OSA Imaging and Applied Optics Congress 2020 (3D, AOMS, COSI, DH, ISA) © OSA 2020

Multi-sensor lensless imaging: synthetic large-format sensing with a disjoint sensor array

Ellin Zhao, Nicolas Deshler, Kristina Monakhova, Laura Waller
Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
ellinzhao@berkeley.edu, nico.deshler@berkeley.edu, monakhova@berkeley.edu

Abstract: We demonstrate a lensless diffuser-based camera array for large field-of-view imaging. Images are captured from multiple disjoint sensors and the synthetic large format sensing area is recovered by solving a compressive sensing inverse problem.

1. Introduction

Large-format camera sensors enable wider fields-of-view (FoV) and larger baselines for 3D imaging, which could be useful for applications such as robotic vision and scientific imaging. Despite these advantages, large imaging sensors are expensive and difficult to manufacture [1]. Furthermore, the size, weight, and complexity of lens assemblies scale with sensor size, resulting in large, bulky systems. Tiling multiple small sensors together is a cost-effective and scalable approach to emulating a larger sensor. However, tiled sensor arrays exhibit gaps in the effective sensing area, rendering them incompatible with a standard lens as points in the scene that focus to a gap between sensors in an array would be unrecoverable. Here, we propose replacing the camera lens with a pseudorandom phase diffuser which maps each point in the scene to a high-contrast intensity pattern on the sensing plane [2]. We leverage compressive sensing techniques conditional on these multiplexed intensity measurements to realize a large synthetic sensing area from tiled sensors. The diffuser is placed close to the sensor array giving rise to a lightweight and compact imager that recovers the scene by solving a sparsity-constrained inverse problem. In this paper we present an experimental prototype system featuring a 2×2 sensor array built from commodity hardware. Our prototype reconstructs images using an effective sensing area that covers merely 8.6% of the total synthesized sensing area.

2. Theory

Our prototype is based on DiffuserCam, a lensless camera that consists of a phase mask placed over a sensing element [2]. Here, we consider a non-contiguous sensor array behind a single phase mask. The camera measurements, $\mathbf{b}$, can be modeled as a masked 2D convolution between the scene, $\mathbf{v}[x,y]$, and the point spread function (PSF), $\mathbf{h}[x,y]$, where $\mathbf{M}[x,y]$ is the binary mask. This can be written as a matrix-vector multiplication, where $\mathbf{H}$ represents the optical forward model and $*$ denotes the 2D convolution operator:

$$\mathbf{b} = \mathbf{M}[x,y]\cdot\Big(\mathbf{h}[x,y]*\mathbf{v}[x,y]\Big) = \mathbf{MHv}. \tag{1}$$

The scene is recovered by solving a non-negative sparsity-constrained inverse problem using the fast iterative shrinkage-thresholding algorithm (FISTA) [3], where Ψ is a sparsifying transform (e.g. 2D total variation) and τ is a tuning parameter:

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}\geq 0}\frac{1}{2}\|\mathbf{b}-\mathbf{MHv}\|_2^2 + \tau\|\Psi\mathbf{v}\|_1. \tag{2}$$

3. Prototype and Results

The hardware setup (Fig. 1 (a, b)) for the multi-sensor prototype consists of an off-the-shelf diffuser (Luminit 0.5°) placed over a 2×2 array of camera sensors (Raspberry Pi AdaFruit SpyCams). Calibration of the system occurs in two stages: PSF acquisition and sensor orientation. The first stage, PSF acquisition, obtains a PSF corresponding to a synthetic sensing area that is larger than the effective sensing area. We laterally raster the diffuser over the array and image the caustic patterns formed by illumination from a distant point source. The resulting set of images is run through a panoramic stitching algorithm to create the "full" PSF. The second stage of calibration identifies the positions and orientations of each sensor in the array relative to the diffuser. Each sensor images a distant point source centered above the array. A semi-automated image registration algorithm then uses these images to determine the rigid transformation that overlays each sensor image onto the full PSF [4]. These transformations are used to generate subsequent sensor array measurements, **b**, and to create the masking operator, **M**. The described calibration sequence can handle arbitrary array configurations and is robust to imprecise sensor placements.

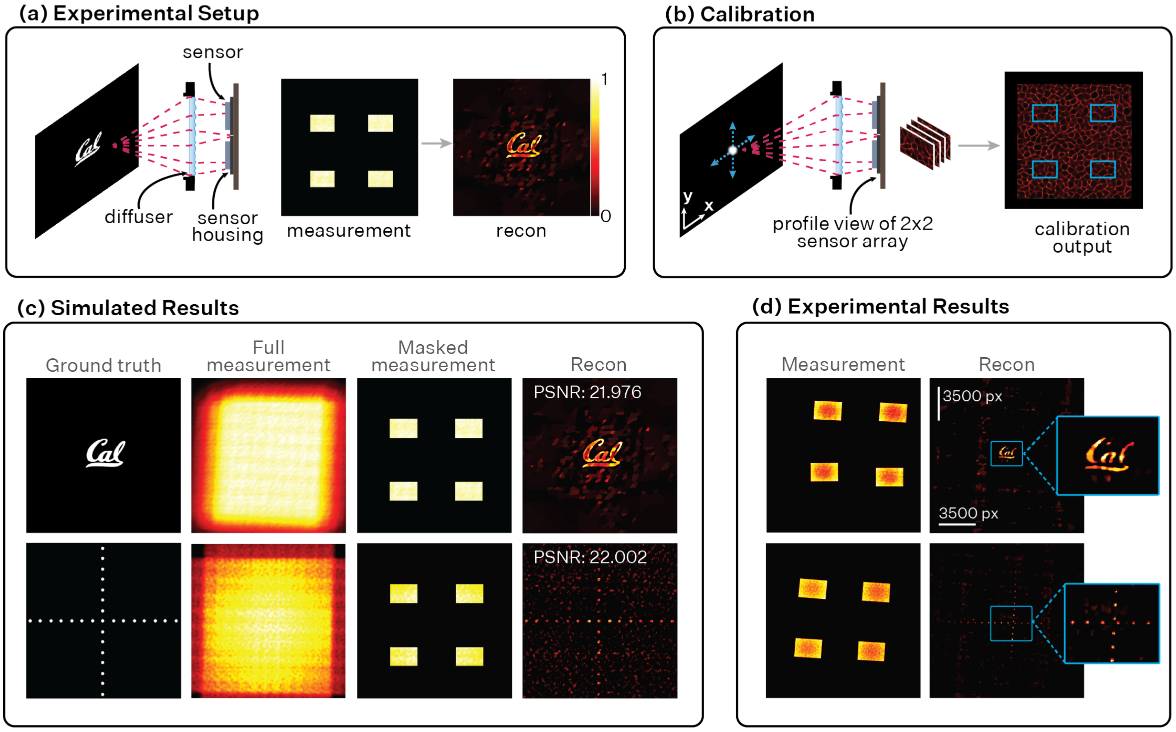

Fig. 1. **(a)** Multi-sensor DiffuserCam: 2×2 sensor array produces a non-contiguous DiffuserCam measurement from which the scene is reconstructed. **(b)** Calibration: relative positions of the four sensors are found using a point-matching algorithm that defines rigid transformations mapping each sensor (blue boxes) onto the full PSF. **(c)** Simulated results: image reconstructions for a synthetic large-format sensing area $12\times$ larger than the effective sensing area of the disjoint sensor array. **(d)** Experimental results: image reconstructions of Cal logo and sparse point sources located along the gaps in the sensor array.

Figure 1(c) shows successful recovery of simple scenes from simulated measurements. Here, the four sensors in the array cover only 8.6% of the synthetic sensing area suggesting that Multi-sensor DiffuserCam could be ideal for systems with limited memory capacity or data rate bottlenecks. We successfully recover content on the edges of the FoV, providing a $10\times$ increase in the lateral FoV from the single-sensor case. Figure 1(d) shows experimental reconstructions, which due to the placement of the objects, could not have been recovered by a lensed sensor array. Our reconstructions use 3000 iterations of FISTA with TV regularization. Experimentally, we are able to recover a number of simple objects, but the quality of reconstructions suggests only a $5\times$ increase in the FoV from the single-sensor case due to model mismatch. The resolution of the system is dependent on scene complexity; sparse objects have a higher recovery rate than dense scenes. More sophisticated data-driven regularization methods could be used to improve the reconstructions and potentially recover more dense scenes [5].

References

1. G. H. Chapman, J. Leung, R. Thomas, A. Namburete, Z. Koren, and I. Koren, "Projecting the rate of in-field pixel defects based on pixel size, sensor area, and ISO," in *Sensors, Cameras, and Systems for Industrial and Scientific Applications XIII,* (SPIE, 2012).
2. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," Optica **5**, 1–9 (2018).
3. A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," IEEE Transactions on Image Process. **18**, 2419–2434 (2009).
4. A. Goshtasby, "Image registration by local approximation methods," Image Vis. Comput. **6**, 255–261 (1988).
5. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," Opt. express **27**, 28075–28090 (2019).