rect}(y/\Delta_{\mathrm{y}} - n) \, \mathrm{d}x \, \mathrm{d}y. \quad (1)$$

978-1-5386-0951-4/18/$31.00

If we consider the sensor plane at a distance of κF for any scalar $\kappa \in (0, \infty)$, we can similarly find the photograph taken, $I_{\kappa F}[m, n]$, by integrating $L_{\kappa F}(x, y, u, v)$ which relates to our original parameterized light-field by a shearing operation [1]

$$L_{\kappa F}(x, y, u, v) = L_F(\kappa^{-1}x - (1 - \kappa^{-1})u, \\ \kappa^{-1}y - (1 - \kappa^{-1})v, u, v). \quad (2)$$

These light-field parameterizations are continuous and intractable for computation. Since we are interested in reconstructing a discrete light-field that approximates the continuous function, we will assume that $L_F(x, y, u, v)$ can be written as a sum of rectangular basis functions

$$L_F(x, y, u, v) \\ \approx \sum_m \sum_n \sum_k \sum_l L[m, n, k, l] \text{rect}(x/\Delta_\text{x} - m) \\ \text{rect}(y/\Delta_\text{y} - n)\, \text{rect}(u/\Delta_\text{u} - k)\, \text{rect}(v/\Delta_\text{v} - l)\,, \quad (3)$$

where we take the basis coefficients $L[m, n, k, l]$ as our discrete light-field. By plugging (3) and (2) into (1), we can find the digital photograph taken at any sensor distance κF by

$$I_{\kappa F}[m', n'] = (L * g)[m, n, k, l]\Big|_{m=\kappa^{-1}m', n=\kappa^{-1}n', k=0, l=0}, \quad (4)$$

with

$$g[m, n, k, l] = \{s * t\}(m\Delta_\text{x}, n\Delta_\text{y}, k\Delta_\text{u}, l\Delta_\text{v}) \\ s(x, y, u, v) = \text{rect}(x/\Delta_\text{x})\, \text{rect}(y/\Delta_\text{y}) \\ \text{rect}(u/\Delta_\text{u})\, \text{rect}(v/\Delta_\text{v}) \\ t(x, y, u, v) = \kappa \text{rect}\big(\kappa/\Delta_\text{x}\,(x - (1 - \kappa^{-1})u)\big) \\ \text{rect}\big(\kappa/\Delta_\text{y}\,(y - (1 - \kappa^{-1})v)\big)\,.$$

Thus the projection of the sheared, discretized light-field can be factored as a 4D convolution followed by a 2D slicing and magnification. In practice we compute a set of 2D convolutions across the 2D slice for a more memory efficient implementation with the same computational cost. We note that the slice indices for the discrete function in (4) are not necessarily integers in general, but in many contexts $\kappa^{-1} \approx 1$ and so we forego the interpolation step.

Traditionally, focal stacks have been measured by physically moving the imaging sensor and taking separate exposures across time, but recent advances in sensor technology have allowed for transparent photosensors [6]. A collection of these transparent photosensors at different depths will allow us to capture a focal stack in a single exposure, making a practical focal stack camera. We obtain a 3D set of measurements by stacking the 2D images taken at different values of κ.

III. Image Reconstruction: $\mathcal{L} + \mathcal{S}$

Reconstructing a 4D dataset from a 3D set of measurements is an extremely underdetermined problem. To overcome this limitation, we model a light-field as a low-rank plus sparse tensor. While matrices are two-way arrays, tensors generalize this idea to higher dimensions. This allows us to model higher-dimensional data without flattening the data to only two-dimensions that would possibly lose the structural relationship present between dimensions.

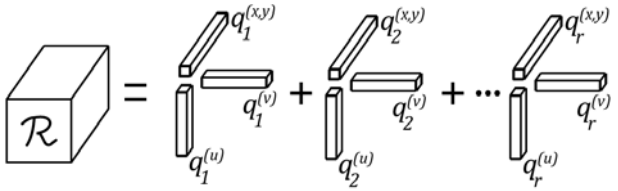

Fig. 1. Illustration of the CP Decomposition for a 3D tensor $\mathcal{R}$ as sum of r rank 1 tensors. Each rank 1 tensor is a tensor outer product of 3 vectors.

The idea of rank is a bit more nuanced for tensors then it is for matrices [7]. This work focuses on the Canonical Polyadic Decomposition (CPD) that decomposes a tensor into a set of r rank 1 tensors (unlike the SVD, these component vectors are not necessarily mutually orthogonal). The CPD decomposition can be used to find the best rank r approximation of a tensor, $\mathcal{R}$, as follows:

$$\mathcal{R} = \sum_{i=1}^{r} q_i^{(x,y)} \circ q_i^{(u)} \circ q_i^{(v)} := [\![Q^{(x,y)}, Q^{(u)}, Q^{(v)}]\!], \quad (5)$$

where we have used $\circ$ to represent a generalized outer product (see Fig. 1 for visual depiction) and $[\![\cdot]\!]$ represents a compact notation where $Q^{(z)}$ represent the concatenation of each $q_i^{(z)}$. In our low-rank tensor approximation, we unroll the spatial dimensions into a single dimension. While the data along the angular dimensions tends to be slowly varying – lending itself well to a low-rank approximation – we make no such assumptions on the spatial variations of the scene. By using a 3-way tensor model, we are able to achieve a lower rank than a full 4-way tensor model.

Inspired by the formulation for Robust PCA in [8] and its application to inverse problems in [4], we optimize an inverse problem where the low-rank plus sparse tensors are enforced as a regularizer with the following objective:

$$\arg\min_{\mathbf{x} \geq 0} \min_{Q^{(x,y)}, Q^{(u)}, Q^{(v)}, \mathcal{S}} \|\mathbf{y} - A\mathbf{x}\|_2^2 \\ + \alpha \left\| \mathbf{x} - \text{vec}\Big([\![Q^{(x,y)}, Q^{(u)}, Q^{(v)}]\!] + \mathcal{S} \Big) \right\|_2^2 \\ + \beta \|W \text{vec}(\mathcal{S})\|_1 \quad (6)$$

where A is our system model obtained by stacking the convolutions in (4) for each κ, W is a 4D unitary sparsifing transform (e.g., canonical, Fourier, wavelet bases, etc.), $\mathbf{x}$ and $\mathbf{y}$ represent our vectorized reconstruction and measurements, respectively, $\mathcal{S}$ represent the sparse tensor components, and each $Q^{(z)}$ represent the tensor component matrices that define our low-rank tensor $\mathcal{R}$.

Compared with traditional matrix-based robust PCA, finding a low-rank tensor approximation of a light-field is a non-convex problem due to the product of $Q^{(z)}$ terms. This

challenge has hindered algorithms in previous works from providing a convergence guarantee. Here we introduce an image reconstruction algorithm based on *Block Proximal Gradient method using a Majorizer* (BPG-M) [9], [10], [11] that guarantees (local) convergence under mild assumptions (e.g., continuity, lower-boundedness, etc. [9]).

IV. Convergent Image Reconstruction Algorithm: BPG-M

We optimize our objective (6) by BPG-M using five blocks x, $Q^{(x,y)}, Q^{(u)}, Q^{(v)}$ and $\mathcal{S}$, and note that (6) is a block multi-convex problem (i.e., (6) is convex in each block, when all other blocks are fixed). We update each block sequentially.

A. Image Update

For the image update step, we first construct a quadratic surrogate function using a diagonal majorization matrix, M, at a linearly extrapolated point, $\acute{\mathbf{x}}^{(i+1)}$. We then solve the corresponding majorized problem. Our proximal mapping problem is given by:

$$\underset{\mathbf{x}\,:\,\mathbf{x}\geq 0}{\arg\min}\ \|\mathbf{x}-\xi^{(i+1)}\|_M^2+\alpha\|\mathbf{x}-\mathrm{vec}(\mathcal{R}+\mathcal{S})\|_2^2 \qquad (7)$$

where

$$\begin{aligned}
\xi^{(i+1)} &= \acute{\mathbf{x}}^{(i+1)} - M^{-1}A^T(A\acute{\mathbf{x}}^{(i+1)}-\mathbf{y})\\
\acute{x}^{(i+1)} &= \mathbf{x}^{(i)} + w^{(i+1)}(\mathbf{x}^{(i)}-\mathbf{x}^{(i-1)})\\
w^{(i+1)} &= \frac{\theta^{(i)}-1}{\theta^{(i+1)}}\\
\theta^{(i+1)} &= \frac{1+\sqrt{1+4(\theta^{(i)})^2}}{2},
\end{aligned}$$

and $w^{(i+1)}$ is the increasing momentum-coefficient used in [9, (11)]. For our majorization matrix, we use $M = \mathrm{diag}\big(|A^T||A|1\big)$ [9], [10]. Our image update is given by a closed-form solution to (7):

$$\mathbf{x}^{(i+1)} = \Big[(M+\alpha I)^{-1}(M\xi^{(i+1)}+\alpha\mathrm{vec}(\mathcal{R}+\mathcal{S}))\Big]_{\geq 0},$$

where $[\cdot]_{\geq 0}$ thresholds negative values to zero.

B. Sparse Tensor Update

We update the sparse tensor by using the well-known soft-thresholding solution for an ℓ_1 proximal update:

$$\mathrm{vec}\Big(\mathcal{S}^{(i+1)}\Big) = W^T\mathcal{T}_{\beta/\alpha}(W(x-\mathrm{vec}(\mathcal{R})))$$

where $\mathcal{T}_a(z) := [z-a]_{\geq 0}\cdot\mathrm{sign}(z)$.

C. Low-Rank Tensor Update

For the low-rank tensor updates, we forgo the extrapolation and majorization and solve for each component exactly in a BCD fashion using least-squares, where each update is given by:

$$\begin{aligned}
&Q^{(x,y)(i+1)}\\
&= (X_{(x,y)} - S_{(x,y)})(Q^{(v)(i)}\odot Q^{(u)(i)})\\
&\quad ((Q^{(v)(i)})^TQ^{(v)(i)}\cdot\!*(Q^{(u)(i)})^TQ^{(u)(i)})^{\dagger},\\
&Q^{(u)(i+1)}\\
&= (X_{(u)} - S_{(u)})(Q^{(v)(i)}\odot Q^{(x,y)(i+1)})\\
&\quad ((Q^{(v)(i)})^TQ^{(v)(i)}\cdot\!*(Q^{(x,y)(i+1)})^TQ^{(x,y)(i+1)})^{\dagger},\\
&Q^{(v)(i+1)}\\
&= (X_{(v)} - S_{(v)})(Q^{(u)(i+1)}\odot Q^{(x,y)(i+1)})\\
&\quad ((Q^{(u)(i+1)})^TQ^{(u)(i+1)}\cdot\!*(Q^{(x,y)(i+1)})^TQ^{(x,y)(i+1)})^{\dagger},
\end{aligned}$$

where $X_{(z)}$ and $S_{(z)}$ represents a matrix created by unfolding our reconstructed tensor and sparse tensor along the z dimension, $\odot$ represents the Khatri-Rao product or "column-wise Kronecker product" and $\cdot\!*$ represents the Hadamard or elementwise product. (See [7] for additional information).

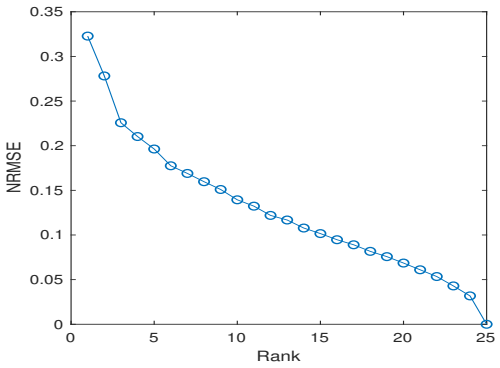

Fig. 2. Tensor approximation accuracy vs tensor rank for the two disc dataset. The true tensor is full rank even in the absence of noise.

V. Numerical Experiments

A. Setup: Imaging and Image Reconstruction

We tested our model on a simulated scene of two target discs at 1 and 2 meters from the camera (see Fig. 3(a-True)). From our simulated scene, we generated five noiseless 151×151 images at different focus settings through a 50mm focal length lens, and we reconstructed a light-field with 5×5 angular views. The focal planes were placed at regular intervals in the scene.

For comparison, we reconstructed the light-field using several general methods, including backprojection (BP) and 4D edge-preserving (EP) regularization. For EP regularized reconstruction, we used a 4D first-order finite difference regularizer with hyperbola penalty, i.e., $\varphi(t) := \delta^2(\sqrt{1+(t/\delta)^2}-1)$ ($\delta = 10^{-2}$). We chose the regularization parameter (balancing the data fitting term and the regularizer) as 10^{-7}. For the proposed $\mathcal{L}+\mathcal{S}$ reconstruction method, we chose $W = I$, and finely tuned the regularization parameters α, β to achieve good image quality – $\alpha = 10^{-9}$ and $\beta = 10^{-10}\times\alpha$. For the tensor rank, we chose $r = 20$. Fig. 2 shows the accuracy vs rank of a low-rank tensor approximation of the true light-field.

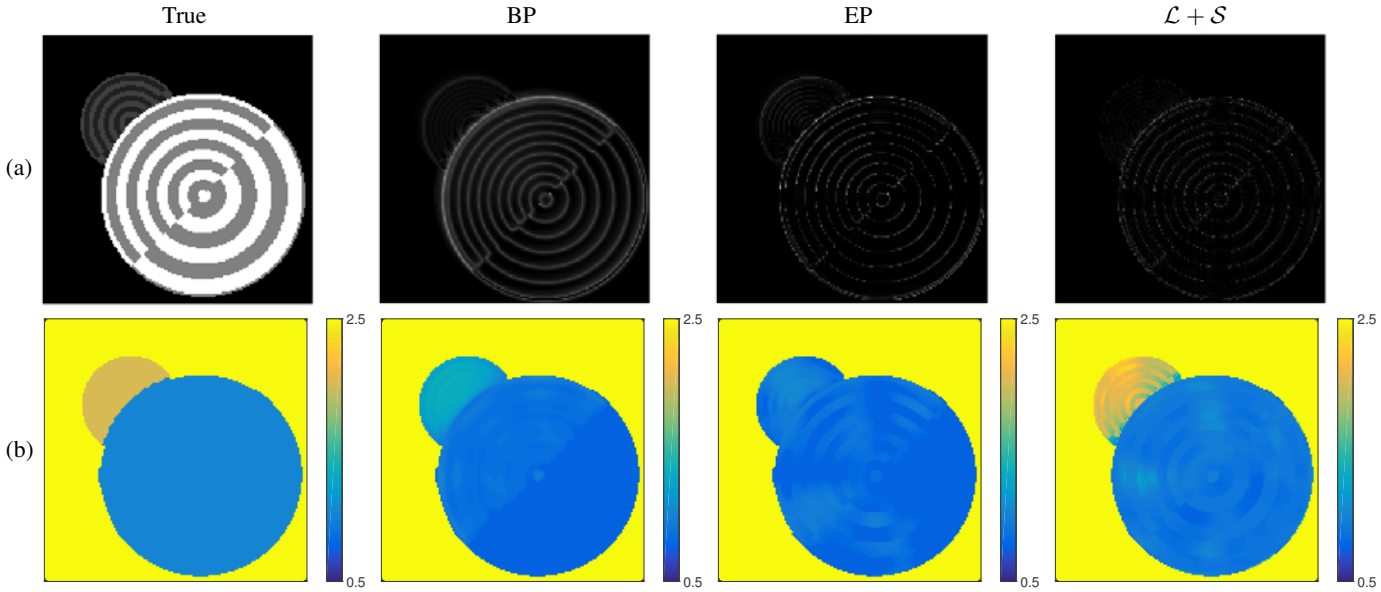

Fig. 3. Comparison of reconstructed light-fields and estimated depths with different reconstruction methods on noiseless measurements. (a) Error maps of reconstructed 1,1 light-field view (farthest from the center view), and (b) Estimated depth maps

TABLE I
ACCURACY OF IMAGE RECONSTRUCTION AND DEPTH ESTIMATION WITH DIFFERENT IMAGE RECONSTRUCTION METHODS ON NOISELESS MEASUREMENTS

	BP	EP	$\mathcal{L}+\mathcal{S}$
PSNR (dB)	21.0	26.1	28.7
Depth RMSE (m)	0.31	0.40	0.16

This plot was typical of the light-field data we analyzed which motivated a larger choice for r. We used the MATLAB Tensor Toolbox [12] for low-rank tensor updates in Section IV-C. We ran 1,000 BPG-M iterations to ensure sufficient convergence.

B. Results: Image Reconstruction

Fig. 3(a) shows the error in the top leftmost view of the light-field. Edges are the most difficult part to reconstruct. While backprojection clearly blurred edges a lot, other methods performed better. Table I compares the performance in terms of PSNR of the three different reconstructions. The proposed $\mathcal{L}+\mathcal{S}$ method improves the PSNR of the EP reconstruction by over 2.5dB in the noiseless measurements case. In the noisy case, we see a sharp decrease in improvement in PSNR to only 0.5dB over the EP method. We are currently working to resolve this apparent sensitivity to noise.

C. Result: Depth Estimation

Our goal in reconstructing the light-field is not for the sake of the data itself, but to enable the interesting applications of light-fields to imaging. Such applications are perspective shifts, digital refocusing, and monocular depth estimation. While our light-field error measures well the ability of our reconstruction to shift perspectives, here we test how well it performs as input to a depth estimation algorithm. Ideally our reconstruction will distribute its error in such a way as to have little effect on depth estimation.

For a depth estimation algorithm, we have chosen to use the Spinning Parallelogram Operator (SPO) method for robust depth estimation [13]. This state-of-the-art method performs well on benchmark data and has code available publicly online.

Fig. 3(b) compares the depth maps attained by applying SPO to BP, EP, and $\mathcal{L}+\mathcal{S}$ reconstructions. While the light-field obtained via EP reconstruction performed better than BP in terms of PSNR, it is not useful for determining depth as it leads to the incorrect classification of the back disc as being at the same depth as the front. Our proposed method $\mathcal{L}+\mathcal{S}$ obtains high PSNR and still allows SPO to distinguish the depths of each disc.

VI. CONCLUSIONS

This paper proposed a regularized reconstruction method for light-fields using low-rank plus sparse tensors. The proposed $\mathcal{L}+\mathcal{S}$ reconstruction model captures well the redundancy and structure inherent in light-field data, leading to improved reconstruction of light-fields from focal stack measurements, in terms of both PSNR and performance in depth estimation algorithms.

In further work, we hope to overcome an apparent sensitivity of the method to noise and to speed up the reconstruction algorithm by tightening the majorizer used for optimizing with BPG-M. Appropriate choice of sparsifying transform W may further improve the image reconstruction accuracy of $\mathcal{L}+\mathcal{S}$.

VII. ACKNOWLEDGEMENT

The authors thank Zhengyu Huang for help with the depth estimation, and acknowledge Miao-Bin Lien, Ted Norris and Zhaohui Zhong for collaboration on focal stack imaging with transparent sensors.

REFERENCES

[1] R. Ng, "Fourier slice photography," *ACM Trans. on Graphics*, vol. 24, no. 3, pp. 735–44, July 2005.

[2] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*, 2010, pp. 1831–8.

[3] A. Mousnier, E. Vural, and C. Guillemot, "Partial light field tomographic reconstruction from a fixed-camera focal stack," 2015, arxiv 1503.01903.

[4] R. Otazo, E. Candes, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Mag. Res. Med.*, vol. 73, no. 3, pp. 1125–36, Mar. 2015.

[5] M. H. Kamal, B. Heshmat, R. Raskar, P. Vandergheynst, and G. Wetzstein, "Tensor low-rank and sparse light field photography," *Comp. Vision & Im. Understanding*, vol. 145, pp. 172–81, Apr. 2016.

[6] C-H. Liu, Y-C. Chang, T. B. Norris, and Z. Zhong, "Graphene photodetectors with ultra-broadband and high responsivity at room temperature," *Nature Nanotechnology*, vol. 9, pp. 273–8, 2014.

[7] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[8] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. Assoc. Comput. Mach.*, vol. 58, no. 3, pp. 1–37, May 2011.

[9] Il Yong Chun and Jeffrey A Fessler, "Convolutional dictionary learning: Acceleration and convergence," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1697–1712, Apr. 2018.

[10] Il Yong Chun and Jeffrey A Fessler, "Convolutional analysis operator learning: Acceleration, convergence, application, and neural networks," submitted, Jan. 2018.

[11] Il Yong Chun and Jeffrey A. Fessler, "Convergent convolutional dictionary learning using adaptive contrast enhancement (CDL-ACE): Application of CDL to image denoising," in *Proc. Sampling Theory and Appl. (SampTA)*, Tallinn, Estonia, Jul. 2017, pp. 460–464.

[12] Brett W. Bader, Tamara G. Kolda, et al., "Matlab tensor toolbox version 2.6," Available online, February 2015.

[13] Shuo Zhang, Hao Sheng, Chao Li, Jun Zhang, and Zhang Xiong, "Robust depth estimation for light field via spinning parallelogram operator," *Computer Vision and Image Understanding*, vol. 145, pp. 148–159, 2016.