Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/12/e1701548/DC1

Supplementary Materials for

Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope

Jesse K. Adams, Vivek Boominathan, Benjamin W. Avants, Daniel G. Vercosa, Fan Ye, Richard G. Baraniuk, Jacob T. Robinson, Ashok Veeraraghavan

Published 8 December 2017, *Sci. Adv.* **3**, e1701548 (2017)
DOI: 10.1126/sciadv.1701548

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at advances.sciencemag.org/cgi/content/full/3/12/e1701548/DC1)

- movie S1 (.mp4 format). Digital focusing of simulated resolution target.
- movie S2 (.mp4 format). Fluorescent beads flowing in microfluidic channels of different depths.

SUPPLEMENTARY MATERIALS

section S1. Derivation of T2S model

Mask design

Two-dimensional (2D) separable mask patterns are constructed via the outer product of two 1D sequences, $m_1$and m_2 of length N, where each entry in the sequences is either -1 or $+1$. In the resulting 2D matrix, -1 entries are assigned to closed apertures of the mask and $+1$ entries are assigned to open apertures of the mask (*14*, *16*). Since closed apertures block light, they are instead assigned value 0 in the 2D matrix. The 2D mask pattern can then be written as the following matrix of dimensions $N \times N$

$$M = \frac{1+m_1 m_2^T}{2} \tag{S1}$$

Imaging model

Image formation for FlatScope can be described as follows. Let the scene, mask, and sensor be at the distances shown in fig. S7A, where d_1 is the distance between scene and mask, and d_2 is the distance between mask and sensor. Let the discretized scene be written as X, discretized mask as M and discretized sensor measurement as Y. For convenience, we will consider these quantities as matrices and index them by rows and columns. For example, $X(u, v)$ is the u^{th} row and v^{th} column of X. For convenience, we will assume that X, M, Y are of the same dimension $N \times N$. In the following paragraphs, we will derive the image formation model for a scene at a single depth and simplify the notations by dropping the subscript d in Eq. 1 of main text.

When the scene element $X(u, v)$ is active (or illuminated), the sensor records a magnified version of the mask centered around $M(u, v)$ and scaled by the scene intensity $X(u, v)$. The magnification $\alpha = \frac{d_1+d_2}{d_1}$ can be calculated using similar triangles as shown in fig. S7A. An additional term that needs to be considered is the sensor pixel's response to light rays at different angles. Figure S7B shows the pixel's response profile (PRP). Attenuation due to the pixel's response causes the formation of a local pattern on the sensor when a point source, such as a fluorescent bead, is imaged by FlatScope (see Main Text, Fig. 2A,B). We will denote the pixel's response by C. The sensor measurement at (u', v') can then be written as

$$Y(u', v') = \sum_u \sum_v X(u, v)\, M\left(u + \frac{u'-u}{\alpha}, v + \frac{v'-v}{\alpha}\right)\, C(u' - u, v' - v)\,. \tag{S2}$$

The mask design procedure above enables us to write the following

$$M\left(u + \frac{u'-u}{\alpha}, v + \frac{v'-v}{\alpha}\right) = \frac{1}{2} + \frac{m_1\left(u + \frac{u'-u}{\alpha}\right)\, m_2\left(v + \frac{v'-v}{\alpha}\right)}{2}$$

Since the pixels are rectangular, each pixel's response is separable as well

$$C(u' - u, v' - v) = c_1(u' - u)\, c_2(v' - v)$$

where c_1 is the pixel's response along rows and c_2 is the pixel's response along columns (as show in fig. S7B). Substituting the above two equations in eq. S2 yields

$$Y(u',v') = \sum_u \sum_v \left[\frac{1}{2} c_1(u'-u) X(u,v) c_2(v'-v) + \frac{1}{2} c_1(u'-u)\, m_1\left(u + \frac{u'-u}{\alpha}\right) X(u,v)\, m_2\left(v + \frac{v'-v}{\alpha}\right) c_2(v'-v) \right]$$

$$\Rightarrow Y(u',v') = \frac{1}{2}\sum_u c_1(u'-u) \left[\sum_v X(u,v) c_2(v'-v) \right] + \frac{1}{2}\sum_u c_1(u'-u)\, m_1\left(u + \frac{u'-u}{\alpha}\right) \left[\sum_v X(u,v)\, m_2\left(v + \frac{v'-v}{\alpha}\right) c_2(v'-v) \right]$$

Since the summation along the rows and summation along the columns have separated in the above equation, we can rewrite it concisely as

$$Y = P_o X Q_o^T + P_c X Q_c^T \tag{S3}$$

where P_o and Q_o implement the effects of c_1 and c_2, respectively, and P_c and Q_c implement the effects of m_1 and m_2, respectively. Equation S3 shows that the sensor measurements can be written as a superposition of two separable functions. We call this superposition the **Texas Two-Step model (T2S)**.

Note that P_o and Q_o are not functions of the mask and, in fact, the first term in eq. S3 models the effect when an open (or no) mask is placed in front of the sensor. On the other hand, P_c and Q_c are functions of the mask pattern (i.e., m_1, m_2) and hence the second separable term in eq. S3 models the effect due to the coding of the mask. The subscripts *o* and *c* refer to "open" and "coding", respectively. This interpretation is visually shown in Fig. 2. Also note that, P_o and Q_o have positive entries (since the sensor pixel's response is always positive), while P_c and Q_c have both positive and negative entries.

The 1D sequences m_1 and m_2 are chosen to have half of their entries equal to $-1/+1$. Therefore, m_1 and m_2 are orthogonal to a sequence with all $+1$ entries. Consequently, corresponding columns of P_o (no mask component) and P_c (coding component) are orthogonal. The same is true for corresponding columns of Q_o and Q_c. We exploit this fact to expedite the calibration process (see Model calibration).

section S2. Computational tractability

There are two aspects to consider when discussing computational tractability: memory requirements and runtime. Either large memory requirement or long run time is undesirable in a computational imaging system, particularly if real time processing is desired. In the following discussion, we will show that a separable mask that conforms to the T2S model requires far less memory and run time as compared to an arbitrary mask.

First, we consider the advantages of the T2S model in terms of the total number of parameters in the linear mapping from the scene to the measurements. This both reduces the effort for system calibration and speeds up the image reconstruction run time. To be general, consider a scene X and sensor measurement Y as matrices of size (or dimensions) $M \times N$. For an arbitrary amplitude mask, the following generalized linear model holds

$$y = \Phi x \tag{S4}$$

where, x and y are the vectorized versions of X and Y, respectively, formed by concatenating the columns of each matrix into a single, long vector. Both y and x are of length MN, while Φ is $MN \times MN$. Therefore, Φ contains M^2N^2 or $O(N^4)$ elements if $M \cong N$.

In contrast, a separable mask yields to the T2S model

$$Y = P_o X Q_o^T + P_c X Q_c^T$$

where, P_o and P_c are each of size $M \times M$ and Q_o and Q_c are each of size $N \times N$. This reduces the total number of elements in the mapping from the scene to measurements to $2(M^2 + N^2)$ or $O(N^2)$ if $M \cong N$.

Second, we consider the advantages of the T2S model in terms of the amount of memory required to store the parameters in the linear mapping from the scene to the measurements. For the FlatScope prototype presented here, M = 1000 and N = 1300. Hence, for an arbitrary (i.e., non-separable) mask, the matrix Φ will contain 1.7 trillion elements. To be robust to quantization noise, we represent each element as a 32-bit floating point datatype. Therefore, storing the matrix Φ would require 6 TB of memory, which is beyond the capabilities of both commercially available desktop computers and memory optimized cloud computing services (e.g., Amazon EC2 X1 currently offers memory of 2TB). This necessitates breaking Φ into smaller chunks for computation, which drastically increases the data communication overhead between storage and RAM, thereby superseding any hardware speedups, yielding intractable run times of several weeks to months. Needless to say, multi-depth reconstruction that requires multiple Φ matrices becomes practically impossible.

By using a separable mask, the total memory requirement for the matrices (P_o, Q_o, P_c, Q_c) in the T2S model is just 21 MB, a savings of five orders of magnitude over a non-separable approach. Due to low memory usage and reduced size of matrices, the gradient steps required for the iterative image reconstruction algorithms (see Reconstruction algorithms in Main Text) can be computed repeatedly with low computational cost. Additionally, we can use a parallel implementation on a GPU, which usually has less memory than a CPU. With a sub-optimal Nvidia Tesla GK210 GPU implementation using MATLAB, we achieved a single depth reconstruction in under 10 s and a 41-layer 3D reconstruction in under 15 min. An optimized GPU implementation will provide an even larger speedup.

section S3. Model calibration

FlatScope calibration relies on an observation that if the scene is separable (rank-1), then the FlatScope measurement is rank-2. For example, if the scene has only the i^{th} row active (or illuminated), then the

scene can be written as $X_i = e_i \mathbf{1}^T$, where e_i is a sequence of zeros with only the i^{th} element to be 1 and $\mathbf{1}^T$ is a sequence of all 1s. Then the FlatScope measurement can be written as

$$Y_i = (P_o e_i)(Q_o \mathbf{1})^T + (P_c e_i)(Q_c \mathbf{1})^T = p_{oi}\, q_o^T + p_{ci}\, q_c^T$$

Here, p_{oi} and p_{ci} are the i^{th} columns of P_o and P_c, respectively, and q_o and q_c are the sums of columns of Q_o and Q_c, respectively. As noted in the section S1, p_{oi} and p_{ci} are orthogonal and can be computed (up to a scaling factor) via the Singular Value Decomposition (SVD) of Y_i truncated to the two largest singular values. Since the sensor measurements are always positive, the truncated SVD of Y_i yields one vector with all positive entries and another vector with both positive and negative entries. The positive vector is assigned to p_{oi}, and the other vector is assigned to p_{ci}. By scanning the rows of the scene, we can compute all the entries in P_o and P_c. Similarly, the columns of Q_o and Q_c can be calibrated by scanning along the columns of the scene. Scanning the rows and columns of the scene is physically done by translating a line slit (see Methods), as shown in fig. S1. The transfer functions $\{P_o, Q_o, P_c, Q_c\}$ are dependent on the distance of the scene (d_1) and can be calibrated for each depth by first translating the line slit to the required depth and then scanning the FOV.

Note that the number of calibration images needed is equal to the sum of the number of columns and number of rows of scene X. If the scene is of size $N \times N$, then the number of calibration images needed is $2N$. This is far less than the N^2 number of calibration images needed for a generalized linear model mentioned in eq. S4.

section S4. Gradient direction for iterative optimization

Here we show the gradient direction needed for optimization problems mentioned in the methods section of the main text.

For single depth reconstruction, let us define the forward operator $A_d(\cdot)$ and a transpose operator $A_d^T(\cdot)$ as follows

$$A_d(X_d) = P_{od} X_d Q_{od}^T + P_{cd} X_d Q_{cd}^T, \qquad A_d^T(Y) = P_{od}^T Y Q_{od} + P_{cd}^T Y Q_{cd}$$

Then the gradient direction for optimization Eq. 3 in the main text is

$$\nabla = -2A^T(Y) + 2A^T\big(A(X_d)\big) + 2\lambda_2 X_d$$

The gradient direction for optimization Eq. 4 in the main text is

$$\nabla = -2A^T(Y) + 2A^T\big(A(X_d)\big)$$

For 3D reconstruction, let us define the forward operator $A_D(\cdot)$ and a transpose operator $A_D^T(\cdot)$ as follows

$$A_D(X_D) = \sum_{d=1}^{D} (P_{od} X_d Q_{od}^T + P_{cd} X_d Q_{cd}^T)$$

$$A_D^T(Y) = \sum_{d=1}^{D} (P_{od}^T Y Q_{od} + P_{cd}^T Y Q_{cd})$$

Then the gradient direction for optimization Eq. 5 in the main text is

$$\nabla = -2A_D^T(Y) + 2A_D^T\big(A_D(X_D)\big)$$

section S5. MURA: A separable pattern

Modified Uniformly Redundant Array (MURA) (*27*) has many attractive properties that have motivated the pattern's use in astronomy, fast neutron, and gamma-ray imaging. Here we show that MURA is separable and hence can be used as a mask pattern for FlatScope.

Let M represent the two-dimensional MURA pattern. Given a prime number p of the form $p = 4m + 1$, MURA is of size $p \times p$. Let $i = 0, \dots, p-1$ index the rows and $j = 0, \dots, p-1$ index the columns of M. Then the MURA pattern can be written as(*27*)

$$M_{ij} = \begin{cases} 0 & \text{if } i = 0, \\ 1 & \text{if } j = 0, i \neq 0, \\ 1 & \text{if } C_i C_j = +1, \\ 0 & \text{if } C_i C_j = -1, \end{cases}$$

where

$$C_q = \begin{cases} +1 & \text{if } q \text{ is a quadratic residue modulo } p, \\ -1 & \text{otherwise.} \end{cases}$$

An integer q is called a quadratic residue modulo p if there exists an integer x such that: $x^2 \bmod p = q \bmod p$.

To make MURA pattern M separable, we drop the first row ($i = 0$) and the first column ($j = 0$) and redefine M of size $(p-1) \times (p-1)$ as

$$M_{ij} = \begin{cases} 1 & \text{if } C_i C_j = +1, \\ 0 & \text{if } C_i C_j = -1. \end{cases}$$

If we define $C_I = \{C_i\}_{i=1}^{p}$ and $C_J = \{C_i\}_{i=1}^{p}$, M can then be written in the separable form of eq. S1 as follows

$$M = \frac{1 + C_I C_J^T}{2}$$

section S6. FlatCam model: An approximation to the T2S model

Given a separable mask, the T2S model applies at all scene distances. At large working distances, it can be simplified to the FlatCam (*14*, *16*) imaging model, which features only one and not two separable terms. At large working distances, each scene element acts like a source of plane waves resulting in a global effect on the sensor as opposed to a localized effect. The global effect renders the open mask component ($P_o X Q_o^T$ term in eq. S3) of the T2S model as a scene-dependent constant. This constant can be subtracted using a heuristic method (like averaging), yielding a single separable model: $Y_{\text{FlatCam}} = P_c X Q_c^T$ (second term in eq. S3). FlatCam's single separable model, as proposed by Asif *et al.* (*14*), is an approximation to the T2S model and holds true only for large scene distances. The invariance of the T2S model to scene distance and the restricted capability of the FlatCam model are shown through simulation in fig. S3. Note that the simulated scene in the figure is fairly dense; for a sparser scene, the distance for a quality reconstruction using FlatCam would be far greater than the 2.5 mm shown.

section S7. Diffraction effects and T2S model

The derivation of the T2S model in section S1 is based on geometric optics; but at the scale of our mask features, diffraction effects from wave optics can have a significant impact. To quantify the errors caused when not incorporating wave optics, we compare the raw point spread function (PSF) of our prototype with the T2S approximation of the PSF. The comparison is shown in fig. S10. The raw PSF is captured by illuminating a 5 μm precision pinhole (Edmunds Optics, #38-537) with a wide-angle diffuser. The pinhole acts like a point source when illuminated with a diffuser and is placed at a depth 250 μm from the FlatScope prototype. We see from fig. S10 that there is a small difference between the raw captured PSF and the T2S approximation, with an average error of 12.6%. In our experiments, we use the raw captured images for reconstruction. Notwithstanding the geometrical approximation, we can reliably reconstruct scenes (as shown in Figs. 1,3,4 and 5) using the T2S model.

section S8. Light collection

Fluorophores in samples often provide only a small amount of light, which lensed systems focus onto the sensor. We consider light collection for FlatScope compared to traditional lensed microscopes with comparable resolution capabilities (fig. S11). We calculate $Light\ collection\ \% = 0.5[1 - \cos(\theta)]$, where $\theta = \sin^{-1}(NA)$ for the microscope objectives and θ equal to FWHM/2 of the PRP for FlatScope. For microscope objectives, use of a GFP emission filter was assumed, and for FlatScope the 50% closed aperture amplitude mask and absorption filter were taken into consideration (509 nm wavelength used for calculations). It should be noted, that although FlatScope theoretically performs better than the lensed based systems, these calculations do not account for noise introduced during reconstruction which can degrade the image.

section S9. Impact of sensor saturation

In the absence of a focusing lens, the light from any point in a fluorescent sample is spread across multiple pixels. Hence, for a given exposure duration (those common in fluorescence microscopy), a lensed system might saturate, but FlatScope would likely not. If saturation occurs in a specific region,

the artifacts of reconstruction are localized around that region and don’t affect farther locations. In practice, we want to avoid any nonlinear effects like saturation since our model and reconstructions are based on a linear sensor response. In the future, we may borrow ideas from high dynamic range (HDR) photography to combine multi-exposure captures to ensure quality reconstruction for very high contrast samples.

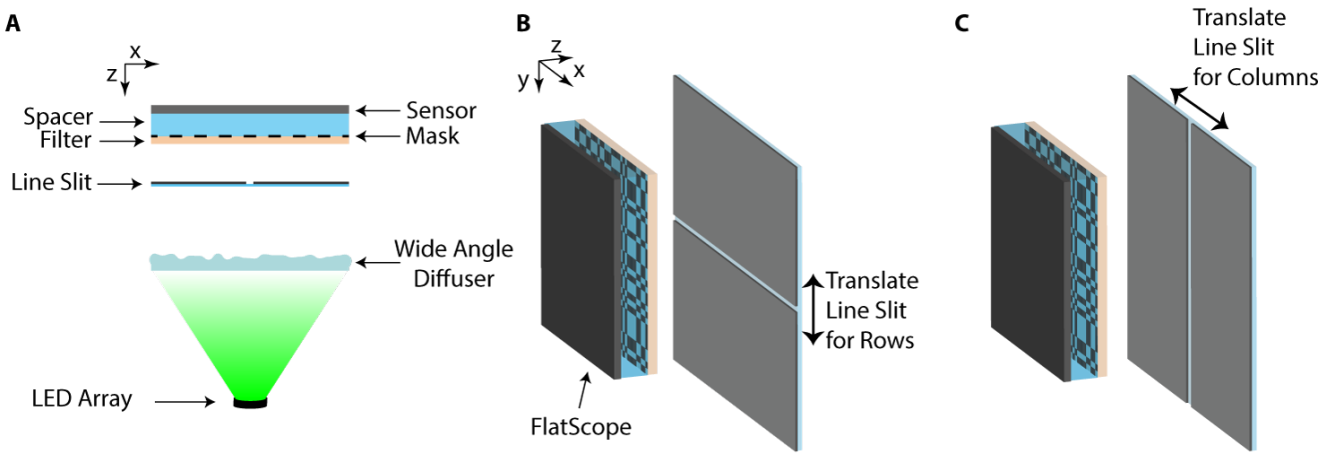

fig. S1. Calibration setup. (**A**) Side view of the calibration setup (not to scale) showing the LED array, wide angle diffuser (~10 cm above the LEDs), target with line slit (~1 cm above the diffuser) and FlatScope (filter, mask, spacer & sensor). (**B**) Calibration of rows by translating the horizontal slit target along the y-axis. (**C**) Calibration of columns by translating the vertical slit target along the x-axis.

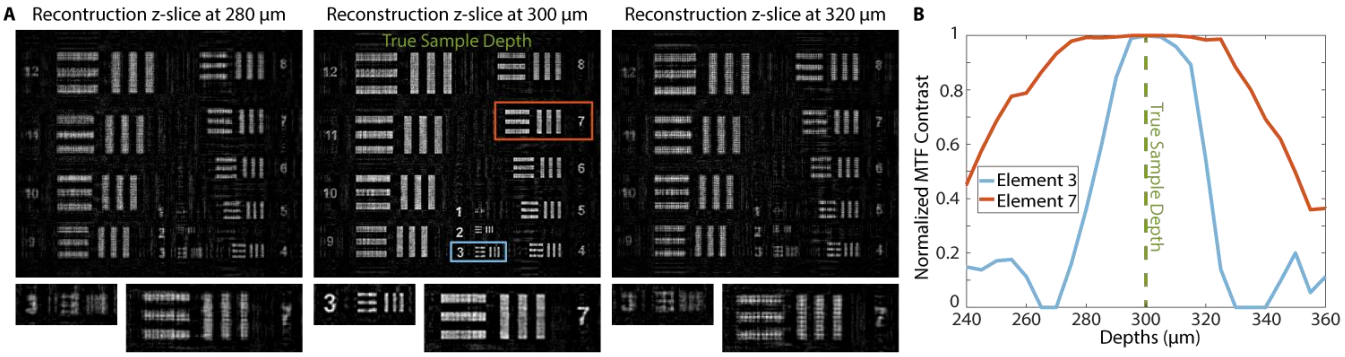

fig. S2. Digital focusing. We perform, by simulation, a focal stack reconstruction of a resolution target placed at 300 µm from FlatScope. (**A**) Three z-slices from the focal stack are shown, 280 µm, 300 µm and 320 µm. We observe that the reconstructed z-slice appears sharpest at the true depth (300 µm) of the target. (**B**) Modulation Transfer Function (MTF) plot for elements 3 and 7. Note that the MTF peaks at the true sample depth.

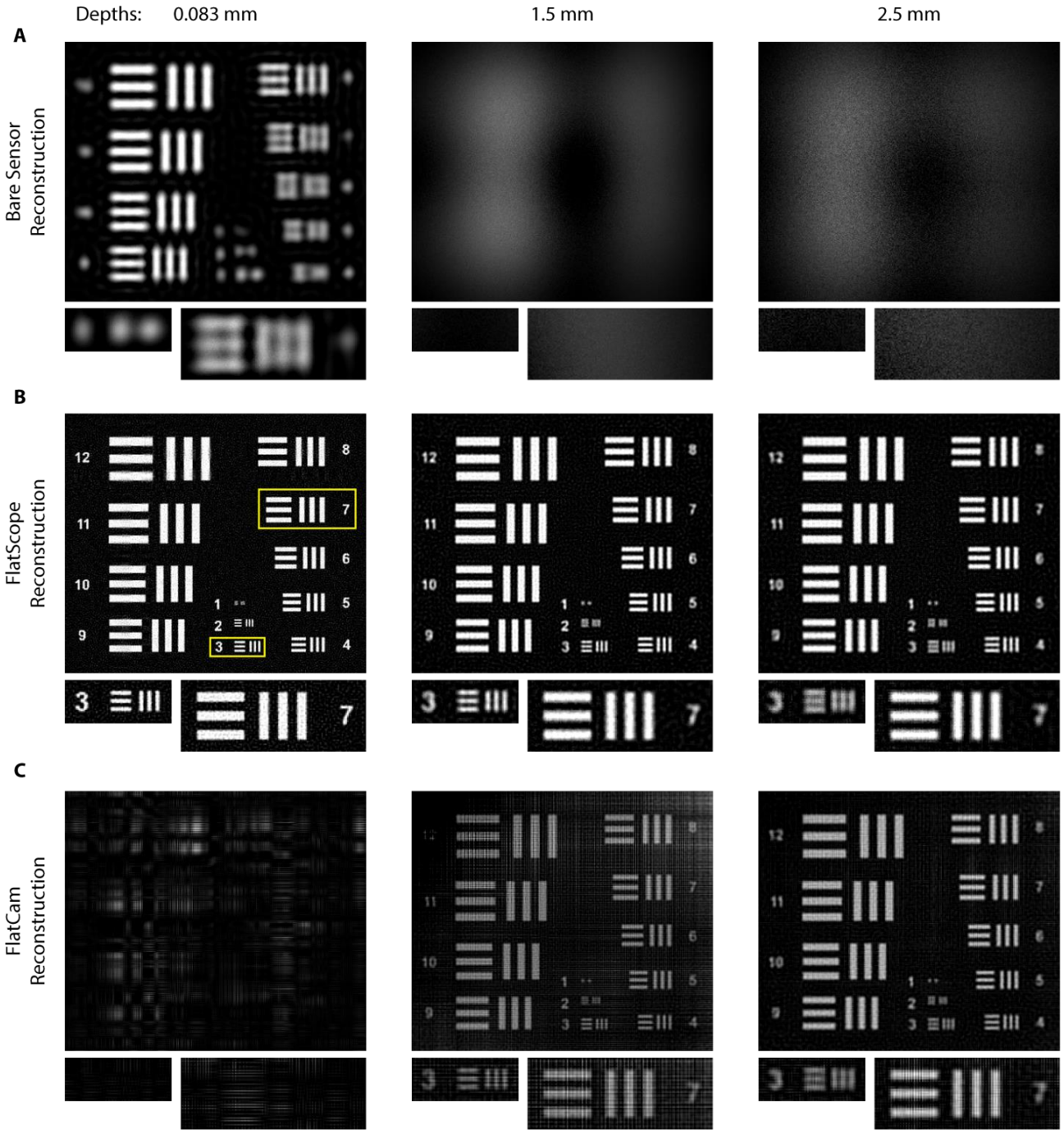

fig. S3. Simulation comparison of bare sensor, FlatScope, and FlatCam. Simulated 2D plane reconstructions are shown at depths 0.083 mm, 1.5 mm and 2.5 mm from respective systems. Images are reconstructed at the known depths. (**A**) Reconstruction of bare sensor decays rapidly with increasing depth. (**B**) Reconstruction of FlatScope shows high resolution and stays fairly stable with increasing depth. (**C**) FlatCam is unable to reconstruct at smaller depths due to limitations of the model with a single separable term. On the other hand, the T2S model of FlatScope can handle all depths. At larger depths, FlatCam and FlatScope reconstructions are comparable.

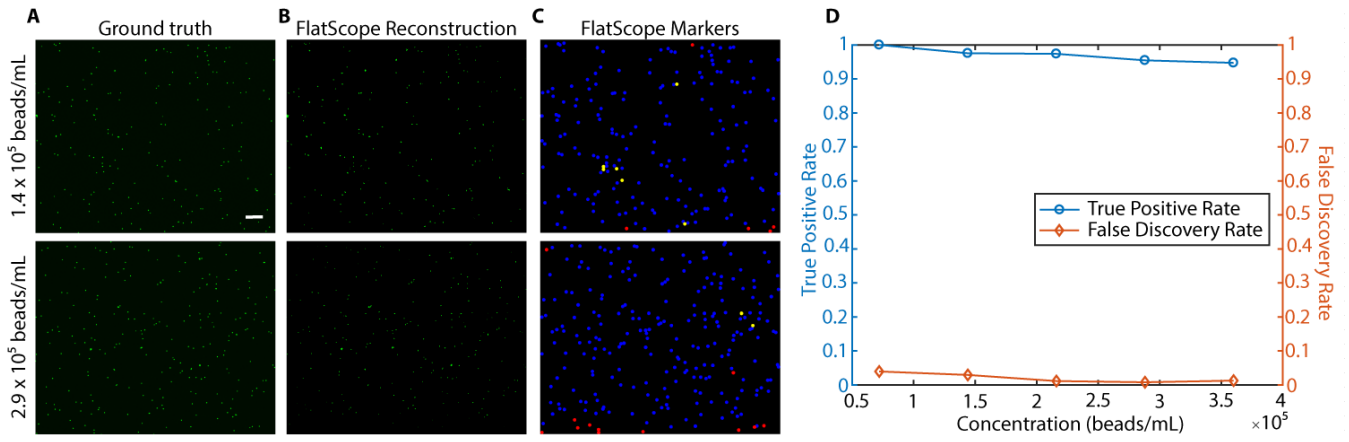

fig. S4. 3D volume reconstruction accuracy. 3D reconstruction of fluorescent beads in PDMS films at different concentrations. (**A**) Maximum projection of the ground truth captured with a confocal microscope (10x objective). (**B**) Maximum projection of FlatScope 3D reconstruction. (**C**) Markers for true positives (blue), misses (red), and false discoveries (yellow) for FlatScope reconstructions. Note that some reconstructed beads appear smaller than ground truth due to regularization. Scale bar, 200 μm. (**D**) Plot of true positive rate and false discovery rate as a function of concentration of fluorescent beads.

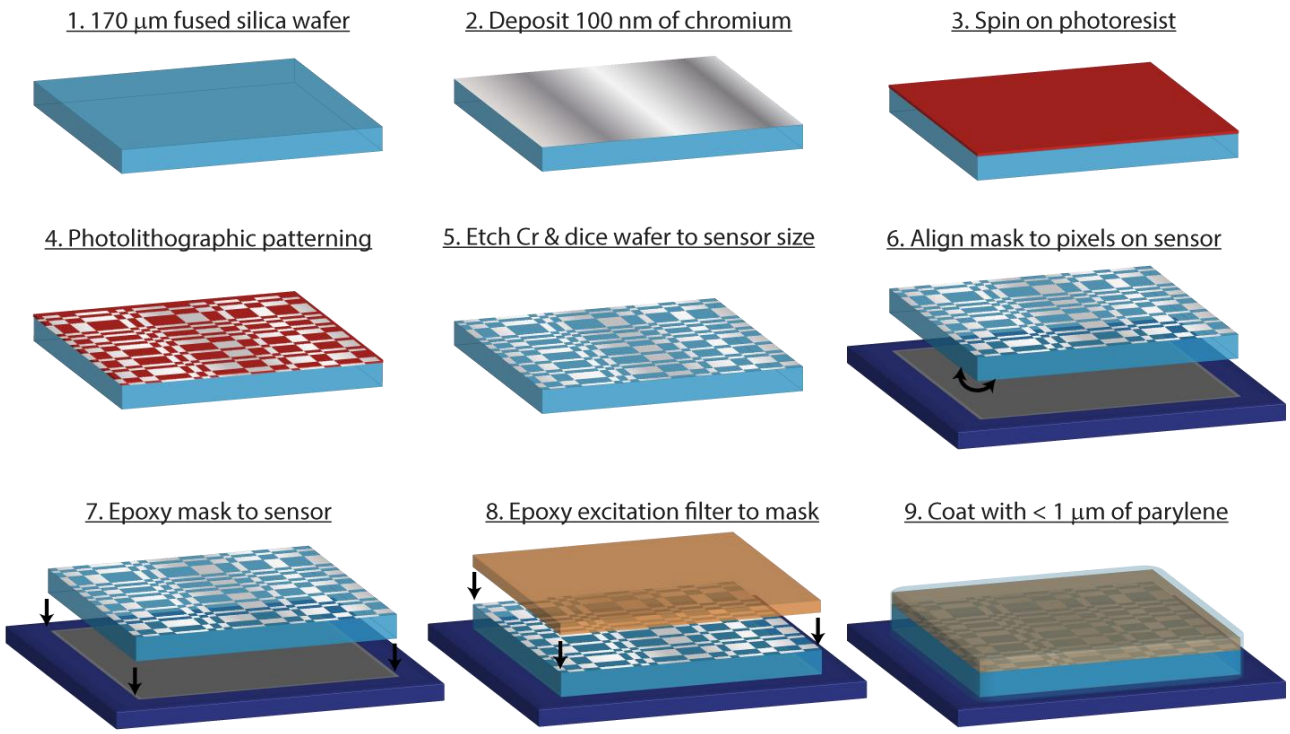

fig. S5. Fabrication of FlatScope. Steps for fabricating the amplitude mask and spacer, aligning and fixing to the imaging sensor, adding the absorptive filter and insulating/protecting the device.

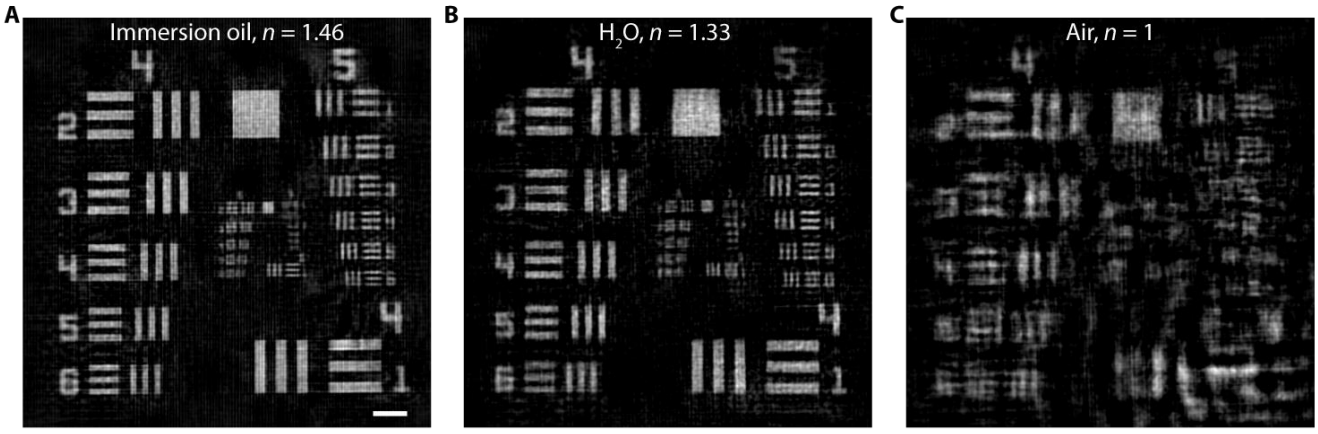

fig. S6. Refractive index matching. Comparison of reconstructed images captured with different media between the surface of the mask and the target, (**A**) immersion oil, (**B**) H_2O, and (**C**) air. Images were captured at a depth of 200 μm. Note that all images for the calibration process were captured with immersion oil between the surface of the mask and the target. Scale bar, 100 μm.

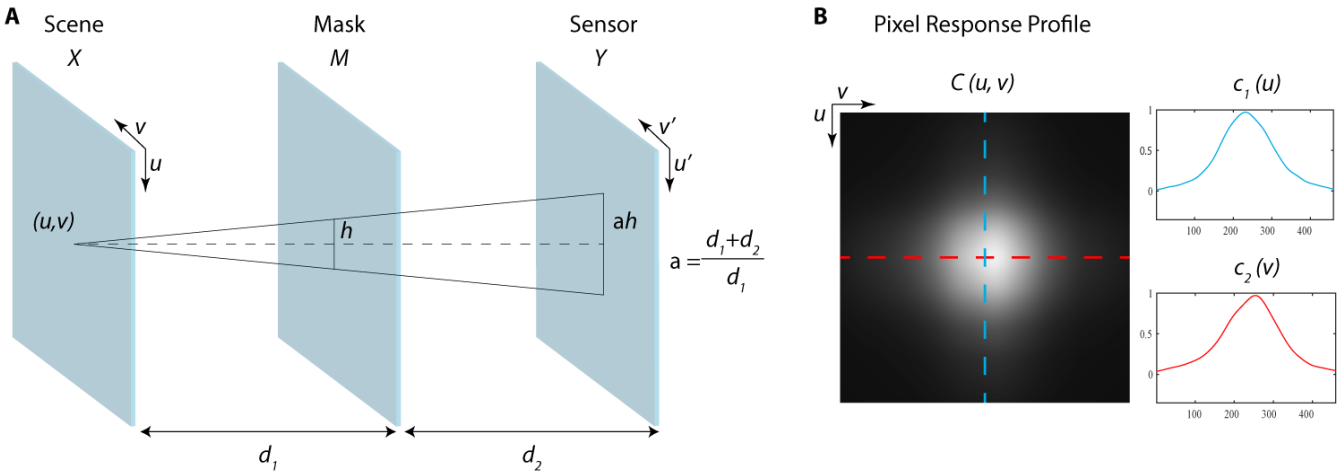

fig. S7. T2S derivation. (**A**) Illustration of scene, mask and sensor positions. (**B**) Sensor pixel's 2D response profile $C(u, v)$, which can be written as an outer product of the pixel's response along rows ($c_1(u)$) and the pixel's response along columns ($c_2(v)$).

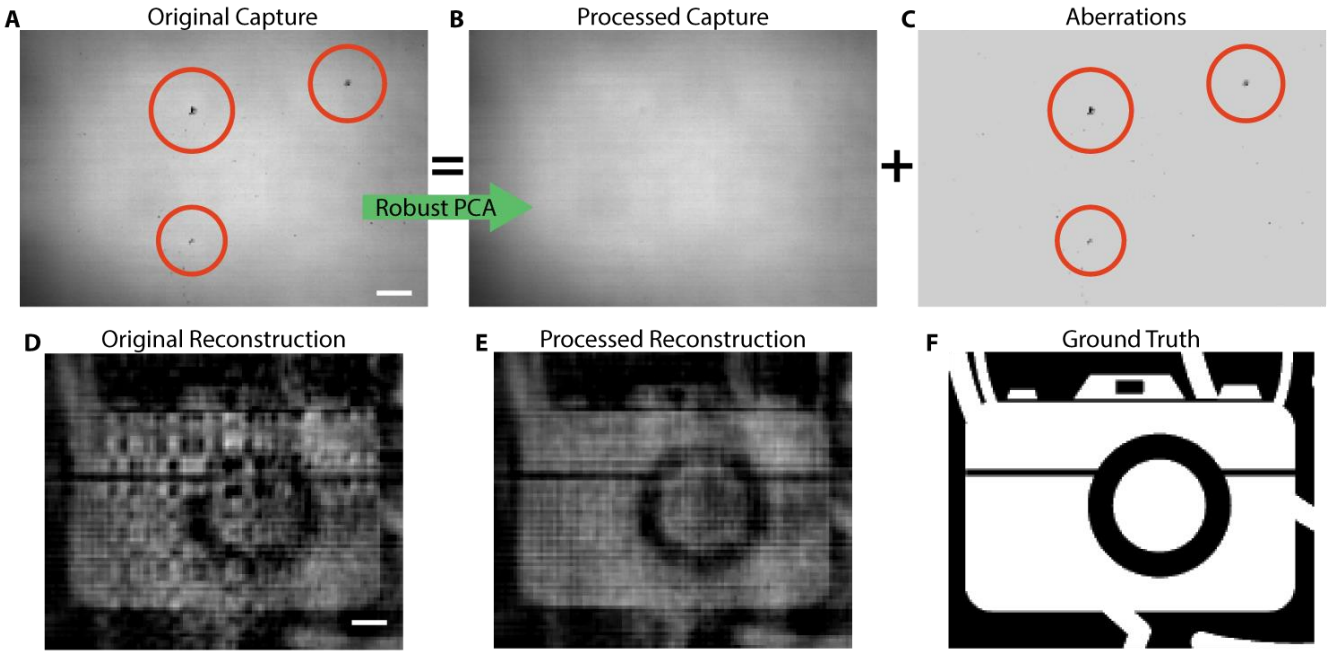

fig. S8. Aberration removal using RPCA. (**A**) Captured FlatScope image before aberration removal. (**B**) Processed capture without aberrations. (**C**) Aberrations removed by using Robust PCA. (**D**) Reconstructed image prior to Robust PCA shows large artifacts. (**E**) Processed reconstruction with aberration removal results in substantially reduced artifacts. (**F**) Ground truth image. Scale bars, 50 μm.

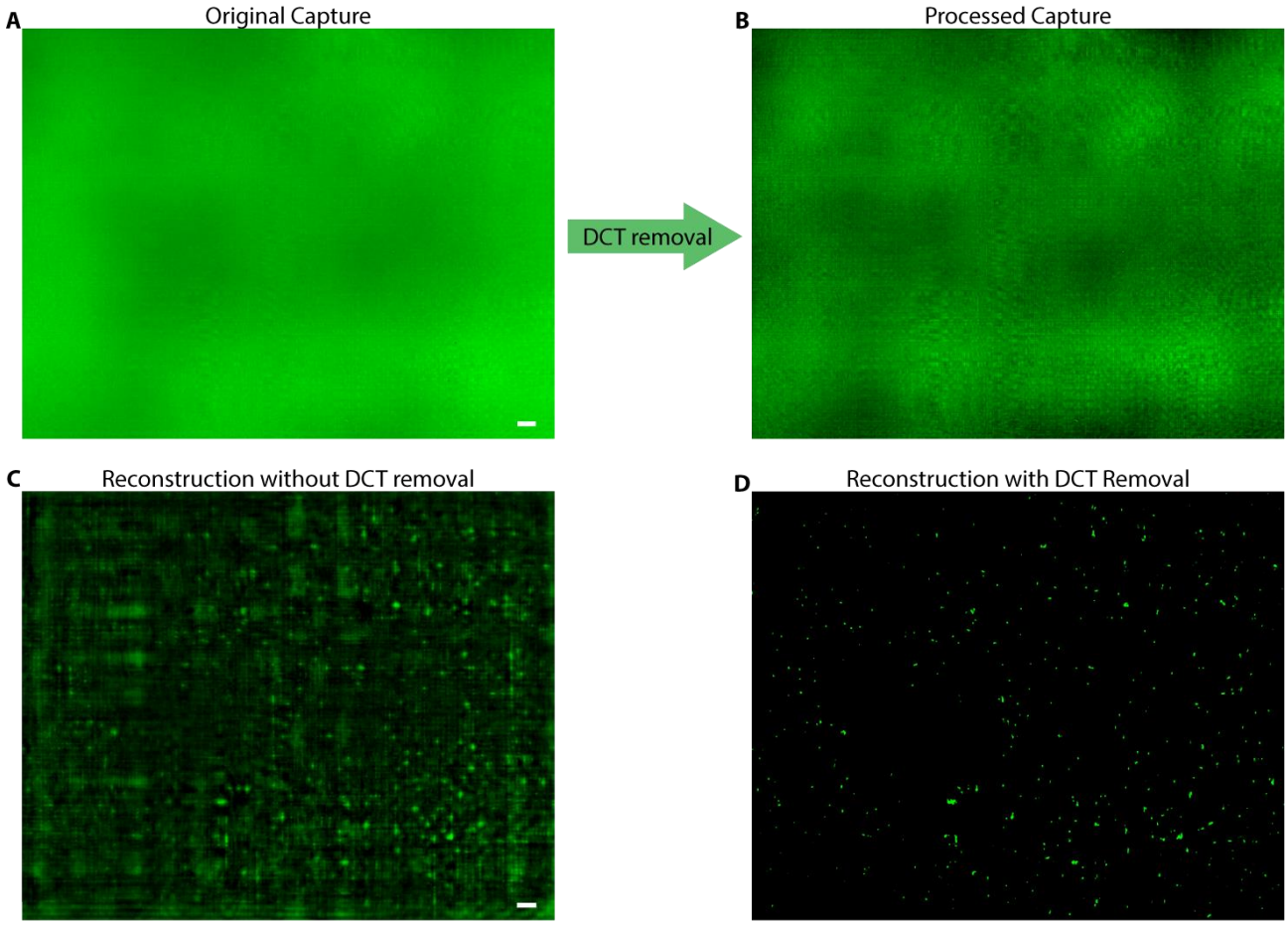

fig. S9. Removing effects from autofluorescence. (**A**) Captured FlatScope image before DCT removal. (**B**) Processed capture image. (**C**) Reconstructed image prior to DCT removal shows excessive noise. (**D**) Processed reconstruction with DCT removal results in substantially reduced noise. Scale bars, 100 μm.

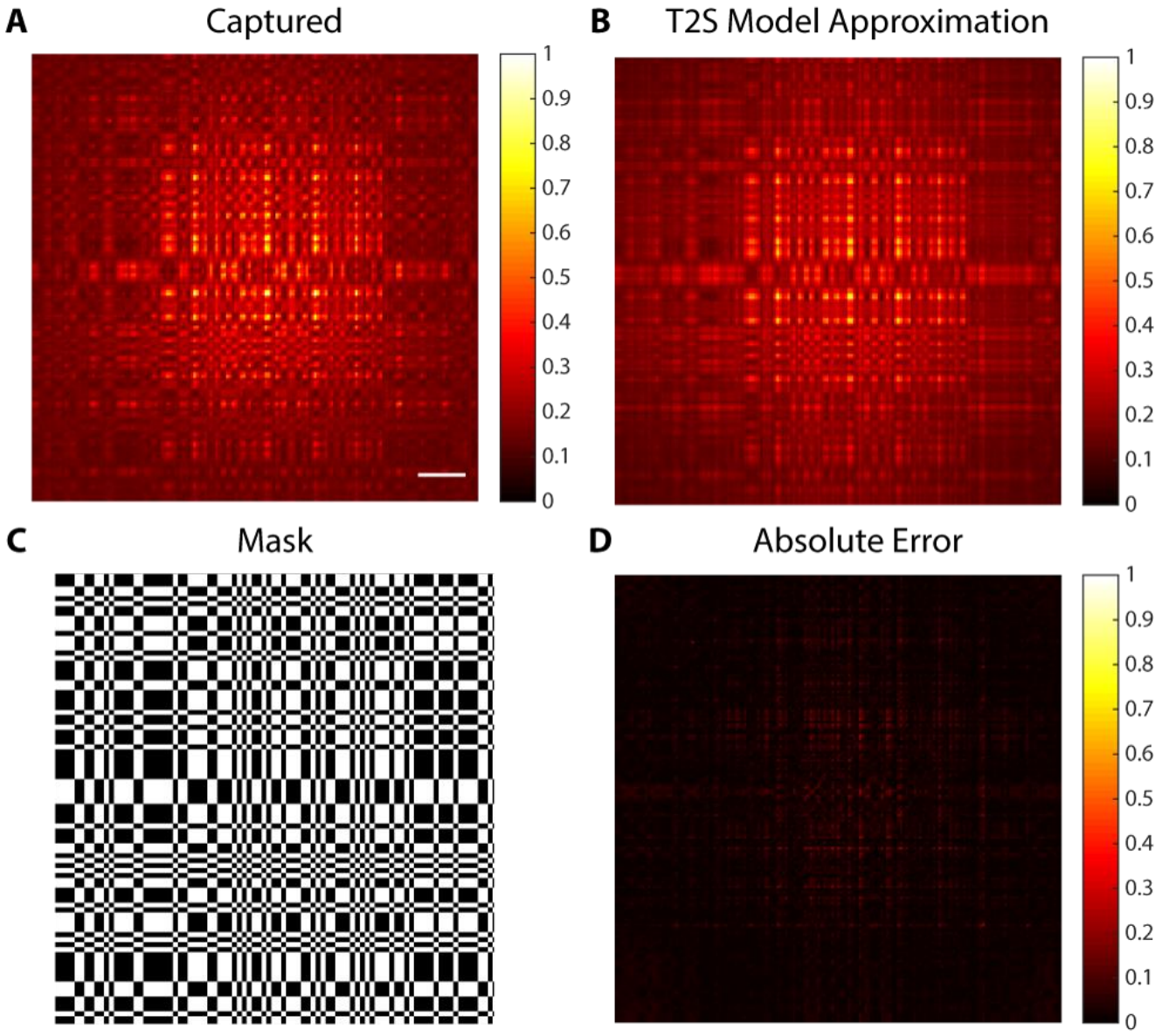

fig. S10. T2S model error from diffraction. (**A**) Raw capture by FlatScope prototype of a point source placed at a depth of 250 μm. The raw capture is cropped to the region of interest. (**B**) T2S model approximation of the capture. (**C**) Mask region illuminated by the point source. The location of the point source was chosen such that the illuminated mask region has a diversity in aperture sizes (white represents open apertures; black represents closed apertures). (**D**) Absolute error difference between raw capture and T2S model approximation. Average error is 12.64%. Scale bar, 40um.

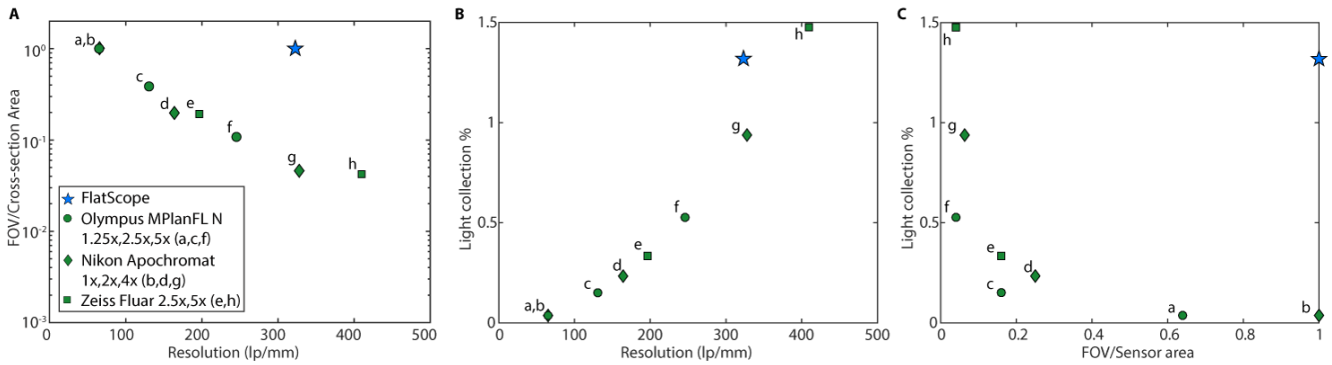

fig. S11. Light collection comparison of FlatScope and microscope objectives. (**A**) FOV/Cross-section area versus resolution for FlatScope and traditional lensed research microscopes with comparable objectives from Olympus (1.25×/2.5×/5×, NA = 0.04/0.08/0.15), Nikon (1×/2×/4×, NA = 0.04/0.1/0.2), and Zeiss (2.5×/5×, NA = 0.12/0.25). Note that cross-section area is a physical constraint imposed by the pupil diameter. (**B**) Light collection versus resolution comparison for FlatScope and the same microscopes/objectives. (**C**) Light collection versus FOV/Sensor area comparison for FlatScope and the same microscopes/objectives, where FOV/Sensor area = 1/magnification^2. Note that the light collection % for FlatScope does not account for noise introduced in reconstruction.

movie S1. Digital focusing of simulated resolution target. Left shows the ground truth image of the simulated resolution target located at 300 μm. Right shows the z-slices through the reconstructed focal stack at depths ranging from 240 μm to 360 μm.

movie S2. Fluorescent beads flowing in microfluidic channels of different depths. The sample is a 3D volume containing fluorescent beads flowing through two separate microfluidics channels (captured at 18 FPS, playback at half-speed). The flowing beads are at estimated depths of 265 μm and 355 μm and are false-colored cyan and red, respectively (dashed lines indicate approximate location of microfluidic channels). Top left shows a render of FlatScope and the microfluidic channels. Bottom left shows the capture by FlatScope. Top and bottom right show the FlatScope 3D reconstructed video from an XY and XZ perspective, respectively.