optica

DiffuserCam: lensless single-exposure 3D imaging: supplementary material

NICK ANTIPA*, GRACE KUO*, REINHARD HECKEL, BEN MILDENHALL, EMRAH BOSTAN, REN NG, AND LAURA WALLER+

Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, California 94720, USA
*Authors contributed equally
+Corresponding Author: lwaller@alum.mit.edu

This document provides supplementary information to "DiffuserCam: lensless single-exposure 3D imaging," https://doi.org/10.1364/OPTICA.5.000001. We provide details on the properties of the diffuser used in our prototype DiffuserCam system. Based on these properties, we validate the field-of-view equation presented in the main text and quantify the similarity between PSFs at different locations in the volume. We also provide ad-ditional details about the algorithm including an explanation of cropping in the forward model, a derivation of the ADMM algorithm used for solving for 3D volumes, and additional details on the implementation of this algorithm. Finally, we compare two different regularizers and demonstrate that the ADMM algorithm is computationally efficient for our problem.

1. SYSTEM PROPERTIES

A. Diffuser Properties

To quantify the properties of our diffuser, we used an LED array microscope to capture a quantitative Differential Phase Contrast (DPC) [1] image of the diffuser phase. After using the index of refraction of the diffuser material (polycarbonate, $n = 1.58$) to convert phase into surface shape, we show in Fig. S1 the measured relative height profile of a small patch on our $0.5°$ diffuser. The surface slope of the diffuser is Gaussian distributed with average magnitude of $0.7°$. The deflection angle at the diffuser surface has a HWHM angle of $0.25°$, which matches the manufacturer specifications. The maximum deflection angle is $\beta = 0.5°$, as shown in the histograms in Fig. S1.

To illustrate the overall size and spread of the caustic PSF patterns in our system, we show in Fig. S2 the full PSF patterns captured for the closest and farthest axial distances used. Note that the closest axial distance is the one at which the caustic pattern just fills the sensor, and therefore depends on the aperture size. The caustics contain high-frequency information in all orientation directions, as evidenced by the sharp lines randomly spread in all directions. This facilitates good resolution at all depths and a highly structured PSF for deconvolution. Our calibration point source is a $30\mu m$ pinhole illuminated by a planar
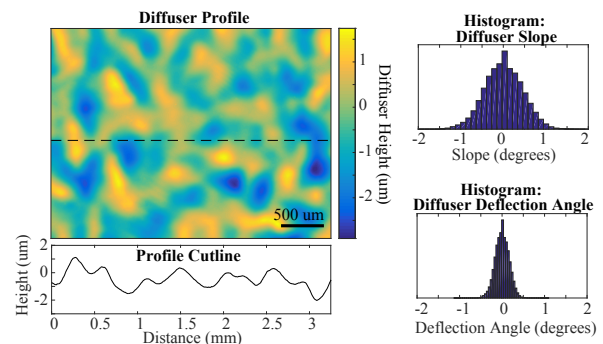

Fig. S1. Left: The thickness profile of a small patch of our diffuser, as measured by quantitative Differential Phase Contrast (DPC) microscopy. Below is a cut-line plot along the dashed line. Right: Histograms of the diffuser slope (top) and the deflection angle of a ray normally incident on the diffuser (bottom). The maximum deflection angle is about $0.5°$.

RGB LED array ($\lambda = 630$ nm, 515 nm, and 460 nm, $\Delta\lambda = 20$ nm, 35 nm, and 25 nm, respectively) placed behind a $80°$ diffuser. As shown in [2], the caustics from narrowband and broadband sources are indistinguishable, and we do not find problems with using narrowband calibration.
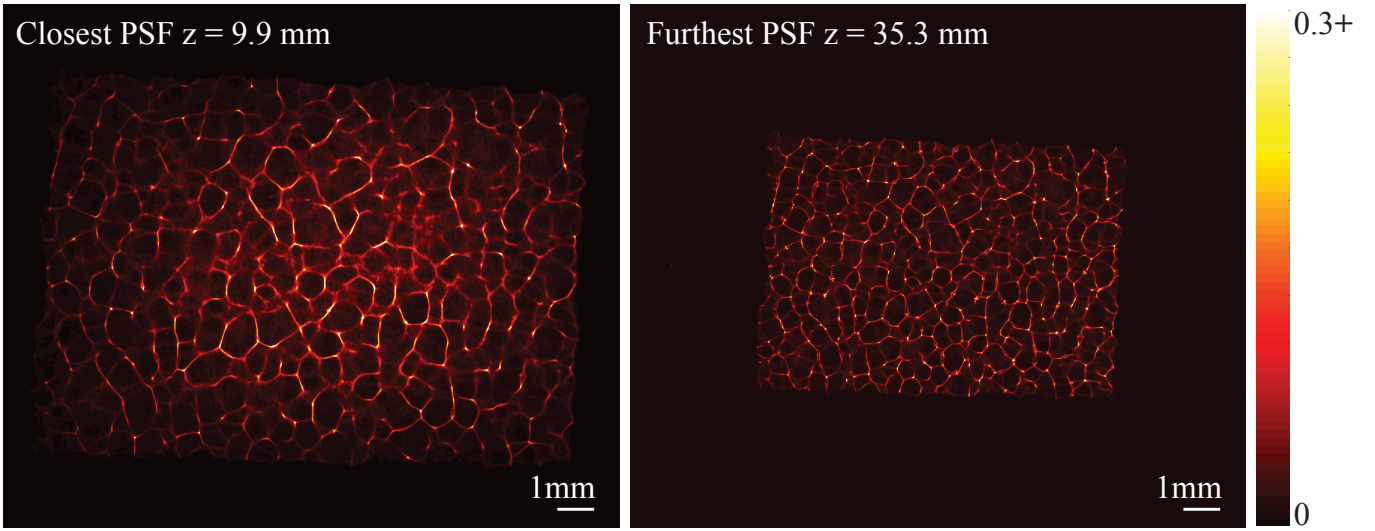
Fig. S2. Un-cropped, false color sensor measurements of PSFs for the closest and farthest planes used in our reconstructions. These were measured by placing a point source on-axis at the front and back of the volume. The closest PSF has a caustic pattern that fills the sensor. Both PSFs have been contrast stretched from 0 to 30% of the max value for visibility.

B. Field-of-View Validation

In the main text in Sec. 3A, we derive the field-of-view (FoV) of our system to be

$$\text{FoV} = \beta + \min[\alpha_c, \tan^{-1}(\tfrac{l+w}{2d})], \tag{S1}$$

where the FoV can be limited by either the geometry of the system (l, w, d) or by the angular acceptance of the pixels (α_c). Here l is the sensor size, w is the aperture size, and d is the distance between the diffuser and the sensor. In our system, $d = 8.9$ mm. In the x-direction, $l_x = 16.6$ and $w_x = 7.5$ mm; the y-direction values are $l_y = 14$ mm and $w_y = 5.5$ mm.

The angular response of the sensor, shown in Figure 3 of the main text, was measured by placing a white LED at optical infinity and rotating the sensor both vertically and horizontally. The average intensity measured at each angle was normalized by the on-axis measurement. We define the angular cutoff, α_c, as the angle at which the response falls to 20% of its on-axis value. For our camera, the x and y cutoffs are $\alpha_{cx} = 41.5°$ and $\alpha_{cy} = 30°$, respectively. Finally, from our diffuser measurements in Fig. S1, we find that the maximum deflection angle of the diffuser, β, is $0.5°$.

Plugging these values into the FoV equation yields a FoV of $42°$ in x and $30.5°$ in y, where the limiting factor is the angular acceptance. Figure S3 shows the recovery of a large, evenly illuminated scene at optical infinity. The angular extent visible in the reconstruction matches our predicted FoV.

C. PSF similarity

We quantify the similarity of the PSF versus shift and scale across the volume to validate our claim that the resulting underdetermined matrix has good properties for sparse recovery techniques. Figure S4 shows the autocorrelation of the PSFs acquired at the minimum and maximum object distances, as well as the cross-correlation between the two. Notice that the PSF autocorrelation maintains a sharp central peak and relatively low sidelobes for all depths within our calibration volume. This means that a shifted version of the PSF is roughly 50% similar to the un-shifted version. Importantly, the cross-correlation has

no values greater than 50%, meaning that the scaled caustics are dissimilar to any shift of the unscaled caustics. To quantify this further, we plot the inner product between the central image in the calibration stack, corresponding to the orange dotted line in Fig. S4b, with all other images in the stack. We again observe a relatively sharp peak and side lobes on the order of 50% in the axial direction. This validates our claim that the caustics produced by any point in the volume are unique.

2. ALGORITHM DETAILS

A. Cropping in Forward Model

In Eq. (4) of the main text, we show that our forward model is a sum of convolutions followed by a crop operation. We would like to emphasize that the crop operation is due directly to the physical cropping caused by the finite sensor size. Consider an off-axis point source, as shown in Fig. S5a. In the experimental measurement from the source (Fig. S5b), half of the on-axis PSF is cut off by the finite size of the sensor. If we do not take this into account in our forward model, our estimate of the measurement would look like Fig. S5c, which is not physical due to the circular boundary conditions. Including the crop operation in our forward model fixes the problem, creating estimates of the measurement that look like the experimental data (Fig. S5d).

B. Derivation of ADMM Inverse Algorithm Formulation

As stated in the main paper in Section 1B, the problem we seek to solve is:

$$\hat{\mathbf{v}} = \underset{\mathbf{v} \geq 0}{\arg\min} \ \tfrac{1}{2}\|\mathbf{b} - \mathbf{H}\mathbf{v}\|_2^2 + \tau\|\Psi\mathbf{v}\|_1. \tag{S2}$$

We transform this into the equivalent problem:

$$\hat{\mathbf{v}} = \underset{w,u,v}{\arg\min} \ \tfrac{1}{2}\|\mathbf{b} - \mathbf{D}v\|_2^2 + \tau\|u\|_1 + \mathbb{1}_+(w)$$

$$\text{s.t. } v = \mathbf{M}\mathbf{v}$$
$$u = \Psi\mathbf{v} \tag{S3}$$
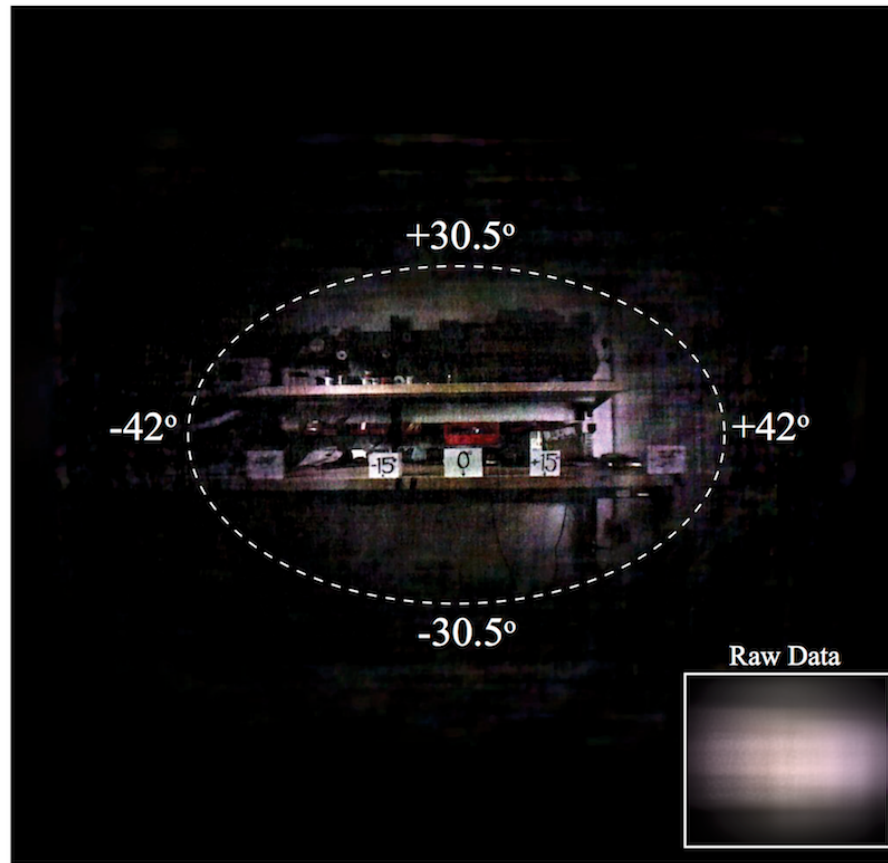$$w = \mathbf{v},$$

Fig. S3. Validation of FoV calculations: based on the measured angular pixel response, α_c, and maximum diffuser deflection angle, β, we calculate our theoretical FoV to be 42° in x and 30.5° in y. This matches our recovered FoV in a scene at optical infinity. The inset shows the raw data.
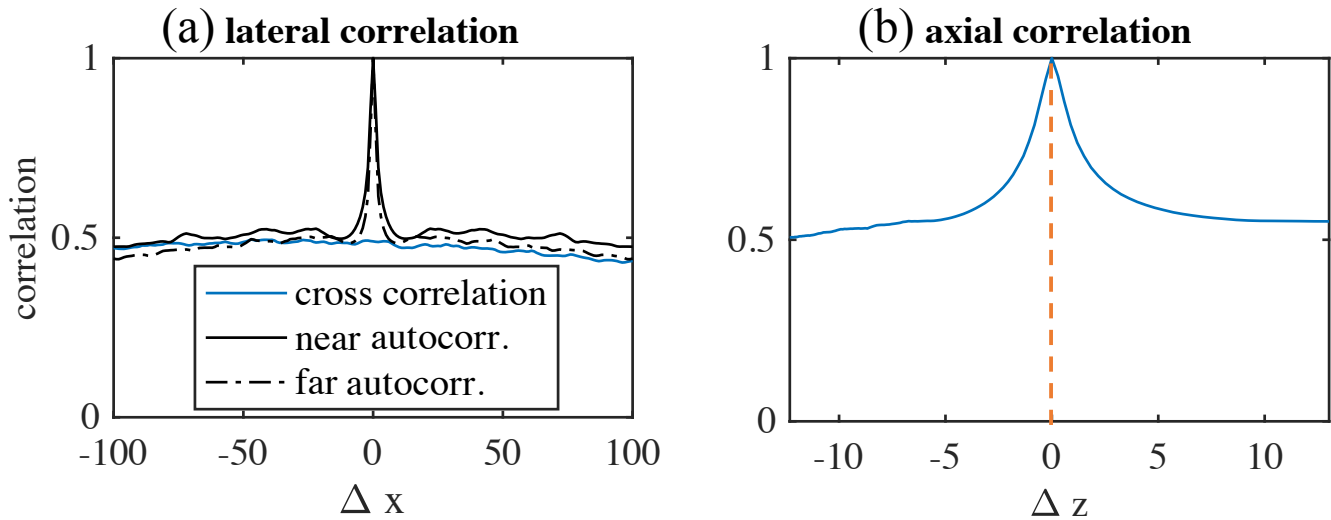

Fig. S4. Correlation of various caustics patterns. (a) The caustics at a given depth are unique over shifting, and caustics from two different depths are not similar to each other, even under translation. The solid black curve is a slice of the autocorrelation of a PSF for a point source near the front of the volume, and the dotted black line is the autocorrelation for a far away point source's PSF. The solid blue line is the cross-correlation between the two. (b) The inner product of the PSF from the middle of the volume (corresponding to the orange dotted line) with all other PSFs at varying depths. In both (a) and (b), shifting or scaling the caustics leads to an inner product of approximately 0.5 compared to a peak value of 1.
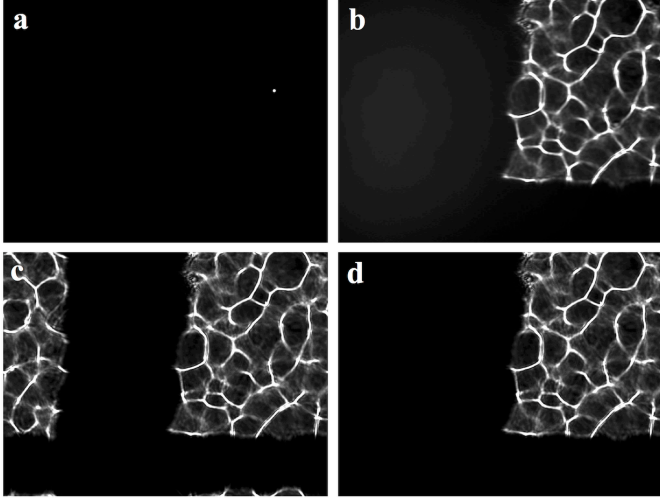
Fig. S5. The crop operation in the forward model accounts for the finite sensor size. (a) Off-axis point source, size exaggerated for visibility. (b) Experimental measurement from the source. (c) Simulated measurement without crop operation. Since the convolution has circular boundary conditions, the PSF wraps around to the opposite side of the sensor. (d) Simulated measurement with crop operation matches the experimental measurement.

where $\mathbb{1}_+(\cdot)$ is the nonnegativity barrier function, which returns 0 when the argument is nonnegative, and ∞ when the argument is negative.

In order to compute the ADMM updates efficiently, we will see that it is useful for both $\mathbf{M}$ and Ψ to represent 3D convolutions. Clearly, when Ψ is the identity matrix, this holds. Additionally, when Ψ is the 3D finite difference operator, it can be expressed as a concatenation of 3D convolutions with the finite difference kernel, oriented in each of the 3 directions. In order to express $\mathbf{M}$ as a 3D convolution, we must choose the diagonal operator, $\mathbf{D}$, such that Eq. (4) can be written as $\mathbf{D}\left(m \overset{(x,y,z)}{*} \mathbf{v} \right)$, where m is a 3D kernel, and $\overset{(x,y,z)}{*}$ represents convolution over the variables, x, y, and z. To accomplish this, we use the fact that a sum of 2D convolutions between an object, $\mathbf{v}(x,y,z)$, and a stack of 2D kernels, $h(x,y;z)$, can be expressed as the first 2D (x,y)-slice in the 3D convolution between the object and a z-flipped version of the kernel stack:

$$\sum_z h(x,y;z) \overset{(x,y)}{*} \mathbf{v}(x,y,z) = \left[h(x,y;-z) \overset{(x,y,z)}{*} \mathbf{v}(x,y,z) \right]\Bigg|_{z=0}.$$
(S4)

For proof, we can take the right hand side of Eq. (S4) and apply the definition of discrete 3D convolution directly:

$$\left[h(x,y;-z) \overset{(x,y,z)}{*} \mathbf{v}(x,y,z) \right]\Bigg|_{z=0}$$
$$= \sum_{z'=0}^{N_z-1} \sum_{y'=0}^{N_y-1} \sum_{x'=0}^{N_x-1} \mathbf{v}(x',y',z') h(x-x',y-y';z'-z)|_{z=0}$$
$$= \sum_{z'=0}^{N_z-1} \mathbf{v}(x,y,z') \overset{(x,y)}{*} h(x,y;z').$$

Using this identity, we can write the forward operator in Eq. (4) as:

$$\mathbf{C}\sum_z \left[\mathbf{v}\left(\frac{-x'}{m}, \frac{-y'}{m}, z \right) \overset{(x,y)}{*} h\left(x',y';z\right) \right]$$
$$= \mathbf{C}\left[\mathbf{v}\left(\frac{-x'}{m}, \frac{-y'}{m}, z \right) \overset{(x',y',z)}{*} h\left(x',y';-z\right) \Bigg|_{z=0} \right]$$
$$= \mathbf{D}\left[\mathbf{v}\left(\frac{-x'}{m}, \frac{-y'}{m};z \right) \overset{(x',y',z)}{*} h\left(x',y';-z\right) \right],$$

where $\mathbf{D}$ is a diagonal operator that simultaneously performs the 2D crop, $\mathbf{C}$, as well as selecting the $z = 0$ slice. Effectively, $\mathbf{D}$ comprises taking the center crop of the first layer of the 3D array resulting from the circular 3D convolution of $h(x',y';-z)$ with $\mathbf{v}$. Note that our definition of z is as a parameter indexing each slice in the 3D array h, not the physical distance to each slice. We assume circular boundary conditions for h, such that $h(\cdot,\cdot;-z) = h(\cdot,\cdot;N_z - z)$ is a z-stack that is flipped in the z-direction.

Using Eq. (S4), we present an efficient method for solving Eq. (S3). We begin by transforming Eq. (S3) into an unconstrained augmented Lagrangian form, and consider the saddle-point problem:

$$\max_{\xi,\eta,\rho}\left[\min_{u,v,w,\mathbf{v}} \frac{1}{2}\|\mathbf{b} - \mathbf{D}v\|_2^2 + \frac{\mu_1}{2}\left\|\mathbf{M}\mathbf{v} - v + \frac{\xi}{\mu_1}\right\|_2^2 \right.$$
$$+ \tau\|u\|_1 + \frac{\mu_2}{2}\left\|\Psi\mathbf{v} - u + \frac{\eta}{\mu_2}\right\|_2^2$$
$$\left. + \mathbb{1}_+(w) + \frac{\mu_3}{2}\left\|\mathbf{v} - w + \frac{\rho}{\mu_3}\right\|_2^2 \right].$$

To solve the above equation using ADMM, we first derive the optimality conditions for each primal variable, assuming the others are fixed:

$$u^{k+1} \leftarrow \underset{u}{\arg\min}\quad \tau\|u\|_1 + \frac{\mu_2}{2}\left\|\Psi\mathbf{v}^k - u + \frac{\eta^k}{\mu_2}\right\|_2^2$$
$$v^{k+1} \leftarrow \underset{v}{\arg\min}\quad \frac{1}{2}\left\|\mathbf{b}^k - \mathbf{D}v\right\|_2^2 + \frac{\mu_1}{2}\left\|\mathbf{M}\mathbf{v}^k - v + \frac{\xi^k}{\mu_1}\right\|_2^2$$
$$w^{k+1} \leftarrow \underset{w}{\arg\min}\quad \mathbb{1}_+(w) + \frac{\mu_3}{2}\left\|\mathbf{v}^k - w + \frac{\rho^k}{\mu_3}\right\|_2^2$$
$$\mathbf{v}^{k+1} \leftarrow \underset{\mathbf{v}}{\arg\min}\quad \frac{\mu_1}{2}\left\|\mathbf{M}\mathbf{v} - v^{k+1} + \frac{\xi^k}{\mu_1}\right\|_2^2$$
$$+ \frac{\mu_2}{2}\left\|\Psi\mathbf{v} - u^{k+1} + \frac{\eta^k}{\mu_2}\right\|_2^2$$
$$+ \frac{\mu_3}{2}\left\|\mathbf{v} - w^{k+1} + \frac{\rho^k}{\mu_3}\right\|_2^2.$$

And update each dual variable as

$$\xi^{k+1} \leftarrow \xi^k + \mu_1(\mathbf{M}\mathbf{v}^{k+1} - v^{k+1})$$
$$\eta^{k+1} \leftarrow \eta^k + \mu_2(\Psi\mathbf{v}^{k+1} - u^{k+1})$$
$$\rho^{k+1} \leftarrow \rho^k + \mu_3(\mathbf{v}^{k+1} - w^{k+1}).$$

The final result is the algorithm outlined in Sec. 2C of the main text.

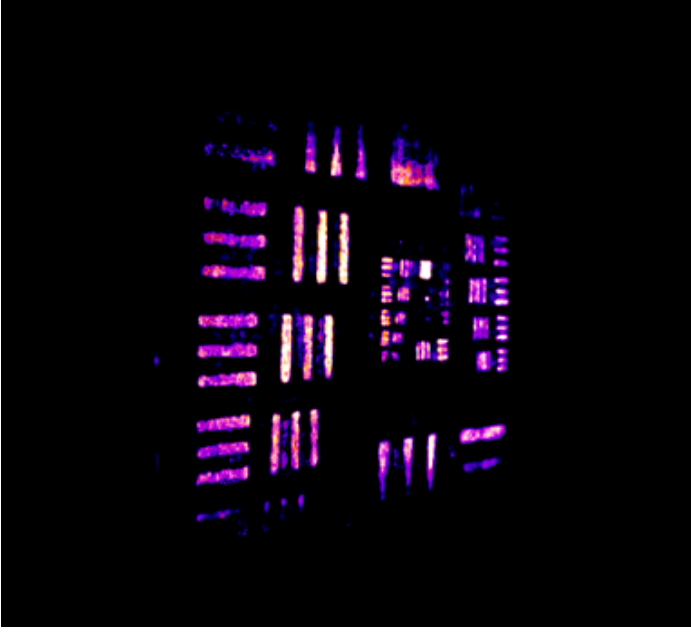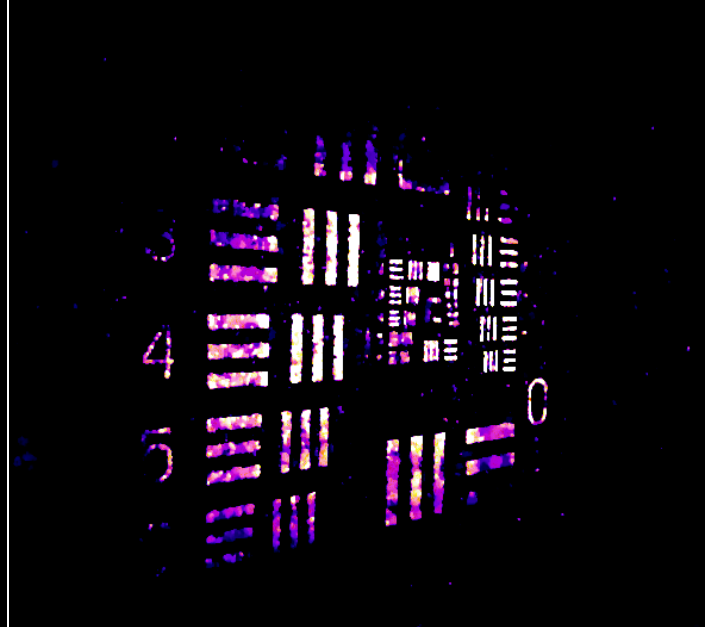**a** FISTA, ℓ_1, 2048x2048x64 (4 hours)   **b** ADMM, 3DTV, 2048x2048x128 (26 mins)

Fig. S6. ℓ_1 vs 3DTV regularization with different algorithm implementations. (a) Max z-projection of FISTA reconstruction using ℓ_1 (soft thresholding on the volume after each iteration). This took 4 hours to run on a Titan X GPU using MATLAB. The soft thresholding has erased some key features. (b) Max z-projection of reconstruction using ADMM with a 3DTV prior. Clearly the result is better, largely due to a better sample-prior match. This reconstruction also required 10x less time to obtain.

C. Implementation details

Evaluation of the cropped discrete convolution at a single depth,

$$\mathbf{C}\left[h(x',y';z) \overset{(x',y')}{*} \mathbf{v}(-x'/m,-y'/m,z)\right](x',y'),$$

is done by zero padding $h(x',y';z)$ to twice its original size in each dimension, then using FFT-based convolution. This ensures that any aliasing artifacts introduced by the circular boundary conditions of the FFT will fall outside the sensor area, causing such artifacts to be removed by the cropping, $\mathbf{C}(\cdot)$. Note that this requires our variable, $\mathbf{v}$, to be approximately twice as many samples in each dimension as our sensor measurement. Interestingly, it is possible for useful information to lie anywhere within this extended FoV. In our prototype, the angular falloff of the sensor means that measurements in the extended region are attenuated too much to be useful. However, a future system using different geometry could leverage this effect to gain even more useful samples in the final reconstruction. In operator notation, the convolution can be evaluated as

$$\mathbf{C}F^{-1}\big\{[FPh(x,y;z)] \cdot [F\mathbf{v}(x,y,z)]\big\} \qquad (S5)$$

where F is the 2D FFT, $\cdot$ is point-wise multiplication, and P is the zero-padding operator.

D. ℓ_1 vs 3D Total Variation

To improve the quality of reconstruction, we use the 3D Total Variation (TV) penalty parameter. This is inefficient to compute as part of a projected gradient technique, because the proximal operator for the TV norm must be computed iteratively. On volumes of the size used here, this requires minutes per outer-loop iteration. Of the priors considered in this work, only native

sparsity and nonnegativity are feasible when using projected gradient methods. To demonstarate the benefit of using ADMM, we show in Fig. S6 the result from ℓ_1 regularized FISTA after running for 4 hours on a GPU using MATLAB compared to our algorithm with 3DTV regularization for 20 minutes. Not only does our algorithm run much faster, but it produces an image with more detail. In particular, note that the ℓ_1 regularization has erased the numbers and eroded the bars, whereas 3DTV runs an order or magnitude faster *and* uses a more sophisticated prior, resulting in categorically better performance.

REFERENCES

1. L. Tian, J. Wang, and L. Waller, "3d differential phase-contrast microscopy with computational illumination using an led array," Optics letters **39**, 1326–1329 (2014).
2. N. Antipa, S. Necula, R. Ng, and L. Waller, "Single-shot diffuser-encoded light field imaging," in "2016 IEEE International Conference on Computational Photography (ICCP)," (2016), pp. 1–11.