



On-chip fluorescence microscopy with a random microlens diffuser

GRACE KUO,¹  FANGLIN LINDA LIU,¹ IRENE GROSSRUBATSCHER,²
REN NG,¹ AND LAURA WALLER^{1,*} 

¹Electrical Engineering and Computer Sciences, University of California, Berkeley 94709, USA

²Helen Wills Neuroscience Institute, University of California, Berkeley 94709, USA

*waller@berkeley.edu

Abstract: We present an on-chip, widefield fluorescence microscope, which consists of a diffuser placed a few millimeters away from a traditional image sensor. The diffuser replaces the optics of a microscope, resulting in a compact and easy-to-assemble system with a practical working distance of over 1.5 mm. Furthermore, the diffuser encodes volumetric information, enabling refocusability in post-processing and three-dimensional (3D) imaging of sparse samples from a single acquisition. Reconstruction of images from the raw data requires a precise model of the system, so we introduce a practical calibration scheme and a physics-based forward model to efficiently account for the spatially-varying point spread function (PSF). To improve performance in low-light, we propose a random microlens diffuser, which consists of many small lenslets randomly placed on the mask surface and yields PSFs that are robust to noise. We build an experimental prototype and demonstrate our system on both planar and 3D samples.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

On-chip microscopy is a powerful imaging modality in which a digital image sensor captures information about the sample without using a traditional microscope objective. These lensless microscopes can be very compact and lightweight for portable or *in vivo* applications, and they typically have simpler hardware than their lensed counterparts. However, many on-chip microscopes are limited to bright-field microscopy [1–3] rather than fluorescence imaging, a critical modality for probing structure and function in a wide range of samples.

As summarized in Greenbaum et al. [1], on-chip fluorescence imaging is challenging for several key reasons. First, fluorophores are incoherent with each other and with background illumination. As a result, digital holography [2,3] and other interferometric methods cannot be applied. Shadow-based techniques [4,5] are also not applicable because fluorescent samples do not necessarily block light. Furthermore, fluorophores emit light uniformly in all directions; in an on-chip system without a main lens, fluorophores therefore become dim and defocused as they move further from the sensor. This results in degradation of both signal-to-noise ratio (SNR) and resolution with increasing distance from the sensor. Prior on-chip microscopes for fluorescence [6–10] mitigate this effect by using very short working distances (less than 500 μm), limiting their applications to samples that can be placed directly on the sensor. In this work, we demonstrate an on-chip fluorescence microscope featuring a practical working distance of over 1.5 mm, suitable for imaging samples on slides or in microfluidic channels.

Our strategy for on-chip fluorescence microscopy involves placing a thin mask between the sample and the sensor. The mask modulates incoming light, indirectly encoding information about the sample, which can then be recovered computationally. Since the mask is placed close to the sensor (3.8 mm), it does not greatly increase the system form factor or hardware complexity (as compared to [11]). Such designs maintain the advantages of an on-chip lensless system and have been demonstrated successfully in both microscopy [10,11] and photography [12–18], and

have been shown to capture higher-dimensional information, such as three-dimensional (3D) [19] or temporal information [20], in a single acquisition.

Here, our mask is a random microlens diffuser [20–22] which has many small lenslets randomly arranged in 2D. Since the lenslets have focusing power, the best performance occurs when the object is in imaging condition with the sensor, enabling practical working distances, over 1.5 mm. In contrast, similar architectures with amplitude masks that have no focusing power [10,12] have the best performance when the object is close to the mask, resulting in short working distances (< 500 μm). Furthermore, unlike amplitude masks, our random microlens diffuser does not block light, making it better suited for fluorescent samples which are typically dim. As in [19], our system can recover 3D structures from a single acquisition; in this work, we demonstrate 8 μm lateral resolution and 50 μm axial resolution, an order of magnitude higher than in [19].

The architecture of our system has many parallels to a 4D light field camera [23–25] or an integral photography system [26,27], which instead uses a periodic microlens array. Similar to a light field camera, each lenslet of our random microlens diffuser can be thought of as imaging the object from a different perspective. However, because our proposed system uses random rather than regular arrays of lenslets, cross-talk between the lenslets can be disambiguated computationally. This allows for increased flexibility in the design of the micro-optics, eliminates the need for a main objective lens, and enables a simple flat architecture that does not require physically isolating each lenslet, as in [16,17,28]. As a result, our system is easy to assemble, compact and portable (total size of 3.5 cm \times 3.5 cm \times 1 cm, limited by the board size of the sensor), and the architecture can easily be extended to larger sensor sizes.

Our microscope requires computational recovery of the image from the raw data. Prior work [19] demonstrated efficient computational image reconstruction by assuming the system point spread function (PSF) is shift-invariant. Unfortunately, this assumption relies on objects being far from the camera (compared to the sensor size). Here, we achieve 10 \times resolution improvement by placing objects closer to the sensor and modeling the spatially-varying PSFs, as described in Sec. 2. Calibration of a spatially-varying system is challenging; one approach is to experimentally measure every PSF in the field-of-view (FoV) [29,30]. However, this brute force calibration approach would necessitate an infeasible number (over 10 million) of calibration images for 3D imaging. Therefore, we derive a calibration scheme in which we measure about 1,000 PSFs in a sparse grid, then intelligently interpolate between them (Sec. 3), resulting in 40,000 \times fewer calibration measurements and 10,000 \times less memory required than a brute force approach. We show how these calibration measurements can be efficiently combined through a local convolution model. In Sec. 4 we introduce the random microlens diffuser and demonstrate its advantages over other masks. Finally, in Sec. 5 we describe our experimental prototype and show results on both planar and 3D fluorescent samples.

2. System overview

Our microscope consists of a thin refractive diffuser placed a few millimeters in front of a traditional image sensor, with color filters between the diffuser and sensor to block excitation light (Fig. 1). Emitted light from each fluorophore in the sample is refracted by the diffuser surface to form a high-contrast pattern on the sensor. Every position in 3D creates a unique pattern, or PSF. We model our scene as a collection of point sources with varying intensity and no occlusions, so we can describe the imaging system with the following linear equation:

$$\mathbf{b}(x', y') = \mathbf{C} \sum_{(x, y, z)} h(x', y'; x, y, z) \mathbf{x}(x, y, z) + \mathbf{g}(x', y'). \quad (1)$$

Here, $\mathbf{b}(x', y')$ is the measurement at position (x', y') on the sensor, $h(x', y'; x, y, z)$ is the PSF taken at position (x, y, z) in the world, $\mathbf{x}(x, y, z)$ is the sample intensity, \mathbf{C} is a crop operator that accounts for the finite sensor size, and $\mathbf{g}(x', y')$ is the background due to unattenuated excitation

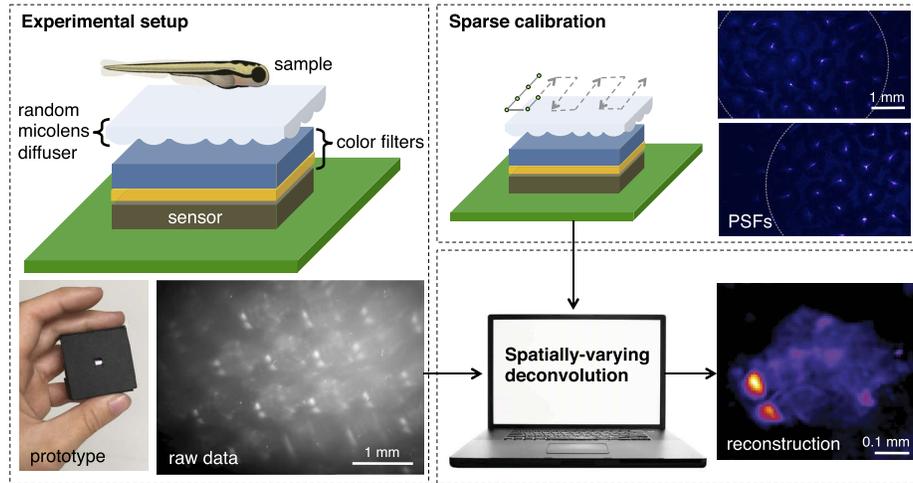


Fig. 1. Our light-weight and portable on-chip microscope consists of a random microlens diffuser placed a few millimeters above an image sensor. Using only a sparse grid of calibration measurements, 3D images are reconstructed with a local convolution model that accounts for the spatially-varying point spread functions (PSFs).

light and autofluorescence from the color filters. Vectorizing the sample, measurement, and background allows Eq. (1) to be written compactly in matrix form: $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{g}$. To recover the object from the sensor measurement, we jointly estimate the sample fluorescence, \mathbf{x} , and background, \mathbf{g} , by solving the regularized least squares problem described in Sec. 3.3.

2.1. Forward model

Reconstructing the sample requires knowing the matrix \mathbf{A} , or equivalently, the PSFs for every point in 3D space. Prior work [19] assumed that the distance between the object and sensor was large relative to the sensor size, making the PSF shift-invariant at each depth; this enabled $h(x', y'; x, y, z)$ to be fully characterized by only one calibration measurement per axial location and enabled \mathbf{b} to be efficiently computed with convolutions. In this work, we place objects closer to the sensor in order to achieve microscopic resolution; however, this breaks the shift-invariance assumption and necessitates accounting for the angular dependence of the sensor.

To capture the dominant optical effects in our system, we introduce what we'll refer to as our "two-part model" in which we split the PSF measurements into two components: (1) A shift-invariant pattern due to the diffuser in which we assume no spatially-varying aberrations, and (2) an angular-dependent response at the CMOS sensor.

The first component of our two-part model is the same as the model in [19]. We define $h_d(x', y'; x, y, z)$ to be the aberration-free PSF from a point source at position (x, y, z) due to the diffuser alone. Assuming paraxial optics, a lateral translation of the point source results in a translation of the PSF in the opposite direction, and an axial translation of the point source results in a scaling of the PSF, as described in [19,31,32]. For notational simplicity, we refer to the on-axis PSF as $h_d(x', y'; z) = h_d(x', y'; 0, 0, z)$. Using the paraxial approximation, we can write any PSF as a transformation of the on-axis PSF taken at depth z^* :

$$h_d(x', y'; x, y, z) = h_d\left(sx' + \frac{sd}{z}x, sy' + \frac{sd}{z}y; z^*\right), \quad (2)$$

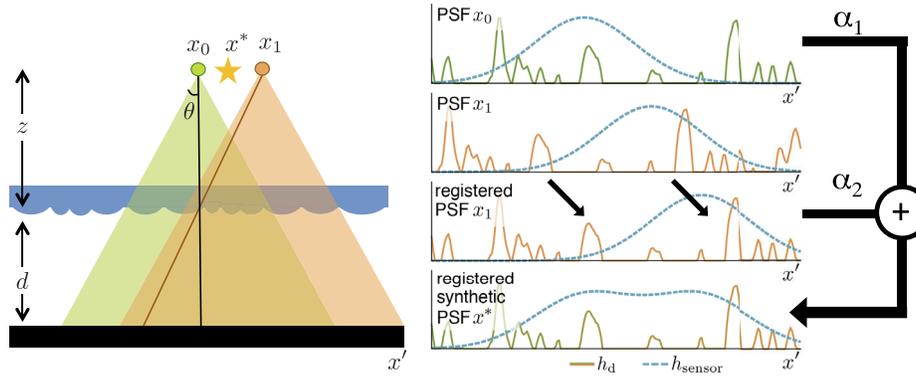


Fig. 2. The two-part forward model, which consists of a diffuser component (h_d) and sensor component (h_{sensor}), informs interpolation between calibration measurements. Consider two calibration points located on-axis at x_0 and off-axis at x_1 . Based on the two-part model, each PSF consists of a high frequency signal due to the diffuser (h_d) multiplied by the angular-dependent sensor response (h_{sensor}). In the off-axis PSF, h_d is translated to the left while h_{sensor} remains directly under the point source. To interpolate between the two calibration measurements, first we register the component due to the diffuser. Then we approximate the PSF at x^* by taking a linear combination of the registered PSFs where the weights (α) are based on the distance between the calibration location and x^* . We extend this to 2D using bilinear interpolation.

where d is the distance between the diffuser and the sensor and s is a scaling factor defined by

$$s = \frac{z(d + z^*)}{z^*(d + z)}. \quad (3)$$

The second component of our two-part model is the sensor response, which was unaccounted for in [19]. Digital image sensors exhibit angle-sensitive responses due to cosine falloff, microlenses on the pixels, and circuitry that blocks light [33]. We assume a simplified model in which the sensor response is summarized by $f(\theta)$ where θ is the angle between the incident light and the normal to the sensor (see Fig. 2). This model is best for backside-illuminated sensors, in which circuitry is beneath the photodiode and does not create spatial variation in the sensor response. Based on the geometry of the system, the PSF due to the sensor response alone is

$$h_{\text{sensor}}(x', y'; x, y, z) = f\left(\tan^{-1}\left(\frac{\sqrt{(x' - x)^2 + (y' - y)^2}}{d + z}\right)\right).$$

Assuming that the diffuser does not bend light by a substantial amount compared to the frequencies in $f(\theta)$, we model the total PSF, $h(x', y'; x, y, z)$, as the product of the sensor and diffuser components, yielding our complete two-part model:

$$h(x', y'; x, y, z) = h_d\left(sx' + \frac{sd}{z}x, sy' + \frac{sd}{z}y; z^*\right) f\left(\tan^{-1}\left(\frac{\sqrt{(x' - x)^2 + (y' - y)^2}}{d + z}\right)\right). \quad (4)$$

These two components can be seen in the experimental PSFs shown in Fig. 1. Note that when the distance to the object, $d + z$, is much larger than the sensor size (recall that x', y' are constrained

to be within the sensor size), the sensor falloff term reduces to

$$h_{\text{sensor}}(x', y'; x, y, z) = f \left(\tan^{-1} \left(\frac{\sqrt{x^2 + y^2}}{d + z} \right) \right),$$

which is independent of x', y' . Therefore, when the object is far from the sensor, as in [19], the falloff term can be implicitly grouped with the sample, and the total PSF consists only of the shift-invariant component due to the diffuser.

The factorization in Eq. (4) suggests that all PSFs can be synthesized from the two underlying functions $h_d(x', y'; z)$ and $f(\theta)$. However, recall that $h_d(x', y'; z)$ is an approximation which assumes no spatially-varying aberrations in the diffuser. In practice, due to spatially-varying aberrations, there is no single $h_d(x', y'; z)$ that accurately describes the diffuser's pattern for all field positions. Therefore, we capture both aberrations and variations due to sensor falloff by measuring the PSF at many locations across the FoV, described in detail in the following section.

3. Calibration and image reconstruction

3.1. Calibration with model-based interpolation

Due to both the sensor falloff term and aberrations from the diffuser, the PSF is shift-varying and cannot be modeled as a convolution, as was done in [19]. To measure these effects, we acquire a collection of calibration measurements throughout the volumetric FoV. In the extreme, one could capture or simulate a calibration measurement at every possible 3D location in the volume, but this is excessive and unnecessary due to the slowly-varying nature of both the diffuser aberrations and the sensor falloff. Instead, we collect a sparse grid of calibration measurements and introduce an interpolation scheme based on the two-part model in Sec. 2.1.

We denote the i -th calibration measurement taken at (x_i, y_i, z_i) as $h_i(x', y') = h(x', y'; x_i, y_i, z_i)$. Our goal is to interpolate between these calibration measurements to synthesize every PSF in the volume. However, naive pixel-wise averaging of neighboring calibration measurements will result in inaccurate blurring of the high-frequency diffuser pattern since it fails to account for the translation and scaling effects described in Eq. (2). Therefore, instead of interpolating between the raw measurements, we first computationally register the calibration measurements to the on-axis PSF taken at depth z by applying the following shifts and scales:

$$\begin{aligned} \tilde{h}_i(x', y'; z) &= h_i(s_i x' - s_i \Delta x'_i, s_i y' - s_i \Delta y'_i) \\ &= h_d(x', y'; z) f \left(\tan^{-1} \left(\frac{\sqrt{(s_i x' - \frac{d}{z} x_i - x_i)^2 + (s_i y' - \frac{d}{z} y_i - y_i)^2}}{d + z_i} \right) \right) \end{aligned} \quad (5)$$

where

$$s_i = \frac{z(d + z_i)}{z_i(d + z)}, \quad \Delta x'_i = \frac{d}{z s_i} x_i, \quad \Delta y'_i = \frac{d}{z s_i} y_i. \quad (6)$$

The registered calibration measurements, denoted $\tilde{h}_i(x', y'; z)$, have the paraxial component of the diffuser pattern aligned (Fig. 2). After this transformation, pixel-wise interpolation between neighboring calibration measurements preserves the high-frequency features of h_d . In a shift-invariant system, all of the registered PSFs are identical and no interpolation is necessary; in a system with spatial variance, like ours, the variations appear as smoothly changing deviations between the registered PSFs. Therefore, we synthetically generate intermediate PSFs by linearly interpolating between the registered calibration measurements. We first consider the case where all calibration PSFs are at a single depth, z , which yields $s_i = 1$ for all i . The synthetically

generated PSF from an arbitrary point (x, y, z) is approximated from the calibration measurements:

$$h(x', y'; x, y, z) \approx \sum_i \alpha_i(x, y, z) \tilde{h}_i(x' + \frac{d}{z}x, y' + \frac{d}{z}y; z). \quad (7)$$

Here $\alpha_i(x, y, z)$ is a weighting factor determined by the distance to the i -th calibration measurement. We choose $\alpha_i(x, y, z)$ to correspond to bilinear interpolation between the four nearest PSFs.

It is also possible to synthesize new PSFs at different depths using an analogous procedure. To generate PSFs at a new depth z^* , nearby calibration measurements are first scaled based on Eq. (5) to generate $\tilde{h}_i(x', y'; z^*)$. Then, the synthetic PSF at z^* is calculated from a linear combination of these scaled calibration measurements, where the weight is once again based on the synthetic PSF's proximity to the calibration points.

Although Eq. (6) requires the precise locations of the calibration measurements, we can circumvent this requirement by directly determining $(\Delta x'_i, \Delta y'_i)$ from the measurements themselves. We find that cross-correlation between neighboring calibration measurements is maximized when the diffuser component of the PSF is aligned. Therefore, we choose a central PSF to act as the on-axis PSF, then calculate $(\Delta x'_i, \Delta y'_i)$ by determining the translation that maximizes the cross-correlation with this central PSF. We find this approach more robust than physically measuring the translation.

3.2. Local convolution model

By applying Eq. (7) we can generate the complete set of calibration measurements needed for the system matrix \mathbf{A} . However, its large size makes \mathbf{A} computationally inefficient to generate and store. Luckily, the linear structure of Eq. (7) allows us to form what we call the *local convolution model* which models the raw data without explicitly generating every PSF. When we plug Eq. (7) into Eq. (1), the convolutional structure becomes apparent:

$$\begin{aligned} \mathbf{b}(x', y') &= \mathbf{C} \sum_{(z,i)} \sum_{(x,y)} \alpha_i(x, y, z) \mathbf{x}(x, y, z) \tilde{h}_i(x' + \frac{d}{z}x, y' + \frac{d}{z}y; z) \\ &= \mathbf{C} \sum_{(z,i)} \left[\alpha_i\left(-\frac{z}{d}x', -\frac{z}{d}y', z\right) \mathbf{x}\left(-\frac{z}{d}x', -\frac{z}{d}y', z\right) \right] * \tilde{h}_i(x', y'; z). \end{aligned} \quad (8)$$

Here $*$ denotes a 2D convolution in the sensor coordinates; note that the sensor coordinates, (x', y') , and world coordinates, (x, y) , are related by the system magnification, d/z . In Eq. (8) we assume that we have calibration measurements at every depth of interest since we can generate PSFs at new depths using the procedure outlined in the previous section.

With this model we can efficiently interpolate between the calibration measurements as we compute the forward model. We refer to this as the *local convolution model* because we can think of $\alpha_i(x, y, z)$ as choosing a region around the i -th calibration measurement where the measurement is valid, then performing a convolution in this region. If the support of the object is known, computational efficiency can be further improved by using the subset of the calibration measurements corresponding to the 3D object support.

Any interpolation scheme that can be written in the form of Eq. (7) is compatible with the local convolution model, which has the distinct advantage of only requiring that the calibration measurements themselves be stored in memory. Although other interpolation schemes could be used in place of Eq. (7), they would require pre-computing and storing every possible PSF in the volume, using on the order of $10,000\times$ more memory. Computing PSFs on-the-fly is too computationally expensive for practical use in an iterative optimization process.

3.3. Inverse problem with background estimation

To recover the object from the raw data, we formulate a regularized inverse problem using the local convolution model described in Eq. (8). Since many fluorescent samples have a sparse structure, we use an ℓ_1 loss on the object for regularization. In addition, there is frequently autofluorescence and unattenuated excitation light hitting the sensor, which does not match our model and corrupts the reconstruction. Therefore we jointly estimate a low-frequency background component along with the object by solving the following minimization problem:

$$\hat{\mathbf{x}}, \hat{\mathbf{g}} = \underset{\mathbf{x}, \mathbf{g}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{b} - (\mathbf{A}\mathbf{x} + \mathbf{g})\|_2^2 + \tau \|\mathbf{x}\|_1 \quad (9)$$

s.t. $\mathbf{g} \geq 0, \mathbf{x} \geq 0, \mathbf{D}\mathbf{g} = 0$ outside low frequency support.

Here \mathbf{g} is the estimated background that cannot be well-explained by the forward model, \mathbf{D} is the 2D discrete cosine transform (DCT) operator, and τ is a tuning parameter. Without constraints on \mathbf{g} , a trivial solution to Eq. (9) is $\mathbf{x} = 0$ and $\mathbf{g} = \mathbf{b}$. To prevent this, we constrain \mathbf{g} such that its DCT coefficients are zero outside some low-frequency support, typically the 5×5 lowest frequency components. We jointly solve for \mathbf{x} and \mathbf{g} using the fast iterative shrinkage-thresholding algorithm (FISTA) [34]. We find that it helps convergence to initialize the estimated background with a low-pass filtered version of the raw data.

Solving for a 3D sample from a single 2D measurement is an under-determined problem, and compressed sensing theory [35–37] can provide guidance regarding what samples will be accurately reconstructed. Without regularization, there are infinite possible 3D distributions that match the raw data. Therefore, the ℓ_1 regularization term and non-negativity constraints are critical to guide the optimization. As a result, we expect the highest-quality results when the sample matches the underlying assumptions, mainly that the sample is natively sparse. For dense samples, the object could be transformed into a different basis, as described in [36].

3.4. Determining spacing between calibration measurements

The local convolution model relies on experimentally measuring the PSF at discrete calibration points, then interpolating to synthesize the remaining PSFs. Here we apply Nyquist sampling theory to determine the appropriate spacing between calibration measurements.

First, we consider lateral sampling using the two-part model presented in Sec. 2.1, and we assume that all calibration PSFs are captured at the same axial location, $z_i = z$ for all i . Registering the calibration measurements based on Eq. (6) aligns the diffuser component of the PSF, but not the sensor falloff component. The second term in Eq. (5) describes the spatial variance of the falloff function after registration. This is the function that we are interpolating, so we must Nyquist sample this term to enable robust interpolation of PSFs. Since we assumed the sensor component is rotationally symmetric and all pixels have the same falloff response (a good assumption for backside-illuminated sensors), we reduce our analysis to the 1D case and only consider a single representative pixel at $x' = 0$. This yields the registered falloff function $\tilde{f}(x)$:

$$\tilde{f}(x) = f\left(\tan^{-1}\left(\frac{x}{z}\right)\right). \quad (10)$$

If $\tilde{f}(x)$ is sampled at the Nyquist frequency, we can robustly interpolate between samples at different positions. To determine the lateral sampling requirements, we experimentally measure $f(\theta)$, shown in Fig. 3(a). Figure 3(b) shows the magnitude of the Fourier transform of $\tilde{f}(x)$ with its maximum frequency declared to be the frequency at which the magnitude falls to 0.1% of maximum. This corresponds to a Nyquist sampling period of about 400 μm .

To test this, we simulate raw data using the experimentally measured $f(\theta)$, then deconvolve using the local convolution model. We simulate calibration measurements at varying spacings

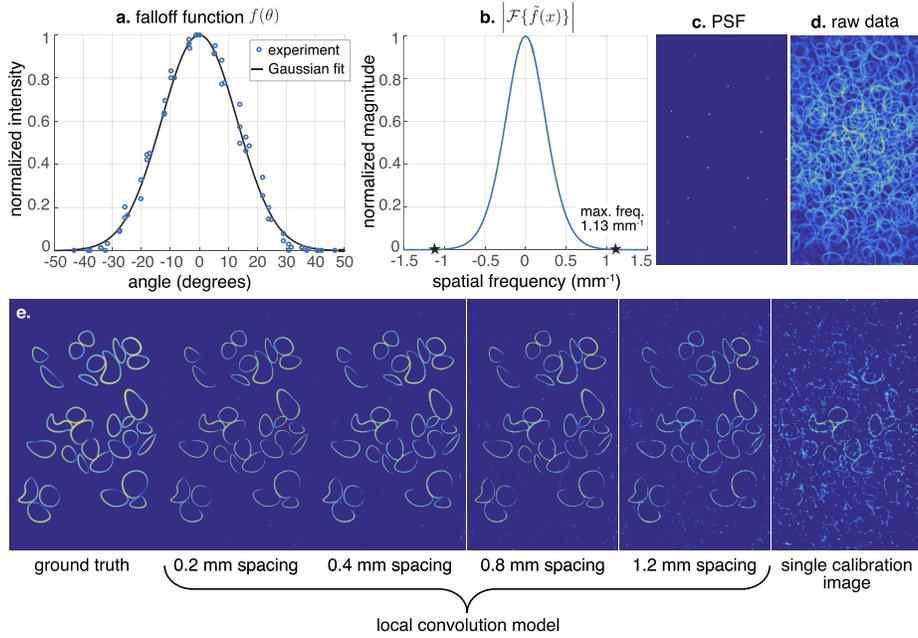


Fig. 3. The number of calibration measurements needed for our local convolution model is determined by the angular falloff of the sensor. (a) Experimentally measured angular falloff, $f(\theta)$, fit to a Gaussian curve with $\sigma = 13^\circ$. (b) Fourier transform after $f(\theta)$ is transformed based on Eq. (10), which is used to determine the maximum frequency, and thus the Nyquist sampling, in order to determine the necessary calibration sampling. (c) To test this, we simulate PSF measurements and (d) raw data using the Gaussian approximation of $f(\theta)$. (e) We then reconstruct the sample using the local convolution model with varying spacing between the calibration measurements. When calibration images satisfy the Nyquist sampling (0.4 mm apart), the model performs well, but when samples are spaced further apart, the reconstruction degrades.

and find that when the calibration measurements are at the Nyquist sampling rate or closer together, we get good reconstruction quality. However, as the measurements move further apart, substantial artifacts appear. Finally, if a single calibration measurement is used, without any model-based interpolation as in [19], the reconstruction fails, demonstrating the necessity of the local convolution model in the microscopy regime.

Next, we consider axial sampling. Axial changes in the PSF are primarily due to defocus, and therefore cannot be explained with the two-part model of Sec. 2.1. However, by assuming the diffuser has a microlens structure with lenslets of focal length f , we can determine the axial sampling based on the depth-of-field. If the radius of the circle of confusion, $\frac{p(d-f)}{2f} \left| 1 - \frac{f}{z(d-f)} \right|$, changes by no more than the diffraction-limited spot size, $\frac{\lambda}{\text{NA}}$, between neighboring samples, then we have fully sampled the axial defocus function. This yields the condition

$$\frac{pd}{2} \left(\frac{1}{z_1} - \frac{1}{z_2} \right) \leq \frac{\lambda}{\text{NA}} \approx \frac{2d\lambda}{p} \quad (11)$$

$$\frac{1}{z_1} - \frac{1}{z_2} \leq \frac{4\lambda}{p^2},$$

where z_1, z_2 are the axial locations of neighboring calibration images, p is the microlens diameter, and λ is the wavelength of light.

4. Random microlens diffuser

The local convolution model and the calibration scheme described above apply to a wide variety of diffuser designs; the only restrictions are that the spatially-varying aberrations in the PSF change smoothly as a function of position, and that no two points in the volumetric FoV generate identical PSFs. In contrast with prior works [10,14] that require specific mask designs for image reconstruction, we have the freedom to design the diffuser to improve other aspects of the system, in particular, resolution, working distance, and noise sensitivity. Specifically, we propose using a random microlens diffuser, as in [20–22], which consists of small lenslets randomly arranged in 2D. To illustrate the advantages of the random microlens diffuser, we compare with a traditional microscope objective and two types of flat transparent masks: a smooth diffuser [19,38] and a regular microlens array. Figures 4 and 5 show simulations of a small patch of the FoV for PSFs from each type of mask. Raw data for each PSF was simulated based on the model in Sec. 2.1. In Fig. 4, the same quantity of Gaussian noise was added to each simulated measurement, and in

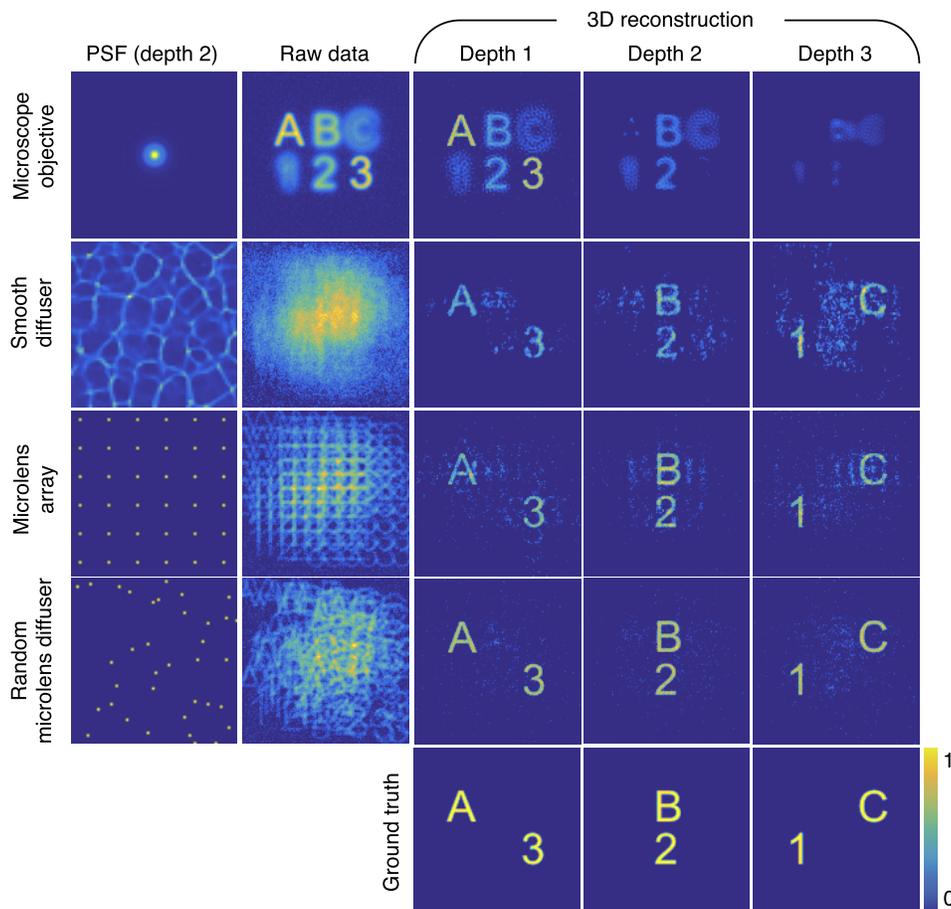


Fig. 4. Simulation comparing PSFs for depth-resolved imaging. A microscope objective has good noise performance but fails to capture 3D information. A smooth diffuser's PSF has significant background light causing noise amplification, as does the periodicity of the regular microlens array. Our random microlens diffuser has a non-periodic PSF with high contrast, resulting in good noise performance and 3D reconstructions. All simulations have the same quantity of Gaussian noise added to the raw data.

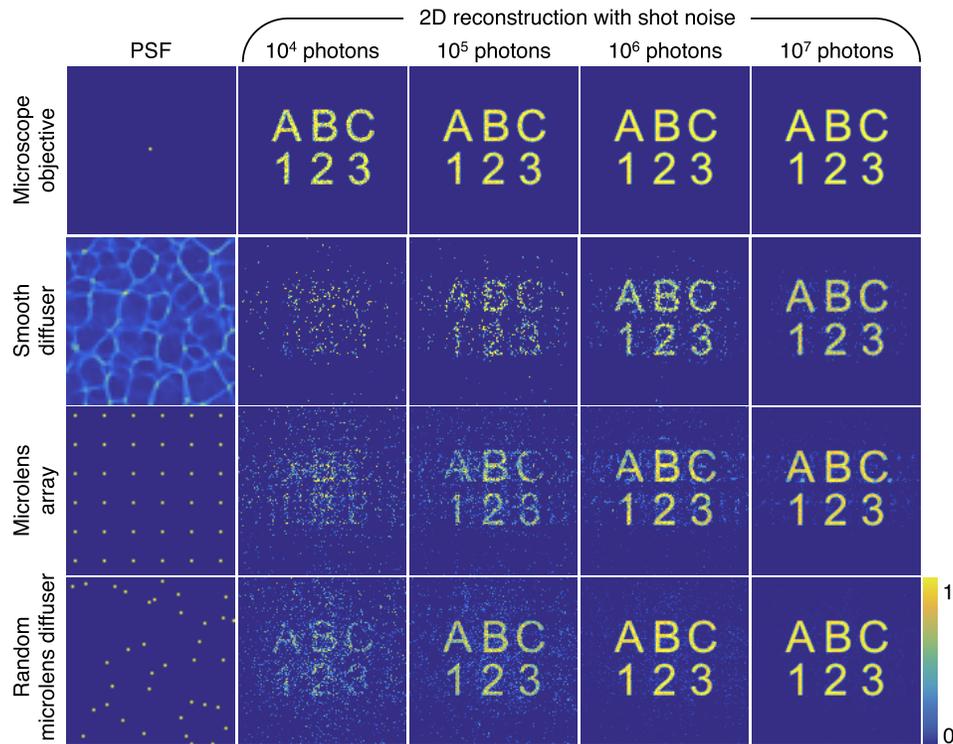


Fig. 5. Simulation comparing PSF robustness to shot noise at a single depth. A microscope objective has the best noise performance for 2D imaging, but it does not extend to depth-resolved imaging nor to miniaturized systems. Of the PSFs with 3D capabilities, the random microlens diffuser is most robust to shot noise.

Fig. 5, Poisson noise (shot noise) was simulated for varying number of collected photons. Each mask spreads photons differently over the sensor, so the number of photons per pixel at the sensor, and thus the shot noise performance, depends on the PSF. In both simulations, the noisy raw data was processed using the method described in Sec. 3.3. Background estimation was omitted and we assumed that the FoV was small enough that one calibration PSF per depth was sufficient.

Traditional microscope objectives are carefully optimized to capture high-quality images at the focal plane. They have good noise performance even at low photon counts for a 2D scene (Fig. 5, top row). However, recovering depth information from a single image is a poorly posed problem; as objects become defocused they lose high-frequency detail which may not be recovered, even with deconvolution (Fig. 4, top row).

Off-the-shelf diffusers [39] are convenient, inexpensive, and can easily be extended to larger sensor sizes. Lensless imagers made with these diffusers have a caustic pattern PSF (Fig. 4, second row). Due to the pseudorandom diffuser surface, any translation or scaling of the PSF *should* result in a substantially different pattern (i.e. the caustics should have a low inner product). However, the large amount of background light between the caustics causes an increased inner product which results in higher noise amplification during deconvolution (Figs. 4 and 5, second row). Other masks with low contrast patterns (e.g. amplitude masks, far field speckle) will suffer from similar noise amplification.

In comparison, a microlens array, which is widely used in light field microscopy, is designed to concentrate all incoming light into diffraction-limited spots beneath each lenslet, resulting in very little background light and a high-contrast pattern. However, if the PSF from a regularly-spaced

microlens array is translated by exactly one period, the shifted PSF is a duplicate of the on-axis PSF, resulting in an increased inner product and higher noise amplification (Figs. 4 and 5, third row). Notice that the reconstruction shows periodic ghosting due to the regularity of the microlens array, and these artifacts are present even in low noise scenarios.

In our system, we use a random microlens diffuser which combines the best properties of the phase masks described above. Like the regular microlens array, our microlens diffuser has a high contrast PSF with low background, and, like the smooth diffuser, our PSF is pseudorandom without periodic ambiguity. The result is reduced noise sensitivity compared to the other flat masks (Figs. 4 and 5, third row). Although a traditional microscope objective has better noise performance at a single plane, our system is better suited for miniaturization and enables reconstruction of 3D information from a single acquisition.

In addition, our microlens-based design is well-suited for resolution enhancement since the focal spot of each lenslet contains high spatial frequencies in all directions. Furthermore, it is easier to design diffraction-limited lenses when the diameter is small [40], allowing each lenslet to have nearly diffraction-limited performance with only a single spherical surface. Finally, since the microlenses focus light, the best performance is obtained when the object is in imaging condition with the sensor, so the lenslet focal length and distance to the sensor can be used to set a practical working distance, over 1.5 mm in our prototype.

5. Experimental results

We built a prototype system using a backside-illuminated monochrome CMOS sensor (UI-3862LE with Sony IMX290 chip) and two color filters (Kodak Wratten #12 and Chroma ET525/50m) designed for green fluorescent probes ($\lambda = 520$ nm). As described in [9], the combination of an absorption and an interference-based color filter is well-suited for removing excitation light at the high angles of incidence potentially present in our system, and any unfiltered light is removed with our computational background estimation (Sec. 3.3). We fabricated our random microlens diffuser with a droplet-based technique, similar to [41,42], since these methods are known for good surface quality. Drops of optical epoxy (Norland 63) were cured on a hydrophobic surface, then transferred onto a glass coverslip to form the diffuser, generating lenslets with approximately $p = 250$ μm diameter. The diffuser was index-matched with polydimethylsiloxane (PDMS) to increase the microlens focal length to about 1.5 mm, and it was placed $d = 3.8$ mm away from the sensor. Based on these physical parameters and the sampling requirements outlined in Sec. 3.4, we require lateral samples every 400 μm and axial samples with $1/z_1 - 1/z_2 \leq 33$ m^{-1} . Rather than sampling dioptrically, we choose to calibrate every 100 μm axially, which satisfies the axial sampling condition for objects $z = 1.7$ mm or further from the diffuser. Calibration images are captured with a 15 μm fluorescent bead, and the lowest frequency 10×10 DCT coefficients were set to zero for all calibration images before further processing to remove background light. Negative values after background subtraction were set to zero. For each calibration point, four measurements were taken and averaged to reduce noise. All images were downsampled by $2\times$ in each direction such that the equivalent pixel size is 5.8 μm .

To characterize the resolution of our system, we solve Eq. (9) on images of two points at varying separation distances, each generated by summing images of a single fluorescent bead. For axial characterization, images are of a 15 μm bead, moved in 10 μm increments; for lateral characterization, we use a 5 μm bead, moved in 2 μm increments. We define the resolution to be the minimum spacing at which there is at least a 20% dip in intensity between neighboring points in the reconstruction. Figure 6(a) summarizes our results, demonstrating 8 μm lateral resolution and under 50 μm axial resolution. We further test our system with a fluorescent USAF resolution target, $z = 2.56$ mm from the diffuser, shown in Fig. 6(b). We can clearly resolve group 6, element 1 with 7.8 μm bars, which matches our two-point resolution experiments and demonstrates an order of magnitude improvement over our previous work [19]. Brute-force calibration at every

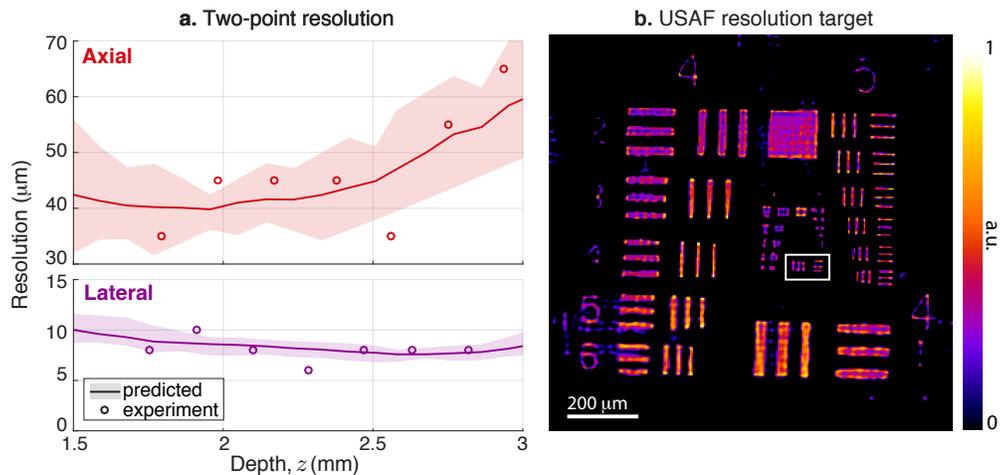


Fig. 6. Experimental resolution characterization. (a) Resolution is measured by determining the minimum separation distance at which two fluorescent beads are resolvable. The experimental resolution can be predicted by calculating inner products of a PSF with shifts and scales of itself; points are considered resolvable when the inner product is below 0.8. The predicted resolution is calculated at 12 different lateral locations in the FoV. The range of values is depicted by the filled area in the plot, and the solid line is the mean. (b) USAF resolution target shows group 6, element 1 containing $7.8 \mu\text{m}$ bars (boxed) is clearly resolvable, which matches the two-point resolution.

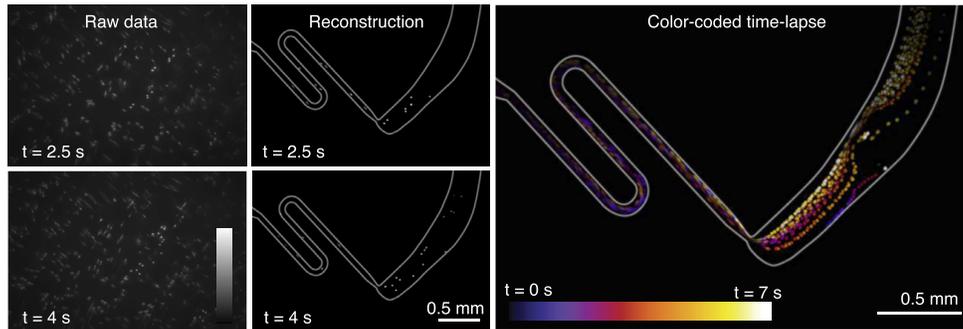
resolvable location in the volume would require Nyquist sampling the two-point resolution, necessitating samples every $4 \mu\text{m}$ laterally and every $25 \mu\text{m}$ axially. This is $100 \times 100 \times 4 = 40,000$ times more calibration measurements than with our sparse calibration scheme and local convolution model, demonstrating the large savings achieved with our model.

In addition, we show that we can predict the two-point resolution from the PSF measurements, without running a full reconstruction. To do this, we shift (for lateral resolution) or scale (for axial resolution) a central PSF and calculate the inner product with the original. A low inner product indicates that neighboring measurements are sufficiently different to be distinguished. For the noise levels in our system, we find that a normalized inner product of 0.8 is a good predictor of the resolution, plotted by the solid line in Fig. 6(a) (average over 12 field positions). This process can be used for system design by simulating PSFs (for example, using Fresnel propagation) then using this inner product metric to predict the final resolution.

Due to fabrication errors and the non-uniform distribution of lenslets, there will be variation between the focal spots under each microlens, which can result in resolution that depends on the object's lateral location. However, since each PSF includes focal spots from many lenslets (about 15 in our prototype), the effect of individual variations is averaged. To quantify this in our system, for each depth we calculate the predicted resolution at 12 locations in the FoV and plot the range of predicted values in Fig. 6(a). We find that the variation in resolving power is low for lateral resolution and modest for axial resolution. We believe resolution variation can be reduced substantially by fabricating the diffuser with more precise methods (e.g. injection moulding).

Since our method is single-shot, the frame rate is only limited by the sensor. To demonstrate, we capture a 10 fps video of $15 \mu\text{m}$ fluorescent beads flowing through a microfluidic channel, shown in Fig. 7(a). Beads were reconstructed at a single depth plane, $z = 2.42 \text{ mm}$, and the full video is available in [Visualization 1](#). We also test our system on a live 6-day-old NeuroD:GCaMP6f larval zebrafish [43] captured at 10 fps and reconstructed at a single depth plane, $z = 2.19 \text{ mm}$.

a. Fluorescent beads in microfluidic channel



b. Larval zebrafish expressing GCaMP

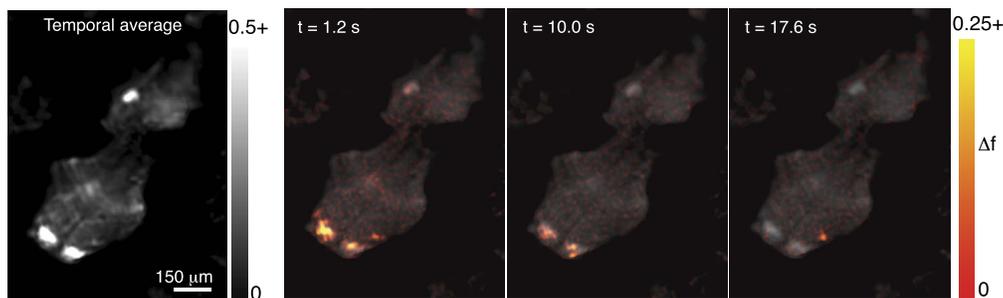


Fig. 7. Experimental videos captured with our diffuser microscope at 10 fps. (a) Fluorescent beads flowing in a microfluidic channel. Channel outlines are superimposed for visualization purposes, and the full video is in [Visualization 1](#). (b) NeuroD:GCaMP6f larval zebrafish, 6 days old. Change in fluorescence (Δf) compared to a 20th percentile baseline is shown in red and indicates neural activity.

Figure 7(b) shows the change in fluorescence (compared to a 20th percentile baseline). Our results qualitatively match the expected neural activity of a larval zebrafish. However, determining whether the reconstructed fluorescence signal is a linear function of the true fluorescence is still an open problem. Compressed sensing theory [37] proves that if the matrix \mathbf{A} fulfills the *restricted isometry property* and the sample is sufficiently sparse, then the signal can be recovered with perfect accuracy. Our design matrix \mathbf{A} is pseudorandom which is expected to fulfill the restricted isometry property with high probability, but the conditions are notoriously hard to verify, and a more rigorous proof of linearity in general cases is the subject of future work.

To highlight the 3D capacity of our system, we created sample containing layers of $15\ \mu\text{m}$ fluorescent beads separated by coverslips. The sample was reconstructed at the three depth planes containing beads, shown in Fig. 8. A focal stack from a traditional fluorescent microscope is shown to validate the bead locations. Note that, unlike with a traditional microscope, our prototype reconstructs the complete 3D distribution of beads from a single acquisition of raw data. Finally, since our system requires that the object be sparse for accurate reconstruction, we test on non-sparse samples to demonstrate that our system still captures the edges and sharp regions of dense samples. We image a fixed brine shrimp sample (Carolina Biological) stained with eosin and reconstructed at 10 depth planes spaced $100\ \mu\text{m}$ apart, processed from a single acquisition of raw data (Fig. 9). In the dense regions of the head of the brine shrimp, we find some inconsistencies between the traditional microscope focal stack and the diffuser microscope reconstruction. In regions where the sample is sparse, especially the shrimp's antennae, our reconstruction matches the 3D locations captured in the traditional microscope focal stack.

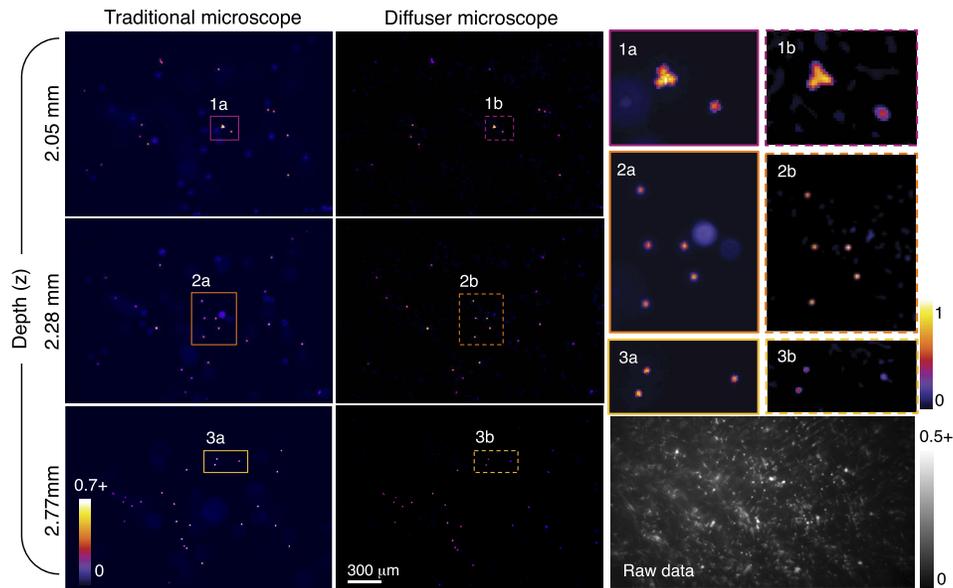


Fig. 8. 3D reconstruction of 15 μm fluorescent beads, axially separated by coverslips. The focal stack from a traditional fluorescence microscope (5 \times , 0.15 NA) is shown for comparison, with close-ups on the right. Our diffuser microscope reconstructs all depth planes from a single acquisition (bottom right) and removes out-of-focus light.

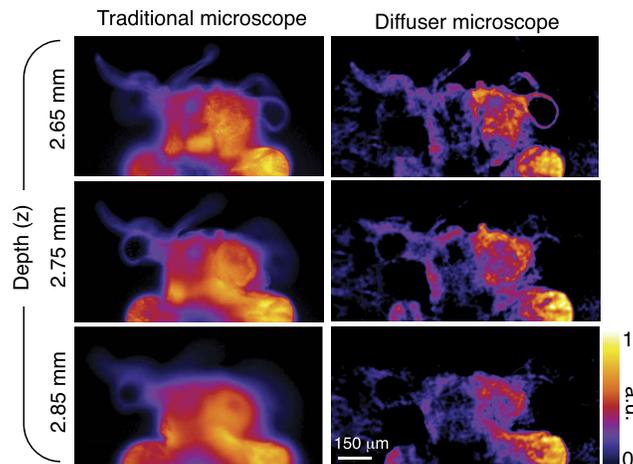


Fig. 9. 3D reconstruction of fixed brine shrimp tagged with eosin, shown at three different axial planes. The focal stack from a traditional fluorescence microscope (10 \times , 0.45 NA) is shown for comparison. Our diffuser microscope reconstructs 3D from a single acquisition and recovers the thin antenna structures at the correct depths.

6. Conclusion

We introduced a novel on-chip microscope which uses a random microlens diffuser to indirectly encode information about the scene which is then recovered by solving an optimization problem. The microscope has small form factor, is easy to assemble, and can capture 3D light distributions in a single acquisition. Brute force calibration would be impractical, so we propose a local

convolution model which requires $40,000\times$ fewer calibration measurements and provides an efficient computational framework. Our device is single-shot, capturing 3D videos at the frame rate of the sensor. We anticipate that our system could be useful for field work requiring portable microscopes, *in vivo* monitoring of fluorescent signals, and 3D tracking of biological organisms.

Funding

National Defense Science and Engineering Graduate Fellowship; National Science Foundation (1617794, DMR 1548924); Alfred P. Sloan Foundation; David and Lucile Packard Foundation.

Acknowledgments

The authors would like to thank Dr. Ehud Isacoff, Nick Antipa, and Pratul Srinivasan for helpful discussions, and Lizzy Griffiths for the zebrafish drawing (Fig. 1) in this manuscript.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. A. Greenbaum, W. Luo, T. Su, Z. Göröcs, L. Xue, S. O. Isikman, A. F. Coskun, O. Mudanyali, and A. Ozcan, "Imaging without lenses: achievements and remaining challenges of wide-field on-chip microscopy," *Nat. Methods* **9**(9), 889–895 (2012).
2. O. Mudanyali, D. Tseng, C. Oh, S. O. Isikman, I. Sencan, W. Bishara, C. Oztoprak, S. Seo, B. Khademhosseini, and A. Ozcan, "Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications," *Lab Chip* **10**(11), 1417–1428 (2010).
3. W. Bishara, T. Su, A. F. Coskun, and A. Ozcan, "Lensfree on-chip microscopy over a wide field-of-view using pixel super-resolution," *Opt. Express* **18**(11), 11181–11191 (2010).
4. G. Zheng, S. A. Lee, Y. Antebi, M. B. Elowitz, and C. Yang, "The ePetri dish, an on-chip cell imaging platform based on subpixel perspective sweeping microscopy (SPSM)," *Proc. Natl. Acad. Sci.* **108**(41), 16889–16894 (2011).
5. X. Cui, L. M. Lee, X. Heng, W. Zhong, P. W. Sternberg, D. Psaltis, and C. Yang, "Lensless high-resolution on-chip optofluidic microscopes for caenorhabditis elegans and cell imaging," *Proc. Natl. Acad. Sci.* **105**(31), 10670–10675 (2008).
6. E. P. Papageorgiou, H. Zhang, B. E. Boser, C. Park, and M. Anwar, "Angle-insensitive amorphous silicon optical filter for fluorescence contact imaging," *Opt. Lett.* **43**(3), 354–357 (2018).
7. A. F. Coskun, I. Sencan, T. Su, and A. Ozcan, "Lensfree fluorescent on-chip imaging of transgenic caenorhabditis elegans over an ultra-wide field-of-view," *PLoS One* **6**(1), e15955 (2011).
8. A. F. Coskun, I. Sencan, T. Su, and A. Ozcan, "Lensless wide-field fluorescent imaging on a chip using compressive decoding of sparse objects," *Opt. Express* **18**(10), 10510–10523 (2010).
9. K. Sasagawa, A. Kimura, M. Haruta, T. Noda, T. Tokuda, and J. Ohta, "Highly sensitive lens-free fluorescence imaging device enabled by a complementary combination of interference and absorption filters," *Biomed. Opt. Express* **9**(9), 4329–4344 (2018).
10. J. K. Adams, V. Boominathan, B. W. Avants, D. G. Vercosa, F. Ye, R. G. Baraniuk, J. T. Robinson, and A. Veeraraghavan, "Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope," *Sci. Adv.* **3**(12), e1701548 (2017).
11. A. Singh, G. Pedrini, M. Takeda, and W. Osten, "Scatter-plate microscope for lensless microscopy with diffraction limited resolution," *Sci. Rep.* **7**(1), 10687 (2017).
12. M. S. Asif, A. Ayremlou, A. Veeraraghavan, R. Baraniuk, and A. Sankaranarayanan, "FlatCam: Replacing lenses with masks and computation," in *Computer Vision Workshop (ICCVW), 2015 IEEE International Conference on*, (IEEE, 2015), pp. 663–666.
13. G. Kuo, N. Antipa, R. Ng, and L. Waller, "DiffuserCam: Diffuser-based lensless cameras," in *Computational Optical Sensing and Imaging*, (Optical Society of America, 2017), pp. CTu3B–2.
14. K. Tajima, T. Shimano, Y. Nakamura, M. Sao, and T. Hoshizawa, "Lensless light-field imaging with multi-phased fresnel zone aperture," in *2017 IEEE International Conference on Computational Photography (ICCP)*, (2017), pp. 76–82.
15. D. G. Stork and P. R. Gill, "Optical, mathematical, and computational foundations of lensless ultra-miniature diffractive imagers and sensors," *Int. J. on Adv. Syst. Meas.* **7**, 4 (2014).
16. J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, "Thin observation module by bound optics: concept and experimental verification," *Appl. Opt.* **40**(11), 1806–1813 (2001).

17. R. Horisaki, S. Irie, Y. Ogura, and J. Tanida, "Three-dimensional information acquisition using a compound imaging system," *Opt. Rev.* **14**(5), 347–350 (2007).
18. K. Fergus, A. Torralba, and W. T. Freeman, "Random Lens Imaging," Tech. rep., Massachusetts Institute of Technology (2006).
19. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," *Optica* **5**(1), 1–9 (2018).
20. N. Antipa, P. Oare, E. Bostan, R. Ng, and L. Waller, "Video from stills: Lensless imaging with rolling shutter," in *2019 IEEE International Conference on Computational Photography (ICCP)*, (IEEE, 2019), pp. 1–8.
21. F. L. Liu, V. Madhavan, N. Antipa, G. Kuo, S. Kato, and L. Waller, "Single-shot 3D fluorescence microscopy with Fourier DiffuserCam," in *Novel Techniques in Microscopy*, (Optical Society of America, 2019), pp. NS2B–3.
22. K. Yanny, N. Antipa, R. Ng, and L. Waller, "Miniature 3D fluorescence microscope using random microlenses," in *Optics and the Brain*, (Optical Society of America, 2019), pp. BT3A–4.
23. E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Machine Intell.* **14**(2), 99–106 (1992).
24. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep. CSTR* **2**, 1–11 (2005).
25. M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-D deconvolution for the light field microscope," *Opt. Express* **21**(21), 25418–25439 (2013).
26. G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.* **7**(1), 821–825 (1908).
27. H. E. Ives, "Parallax panoramagrams made with a large diameter lens," *J. Opt. Soc. Am.* **20**(6), 332–342 (1930).
28. K. Kagawa, K. Yamada, E. Tanaka, and J. Tanida, "A three-dimensional multifunctional compound-eye endoscopic system with extended depth of field," *Electron. Comm. Jpn.* **95**(11), 14–27 (2012).
29. G. Kim, K. Isaacson, R. Palmer, and R. Menon, "Lensless photography with only an image sensor," *Appl. Opt.* **56**(23), 6450–6456 (2017).
30. A. B. Zoubi, K. S. Alguri, G. Kim, V. J. Mathews, R. Menon, and J. B. Harley, "Fast imaging in cannula microscope using orthogonal matching pursuit," in *2015 IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE)*, (IEEE, 2015), pp. 214–219.
31. X. Xie, H. Zhuang, H. He, X. Xu, H. Liang, Y. Liu, and J. Zhou, "Extended depth-resolved imaging through a thin scattering medium with PSF manipulation," *Sci. Rep.* **8**(1), 4585 (2018).
32. P. Berto, H. Rigneault, and M. Guillon, "Wavefront sensing with a thin diffuser," *Opt. Lett.* **42**(24), 5117–5120 (2017).
33. N. Teranishi, H. Watanabe, T. Ueda, and N. Sengoku, "Evolution of optical structure in image sensors," in *2012 International Electron Devices Meeting*, (IEEE, 2012), pp. 24–1.
34. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009).
35. D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006).
36. E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008).
37. E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *CR Math.* **346**(9–10), 589–592 (2008).
38. 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.
39. Luminit, "Technical Data and Downloads," <http://www.luminitco.com/downloads/data-sheets> (2017). [Online; accessed 19-July-2018].
40. A. W. Lohmann, "Scaling laws for lens systems," *Appl. Opt.* **28**(23), 4996–4998 (1989).
41. D. MacFarlane, V. Narayan, J. Tatum, W. Cox, T. Chen, and D. Hayes, "Microjet fabrication of microlens arrays," *IEEE Photonics Technol. Lett.* **6**(9), 1112–1114 (1994).
42. T. Kamal, R. Watkins, Z. Cen, J. Rubinstein, G. Kong, and W. M. Lee, "Design and fabrication of a passive droplet dispenser for portable high resolution imaging system," *Sci. Rep.* **7**(1), 41482 (2017).
43. P. Rupprecht, A. Prendergast, C. Wyart, and R. W. Friedrich, "Remote z-scanning with a macroscopic voice coil motor for fast 3D multiphoton laser scanning microscopy," *Biomed. Opt. Express* **7**(5), 1656–1671 (2016).