IMAGE AND DEPTH ESTIMATION WITH MASK-BASED LENSLESS CAMERAS

Yucheng Zheng and M. Salman Asif

Department of Electrical and Computer Engineering
University of California Riverside

ABSTRACT

Mask-based lensless cameras replace the lens by placing a fixed mask on top of an image sensor. These cameras can potentially be very thin and even flexible. Recently, it has been demonstrated that such mask-based cameras can recover light intensity and depth information of a scene. Existing depth recovery algorithms either assume that the scene consists of a small number of depth planes or solve a sparse recovery problem over a large 3D volume, and lose robustness to complicated scenes consisting of varying depth surface. In this paper, we propose a new approach for depth estimation based on alternating gradient descent algorithm that jointly estimates the continuous depth map and light distribution of a scene. The computational complexity of the algorithm scales linearly with the spatial dimension of the imaging system. We present simulation results on image and depth reconstruction for standard 3D test scenes. The comparison between the proposed algorithm and other method shows that our algorithm is faster and more robust for natural scenes with a large range of depths.

Index Terms— Lensless imaging, depth estimation, non-convex optimization

1. INTRODUCTION

Depth estimation is an important and challenging problem that arises in a variety of applications including computer vision, robotics, and autonomous systems. Existing depth estimation systems are based on disparity frames based on conventional camera or time-of-flight cameras [1–3]. These cameras are often heavy and bulky and require large space for their installation. Therefore, their adoption for portable and lightweight devices with strict physical constraints is still limited.

In this paper, we propose a joint image and depth estimation framework for a computational lensless camera that consists of a fixed, binary mask placed on top of a bare sensor. Such mask-based cameras offer an alternative design for building cameras without lenses. A recent example of mask-based lensless camera is known as FlatCam [4]. In contrast with a lens-based camera that is designed to map every point in the scene to a single pixel on the sensor, every sensor in a FlatCam records light from every point in the scene. A single point source in the scene would cast a shadow of the mask on the sensor, which would shift if the point moves parallel to the sensor plane and expand/shrink if the point source moves toward/away from the sensor plane. The measurements recorded on the sensor thus represent superposition of shifted and scaled versions of the mask shadows corresponding to light sources in different directions and depths. Image and depth information about the scene is thus encoded in the measurements, and we can solve an inverse problem to estimate both of them.

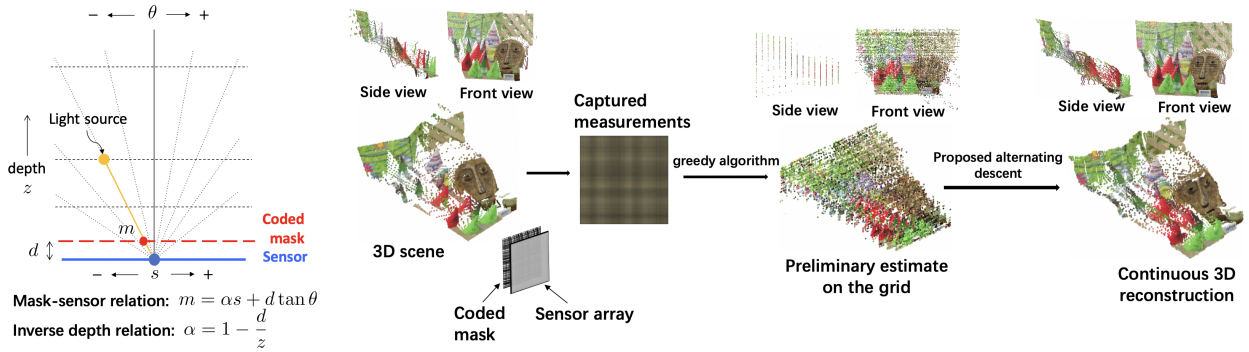
Joint estimation of intensity and depth is an ill-posed problem that multiple solutions exist to fit the measurements. In addition, estimation of depth by itself, even with known scene intensity, is a non-convex problem. To jointly estimate depth and light distribution, we propose a two step approach that consists of an initialization step and an alternating gradient descent step to minimize our objective. To preserve sharp edges in the image intensity and depth map, we include an adaptive regularization penalty in our objective function. An overview of the reconstruction framework is illustrated in Figure 1. We initialize the estimates of image intensity and depth using a greedy algorithm proposed in [5]. Then we refine the estimates by minimizing an objective function with respect to image intensity and depth via alternating gradient descent. To simplify the recovery algorithm, we assume that the mask pattern is smoothly changing and differentiable everywhere. We use adaptive weights to add smoothness regularization on the intensity and depth estimates [6]. Even though the problem of joint estimation of intensity and depth is non-convex, we observed that a simple regularization makes the algorithm robust against local minima of the loss function and improve the performance of the algorithm.

2. RELATED WORKS

The concept of pinhole camera has been known for a very long time. However, the reconstruction quality of pinhole camera is severely affected by noise since the amount of light collected is limited by the pinhole aperture [7]. Mask-based cameras solve this problem by increasing the number of pinholes and allowing more light to reach the sensor [4, 8–10]. Mask-based lensless cameras are often used for imaging light at wavelengths beyond the visible spectrum [9, 10]. With recent advances in computational algorithms and resources, mask-based cameras have been used in computational photography to capture and reconstruct natural images in visible spectrum [11–15].

The imaging model of mask-based camera can be viewed as a linear transform of different points in the scene because the overall measurements are the superposition of the shadow cast by each point source lit alone [5, 8]. With the forward and backward operation known, we can model the joint light and depth estimation problem as an inverse problem. Linear inverse problems arise in various fields such as direction-of-arrival estimation in radar [16], super resolution [17] and compressed sensing [18–20].

The depth recovery algorithms for mask-based cameras have been proposed in [5, 21, 22], all of which are based on on-the-grid signal recovery methods. However, natural scenes contain many light sources located at arbitrary depth and potentially located off-the-grid. Off-the-grid signal recovery methods can be divided into two main types. The first approach formulates the problem as a convex program on a continuous domain and solve it using an atomic norm minimization approach [23, 24]. The second approach models



(a) 1D imaging model for a planar sensor with a coded mask placed at distance d. Light rays from a light source at location (θ, z) are received by all the sensor pixels. A light ray that hits sensor pixel s passes through mask at location m.

(b) An overview of the proposed intensity and depth estimation framework. Consider a natural scene as a 3D point cloud, where each point represents a light source located at a different depth. The camera consists of a fixed, coded mask placed on top of an image sensor. Every point in the scene casts a shadow of the mask on the sensor plane. Each sensor pixels records a linear combination of the scene modulated by the mask pattern. The recovery algorithm consists of two steps. (1) Initialization using a greedy depth selection method. (2) An alternating gradient descent-based refinement algorithm that jointly estimates the light distribution and depth map on a continuous domain.

Fig. 1. The coded mask-based imaging model and an overview of the proposed continuous depth estimation framework.

the signal as a summation of an on-grid approximation and its first-order expansion [17, 25]. Our proposed algorithm is inspired by the second approach.

3. METHODS

3.1. Imaging Model

We assume a 3D scene model that can be divided into $N \times N$ uniformly spaced directions. We use θ_i and θ_j to denote the angular direction of a light source with intensity $l_{i,j}$ and depth $z_{i,j}$ with respect to the center of the sensor. Figure 1(a) depicts the geometry of such an imaging model. A coded-mask is placed on top of the bare sensor array at a small distance d. The $M \times M$ sensor array captures lights coming from the scene modulated by the coded-mask. We use s_u and s_v to denote a sensor pixel on the mask.

Every light source in the scene casts a shadow of the mask on the sensor array, which we denote using basis functions ψ. The shadow cast by a light source with unit intensity at $(\theta_i, \theta_j, z_{i,j})$ can be represented using the following basis function:

$$\psi_{i,j}(s_u, s_v) = \text{mask}\left[\alpha_{i,j} s_u + d\tan(\theta_i), \alpha_{i,j} s_v + d\tan(\theta_j)\right], \quad (1)$$

where mask$[u, v]$ denotes transparency value of the mask pattern at location (u, v) and $\alpha_{i,j}$ is a variable that is related to the physical depth $z_{i,j}$ via the following inverse relation

$$\alpha_{i,j} = 1 - \frac{d}{z_{i,j}}, \quad (2)$$

therefore, we will refer to $\alpha_{i,j}$ as the inverse depth of point in the direction (θ_i, θ_j) in the scene. Let us denote the intensity captured at sensor pixel (s_u, s_v) when the scene consists of a single point source at 3D position $(\theta_i, \theta_j, \alpha_{i,j})$ with intensity $l_{i,j}$ as

$$y(s_u, s_v) = \psi_{i,j}(s_u, s_v) l_{i,j}. \quad (3)$$

The measurements recorded at every sensor pixel represent superposition of measurements from all the point sources in the 3D scene. The intensity of light captured at the sensor pixel (s_u, s_v) is given by the following summation

$$y(s_u, s_v) = \sum_{i=1}^{N} \sum_{j=1}^{N} \psi_{i,j}(s_u, s_v) l_{i,j}. \quad (4)$$

We can write this imaging model in a compact form as

$$\mathbf{y} = \mathbf{\Psi}(\boldsymbol{\alpha})\mathbf{l} + \boldsymbol{e}, \quad (5)$$

where $\mathbf{y} \in \mathbb{R}^{M^2}$ is a vectorized form of an $M \times M$ matrix that denotes sensor measurements, $\mathbf{l} \in \mathbb{R}^{N^2}$ is a vectorized form of an $N \times N$ matrix that denotes light intensity from all the locations $(\theta_i, \theta_j, \alpha_{i,j})$, and $\mathbf{\Psi}$ is a matrix with all the basis functions corresponding to $\theta_i, \theta_j, \alpha_{i,j}$. The basis functions in (5) are parameterized by unknown $\alpha \in \mathbb{R}^{N^2}$. $\boldsymbol{e}$ is the potential noise existing in the observation.

With this imaging model, we formulate the following optimization for fidelity of the imaging model:

$$\hat{\alpha}, \hat{\mathbf{l}} = \underset{\alpha, \mathbf{l}}{\text{argmin}} \frac{1}{2} ||\mathbf{y} - \mathbf{\Psi}(\alpha)\mathbf{l}||_2^2. \quad (6)$$

Note that if we know the true values of α (or we fix it to something), then the problem in (6) reduces to a linear least-squares problem that can be efficiently solved via standard solvers. On the other hand, if we fix the value of $\mathbf{l}$, the problem is nonlinear with respect to α. In the next few sections we discuss our approach for solving the problem in (6) via alternating minimization.

3.2. Image and Depth Estimation

The minimization problem in (6) is not convex, therefore we need a proper initialization to start with. Our approach for initialization is depth pursuit method proposed in [5]. Greedy algorithms are widely used for sparse signal recovery [18–20]. Based on these algorithms, [5] proposed a greedy depth pursuit algorithm for FlatCam depth estimation by iteratively find the depth surface that matches the observed measurements the best. Although this method may not

approximate off-grid point sources well, it produces a good preliminary estimate of the scene without any prior knowledge. We will use its output as the initialization of our minimization problem.

To solve the minimization problem in (6), we alternately update depth and light distribution using gradient information after we obtain a preliminary estimate from the greedy algorithm. The loss function of fidelity is

$$L = \frac{1}{2}\sum_{u=1}^{M}\sum_{v=1}^{M}(y(s_u,s_v) - \sum_{i=1}^{N}\sum_{i=1}^{N}\psi_{i,j}(s_u,s_v)l_{i,j})^2 \quad (7)$$

Let us define $R_{u,v} = y(s_u,s_v) - \sum_{i,j=1}^{N}\psi_{i,j}(s_u,s_v)l_{i,j}$ as the residual approximation error at location (s_u,s_v). The derivatives of loss function with respect to the depth values $\alpha_{i,j}$ can now be written as

$$\frac{\partial L}{\partial \alpha_{i,j}} = = -l_{i,j}\sum_{u=1}^{M}\sum_{v=1}^{M}R_{u,v}\frac{\partial \psi_{i,j}(s_u,s_v)}{\partial \alpha_{i,j}} \quad (8)$$

Then we can compute $\frac{\partial \psi(s_u,s_v)}{\partial \alpha_{i,j}}$ as follows.

$$\frac{\partial \psi_{i,j}(s_u,s_v)}{\partial \alpha_{i,j}} = \frac{\partial \psi_{i,j}(s_u,s_v)}{\partial u_{i,j}}s_u + \frac{\partial \psi_{i,j}(s_u,s_v)}{\partial v_{i,j}}s_v, \quad (9)$$

where $u_{i,j} = \alpha_{i,j}s_u + d\tan(\theta_i)$ and $v_{i,j} = \alpha_{i,j}s_v + d\tan(\theta_j)$. The terms $\frac{\partial \psi_{i,j}(s_u,s_v)}{\partial u_{i,j}}$ and $\frac{\partial \psi_{i,j}s_u,s_v}{\partial v_{i,j}}$ represent the derivatives of the mask pattern along respective spatial direction.

In our method we used a blurred version of the binary mask pattern, which we assume is differentiable everywhere. This assumption is valid because even though the mask pattern printed on the transparency may not be differentiable, the shadow of the mask is smooth and differentiable (mainly because of spreading caused by diffraction). We compute the gradient using finite difference and linear interpolation. We used `minFunc` solver [26] with L-BFGS algorithm [27] to solve the nonlinear optimization problem. We chose the step size using strong Wolfe condition [28]. The algorithm stops when the norm of approximation error converges or maximum number of iterations is reached.

3.3. Regularization Approaches

The optimization problem we are trying to solve (6) is non-convex and the optimization process is sensitive to the initialization. To tackle these issues, we impose a prior on the depth map that the general variation of depth map should be small. To achieve this, we add a quadratic regularization term on the spatial gradients of the depth map to our loss function.

Even though smoothness regularization on depth map removes local minima of loss function, it does not respect the sharp edges in the depth map. To keep sharp discontinuities in depth intact while still enforce the variation of depth to be small, we use an adaptive weighted regularization for depth given as

$$R_W(\boldsymbol{\alpha}) = \sum_{i,j=1}^{N} W_{i,j}^{c,\alpha}(\alpha_{i,j} - \alpha_{i+1,j})^2 + W_{i,j}^{r,\alpha}(\alpha_{i,j} - \alpha_{i,j+1})^2 \quad (10)$$

and the weights are

$$W_{i,j}^{c,\alpha} = e^{-\frac{(\alpha_{i,j}-\alpha_{i+1,j})^2}{2\sigma^2}}, W_{i,j}^{r,\alpha} = e^{-\frac{(\alpha_{i,j}-\alpha_{i,j+1})^2}{2\sigma^2}}.$$

To highlight the effect of the weighted smoothness regularization on depth, we plot the following weighted quadratic function $f(\alpha_i - \alpha_j) = \exp(-(\alpha_i - \alpha_j)^2/\sigma)(\alpha_i - \alpha_j)^2$ in Figure 2, where α_i, α_j stand for inverse depth of neighboring pixels. We plot the weighted function for different values of σ along with a normal quadratic function. The plots show that the quadratic function (without any weights) penalizes large values of depth difference; however, weighted function add small penalty if the neighboring pixels have large depth difference (which indicates the presence of an edge). Thus, weighted regularization forces pixels that have depth within a small range of one another to be smooth and does not penalize the points that potentially lie across an edge. This helps preserve sharp edges in the reconstructed depth estimates. This weighting approach is analogous to bilateral filtering approach for image denoising [29, 30].

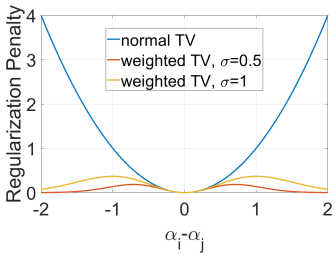


Fig. 2. The weighted regularization function penalizes values that are within a small distance of one another and does not penalize those values that are above certain threshold.

In addition to ℓ_2-based methods, it is also well-known that the ℓ_1 norm regularization of spatial gradient enforces sparsity of the image variation [31]. Ideally, it keeps the edges while sets all the other variations to zero. The ℓ_1-based TV regularization term is

$$R_{TV}(\boldsymbol{\alpha}) = \sum_{i,j=1}^{N} |\alpha_{i,j} - \alpha_{i+1,j}| + |\alpha_{i,j} - \alpha_{i,j+1}|. \quad (11)$$

We solve the nonlinear optimization problem with ℓ_1 regularization term using Split-Bregman steps [32].

Algorithm 1 Weighted TV-ℓ_2 regularized optimization

Input: Sensor measurements: $\mathbf{y}$
Output: Light distribution and depth map of scene: $\mathbf{l}, \boldsymbol{\alpha}$
Initialization via greedy algorithm
Compute $\boldsymbol{\alpha}_{i,j}$ and $\mathbf{l}_{i,j}$ with depth pursuit algorithm
Refinement via alternating gradient descent
for $k = 1 : k_{\max}$ **do**
$\hat{\boldsymbol{\alpha}}^k = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{y} - \boldsymbol{\Psi}(\boldsymbol{\alpha})\mathbf{l}^{k-1}||_2^2 + \lambda R_W(\boldsymbol{\alpha})$
$\hat{\mathbf{l}}^k = \underset{\mathbf{l}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{y} - \boldsymbol{\Psi}(\boldsymbol{\alpha}^k)\mathbf{l}||_2^2$
end for
return $\hat{\mathbf{l}}$ and $\hat{\boldsymbol{\alpha}}$

4. SIMULATION RESULTS

To validate the performance of our algorithm, we simulate an lensless imaging system that a separable MLS sequence is placed 4mm away from the sensor array. The sensor contains 512×512 pixels

and the length of each pixel is 50μm. The chief ray angle of each sensor pixel is $\pm 20^\circ$. We assume that there is no noise added to the sensor measurements. We simulate a 3D scene using the depth data from Middlebury dataset [33]. We sample the scene at uniform angles to create a 128×128 image. We compute the physical depth from α using (2). In our simulation, the depth of the scene ranges from 1m to 1.8m. We use greedy algorithm as our initialization method. We selected 15 candidate depths by uniformly sampling from 1m to 1.8m. The scene and estimation results using conventional methods and proposed methods are shown in Figure 3, along with the root mean squared error (RMSE) of their depth maps. In Figure 3, both image and depth estimation from Greedy [5] contain many spikes due to the model mismatch of the pre-defined depth grid. In contrast, our proposed method with weighted TV-ℓ_2 removes most of the spikes and preserves the edges.

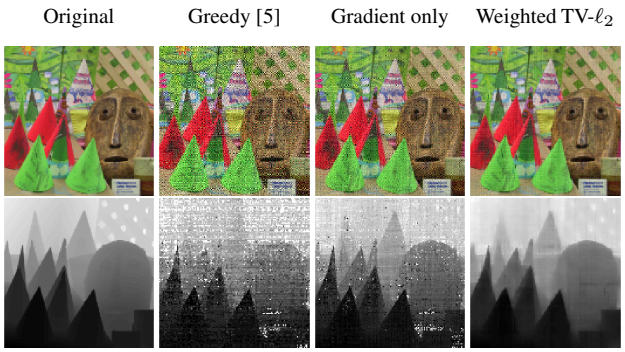


Fig. 3. Left to right: original image and depth of Cones scene (depth range is 1m to 1.8m); image and depth initialized via greedy algorithm [5]; continuous depth estimation via gradient descent without regularization; depth estimation using weighted ℓ_2 regularization.

Comparison of regularization methods. Next, we present a comparison between the three different regularization approaches in Figure 4. As we observe, the optimization with ℓ_2 regularization on spatial gradient is smooth but blur. In opposition to that, the estimatino from TV-ℓ_1 and weighted TV-ℓ_2 preserve the edges of depth map well. Between these two approaches, we observe that weighted TV-ℓ_2 converges faster in our experiment.

Analysis of Noise. To investigate the effect of noise on our algorithm, we present simulation results of reconstructions from the same sensor measurements added with different levels of white Gaussian noise. The plots recording Peak Signal-Noise Ratio (PSNR) of image intensities and root mean squared errors of depth maps are presented in Figure 5. As observed from Figure 5, the quality of both estimated image and depth are improved when the measurements are added with less noise. An example of reconstruction with different levels of noise is shown in Figure 6. As shown in the figures, we are able to reconstruct the general shape of the scene even with presence of noise.

5. CONCLUSION

We proposed a new computational framework for joint estimation of light intensity and depth map from a single image of a mask-based lensless camera. In contrast to existing methods that estimate depth on a grid, our method estimates the depth on a continuous range. Our algorithms consists of a careful initialization step based on greedy pursuit and an alternating minimization step based on gradient descent. We presented different regularization schemes that offer robust recovery on a diverse dataset. Our simulation results demonstrate a significant improvement over existing methods for 3D imaging using coded mask-based lensless cameras.

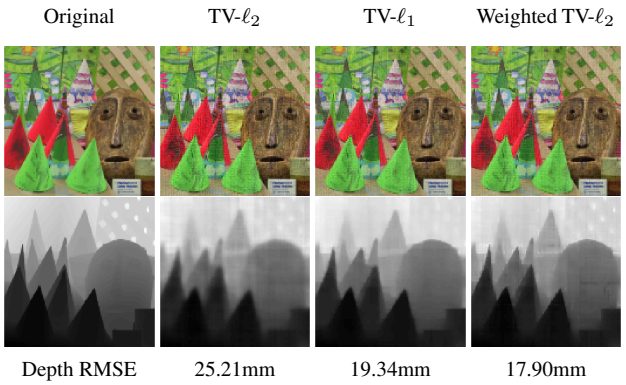
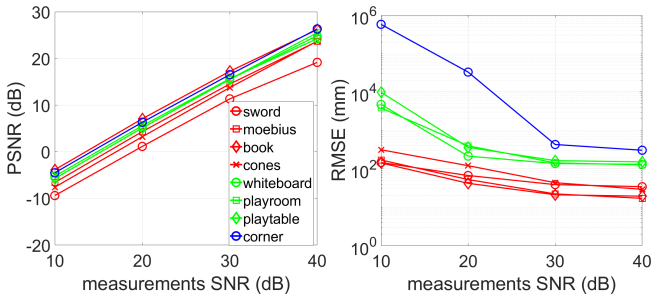


Fig. 4. Comparison between reconstructions using three different regularization approaches from the same measurements.



(a) Image PSNR for different noise levels

(b) Depth RMSE for different noise levels

Fig. 5. Reconstruction from measurements with different levels of noise, the PSNR of image intensities and RMSE of depth maps.

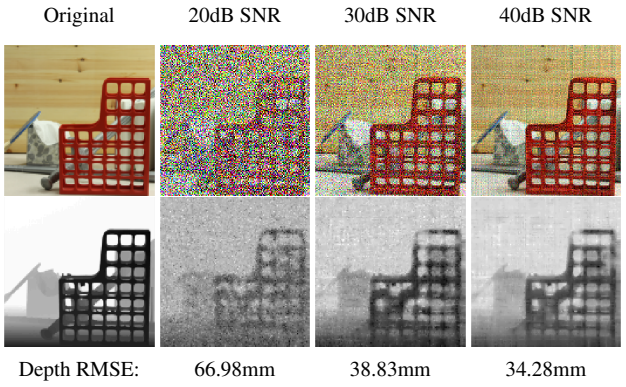


Fig. 6. An example reconstruction from noisy measurements. The sword scene depth range is from 0.6m to 1m.

6. REFERENCES

[1] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor - system description, issues and solutions," in *Conference on Computer Vision and Pattern Recognition Workshop*, June 2004, pp. 35–35.

[2] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[3] Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, pp. 45, 2013.

[4] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, Sept 2017.

[5] M. S. Asif, "Toward depth estimation using mask-based lensless cameras," in *51st Asilomar Conference on Signals, Systems, and Computers*, Oct 2017, pp. 1467–1470.

[6] Yan Liu, Jianhua Ma, Yi Fan, and Zhengrong Liang, "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction," *Physics in Medicine and Biology*, vol. 57, no. 23, pp. 7923, 2012.

[7] Adam Yedidia, Christos Thrampoulidis, and Gregory Wornell, "Analysis and optimization of aperture design in computational imaging," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 4029–4033, April 2018.

[8] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Appl. Opt.*, vol. 17, no. 3, pp. 337–347, Feb 1978.

[9] A. Busboom, H. Elders-Boll, and H. D. Schotten, "Uniformly redundant arrays," *Experimental Astronomy*, vol. 8, no. 2, pp. 97–123, Jun 1998.

[10] T. M. Cannon and E. E. Fenimore, "Coded Aperture Imaging: Many Holes Make Light Work," *Optical Engineering*, vol. 19, pp. 283, June 1980.

[11] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, March 2008.

[12] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *IEEE Computer Vision and Pattern Recognition*, June 2006, vol. 1, pp. 339–346.

[13] Dharmpal Takhar, Jason N. Laska, Michael B. Wakin, Marco F. Duarte, Dror Baron, Shriram Sarvotham, Kevin F. Kelly, and Richard G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," *Proc.SPIE*, vol. 6065, pp. 6065 – 6065 – 10, 2006.

[14] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, July 2007.

[15] Jesse K. Adams, Vivek Boominathan, Benjamin W. Avants, Daniel G. Vercosa, Fan Ye, Richard G. Baraniuk, Jacob T. Robinson, and Ashok Veeraraghavan, "Single-frame 3d fluorescence microscopy with ultraminiature lensless flatscope," *Science Advances*, vol. 3, no. 12, 2017.

[16] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4997–5008, Oct 2014.

[17] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," in *IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2015, pp. 57–60.

[18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[19] Deanna Needell and Joel A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, pp. 93–100, Dec. 2010.

[20] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, April 2010.

[21] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller, "Diffusercam: lensless single-exposure 3d imaging," *Optica*, vol. 5, no. 1, pp. 1–9, Jan 2018.

[22] Jesse K. Adams, Vivek Boominathan, Benjamin W. Avants, Daniel G. Vercosa, Fan Ye, Richard G. Baraniuk, Jacob T. Robinson, and Ashok Veeraraghavan, "Single-frame 3d fluorescence microscopy with ultraminiature lensless flatscope," *Science Advances*, vol. 3, no. 12, 2017.

[23] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, Dec 2012.

[24] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.

[25] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, Jan 2013.

[26] Mark Schmidt, "minfunc: unconstrained differentiable multivariate optimization in matlab," 2005.

[27] Dong C. Liu and Jorge Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, Aug 1989.

[28] P. Wolfe, "Convergence conditions for ascent methods," *SIAM Review*, vol. 11, no. 2, pp. 226–235, 1969.

[29] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Jan 1998, pp. 839–846.

[30] Frédo Durand and Julie Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, July 2002.

[31] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[32] Tom Goldstein and Stanley Osher, "The split bregman method for l1-regularized problems," *SIAM J. Img. Sci.*, vol. 2, no. 2, pp. 323–343, Apr. 2009.

[33] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision*, Dec 2001, pp. 131–140.