Science Advances

Supplementary Materials for

Optical fiber bundles: Ultra-slim light field imaging probes

A. Orth*, M. Ploschner, E. R. Wilson, I. S. Maksymov, B. C. Gibson

*Corresponding author. Email: antony.orth@rmit.edu.au

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/4/eaav1555/DC1)

Movie S1 (.mp4 format). Viewpoint shifting animation of multilayered bead sample
(Fig. 5, A and B) along the *x* axis.
Movie S2 (.mp4 format). Viewpoint shifting animation of multilayered bead sample
(Fig. 5, A and B) along the *y* axis.
Movie S3 (.mp4 format). Viewpoint shifting animation of multilayered bead sample
(Fig. 5, A and B) along a circular trajectory.
Movie S4 (.mp4 format). Viewpoint shifting animation of fluorescent paper sample

(Fig. 5, C and D) along the x axis.

Movie S5 (.mp4 format). Viewpoint shifting animation of fluorescent paper sample (Fig. 5, C and D) along the y axis.

Movie S6 (.mp4 format). Viewpoint shifting animation of fluorescent paper sample (Fig. 5, C and D) along a circular trajectory.

Movie S7 (.mp4 format). Viewpoint shifting animation of proflavine-stained mouse brain sample (Fig. 6, A to C) along the x axis.

Movie S8 (.mp4 format). Viewpoint shifting animation of skin autofluorescence (Fig. 6, D to F) along the x axis.

Supplementary Information

Section S1. Derivation of Eq. 4

Modifying the aperture to simulate defocus

Consider the intensity distribution at the output fiber facet (after pixellation removal) due to a point source located a distance z (ie. source depth) from the input fiber facet. This point spread function (PSF) depends on both the collection aperture of the fiber and the depth of the point source. A useful model for this PSF is a 2D Gaussian (*30*), with a width that scales with object depth and collection aperture

$$PSF(\vec{r}, z, \tan\theta) = \frac{1}{z^2}\exp[-4\ln(2)|\vec{r}|^2/z^2\tan^2\theta] \tag{S1}$$

Note that total energy contained within this PSF is independent of z, and increases with the square of the collection aperture. The factor of $4\ln(2)$ in the exponential sets the full width at half maximum (FWHM) of the Gaussian to $z\tan\theta$. Note that in this approximation, the majority of the energy of the PSF is within the maximum collection angle aperture, as would be expected of a fiber with a finite numerical aperture.

Now consider three imaging scenarios with PSFs, I_0,I_1, and I_2. Images I_0 and I_1 are acquired with identical fiber-to-sample distance z, but with different collection apertures: $\tan\theta_0$ and $\tan\theta_1$, respectively. Image I_2 is acquired with collection aperture $\tan\theta_0$, with the fiber physically moved along the z-axis by a distance Δz

$$I_0 = PSF(\vec{r}, z, \tan\theta_0) = \frac{1}{z^2}\exp[-4\ln(2)|\vec{r}|^2/z^2\tan^2\theta_0] \tag{S2}$$

$$I_1 = PSF(\vec{r}, z, \tan\theta_1) = \frac{1}{z^2}\exp[-4\ln(2)|\vec{r}|^2/z^2\tan^2\theta_1] \tag{S3}$$

$$I_2 = PSF(\vec{r}, z + \Delta z, \tan\theta_0) = \frac{1}{(z+\Delta z)^2}\exp[-4\ln(2)|\vec{r}|^2/(z + \Delta z)^2\tan^2\theta_0] \tag{S4}$$

Our goal is to derive an equation that relates I_1 and I_2. In order to do this, we need to choose a Δz value such that the width of the PSF is equal for I_1 and I_2. This is accomplished by setting $\Delta z = z\left(\frac{\tan\theta_1}{\tan\theta_0} - 1\right)$. With this identification, we can write

$$\alpha I_1 = \frac{1}{(z+\Delta z)^2}\exp[-4\ln(2)|\vec{r}|^2/(z + \Delta z)^2\tan^2\theta_0] = I_2 \tag{S5}$$

with $\alpha = \left(\frac{\tan\theta_0}{\tan\theta_1}\right)^2$. Note that α is also the ratio of the total energy collected in each case: $\alpha = \int I_0 d\vec{r}/\int I_1 d\vec{r}$. Using Eq. S5, the discrete axial derivative of intensity at the fiber facet can be expressed in two ways

$$\frac{\partial PSF}{\partial z} \approx \frac{I_0 - I_2}{\varDelta z} = \frac{I_0 - \alpha I_1}{z\left(\frac{\tan\theta_1}{\tan\theta_0} - 1\right)} \tag{S6}$$

Conveniently, the second expression involves two images, I_0 and I_1, which we can obtain from a single fiber bundle image without physically moving the fiber.

Effective light field moments

Suppose that our image I is now the result of n point sources each at a different depth z_j

$$I(\vec{r}, z, \tan\theta) = \sum_{j=1}^{n} B_j PSF(\vec{r}_j, z_j, \tan\theta) = \sum_{j=1}^{n} B_j PSF_j \tag{S7}$$

For the two aperture settings of interest θ_0 and θ_1

$$I_0 = I(\vec{r}, z, \tan\theta_0) = \sum_{j=1}^{n} B_j PSF(\vec{r}_j, z_j, \tan\theta_0) = \sum_{j=1}^{n} B_j PSF_{0j} \tag{S8}$$

$$I_1 = I(\vec{r}, z, \tan\theta_1) = \sum_{j=1}^{n} B_j PSF(\vec{r}_j, z_j, \tan\theta_1) = \sum_{j=1}^{n} B_j PSF_{1j} \tag{S9}$$

Where B_j is the brightness of point source j. We can relate the intensity resultant from each point source to the axial intensity derivative using Eq. S6

$$\sum_{j=1}^{n} B_j PSF_{0j} - \alpha B_j PSF_{1j} \approx \left(\frac{\tan\theta_1}{\tan\theta_0} - 1\right) \sum_{j=1}^{n} z_j B_j \frac{\partial PSF_j}{\partial z} \tag{S10}$$

From the continuity equation underpinning LMI (Eq. 1 main text), we can also relate the axial intensity derivatives $\frac{\partial PSF_j}{\partial z}$ to the light field moments $\vec{M}_j$[21]

$$\frac{\partial PSF_j}{\partial z} = -\nabla_\perp \cdot PSF_j \vec{M}_j \tag{S11}$$

Where $\nabla_\perp = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right]$. Combining the preceding two equations yields

$$\sum_{j=1}^{n} B_j PSF_{0j} - \alpha B_j PSF_{1j} \approx -\left(\frac{\tan\theta_1}{\tan\theta_0} - 1\right) \sum_{j=1}^{n} z_j \nabla_\perp \cdot B_j PSF_j \vec{M}_j = -\left(\frac{\tan\theta_1}{\tan\theta_0} - 1\right) \nabla_\perp \cdot I\vec{M}_e \tag{S12}$$

$$\rightarrow I_0 - \alpha I_1 = -\left(\frac{\tan\theta_1}{\tan\theta_0} - 1\right)\nabla_\perp \cdot I\vec{M}_e \tag{S13}$$

where the effective light field moment $\vec{M}_e$ is an intensity-weighted average of the light field moments $\vec{M}_j$, scaled by the depth of each point source

$$\vec{M}_e = \sum_{j=1}^{n} z_j B_j PSF_j \vec{M}_j / I \tag{S14}$$

The effective light field moments $\vec{M}_e$ can therefore be calculated with two images of the same scene with the same focus position, but different aperture settings (I_0 and I_1). We will call this approach "aperture-LMI" (aLMI). Note that if the point sources are all located at the same depth, the effective light field moment vector is simply the intensity-weighted average of the light field moment vector arising from each point source. For images containing point sources at different depths, their contributions to $\vec{M}_e$ are amplified by the z-position (depth) of each point source.

Note that I in Eqs. S13-14 must be evaluated at a particular aperture setting. In practice, we use $I = (I_0 + \alpha I_1)/2$. This choice makes the finite difference in Eq. S6 resemble a central finite difference, as opposed to a forward or backward finite difference.

Light field from light field moments

The consequence of using $\vec{M}_e$ to calculate a light field instead of $\vec{M}$ can be illustrated by considering the light field L_j of a point source PSF_j located at a depth of z_j, and centered at the origin. With traditional LMI, we construct a light field using a Gaussian approximation for the angular portion (parameterized by the uv coordinates) of the light field

$$L_j(x,y,u,v) = PSF_j \exp\left[-2\left(u - M_{jx}\right)^2/\sigma^2 - 2\left(v - M_{jy}\right)^2/\sigma^2\right] \tag{S15}$$

where $\vec{M}_j = \left[M_{jx}, M_{jy}\right]$. Instead, with aLMI, we construct something slightly different

$$L_j(x,y,u,v) = PSF_j \exp\left[-2\left(u - z_j M_{jx}\right)^2/\sigma^2 - 2\left(v - z_j M_{jy}\right)^2/\sigma^2\right] \tag{S16}$$

Section S2. Derivation of Eq. 5

Parallax from the light field

The light field moment vector due to a point source can be written down using simple trigonometry (See Fig. 1A in the main text)

$$\vec{M}_j = \left[M_{jx}, M_{jy}\right] = \left[\frac{x}{z_j}, \frac{y}{z_j}\right] \tag{S17}$$

Substituting our 2D Gaussian PSF (Eq. S2) into Eq. S15 gives

$$L_j(x, y, u, v) = \frac{1}{z_j^2} \exp\left[-4\ln(2)(x^2 + y^2)/z_j^2 \tan^2 \theta_0\right] \exp\left[-2\left(\left(u - \frac{x}{z_j}\right)^2 + \left(v - \frac{y}{z_j}\right)^2\right)/\sigma^2\right]$$

(S18)

Thus, L_j is the product of two Gaussians, offset in xy space. The product of these two Gaussians is itself a Gaussian, with a centroid (mean) $\vec{C}_j$ that depends on the chosen viewpoint $[u_0, v_0]$

$$\vec{C}_j = \frac{z_j \tan^2 \theta_0}{\tan^2 \theta_0 + 2\ln(2)\sigma^2}[u_0, v_0]$$

(S19)

$$\rightarrow \vec{C}_j \propto z_j[u_0, v_0]$$

(S20)

This relationship describes the manifestation of parallax: the lateral position of an object shifts linearly with both the object depth and viewpoint.

On the other hand, the light field constructed with aLMI is (substituting Eq. S2 into Eq. S16)

$$L_j(x, y, u, v) = \frac{1}{z_j^2} \exp\left[-4\ln(2)(x^2 + y^2)/z_j^2 \tan^2 \theta_0\right] \exp\left[-2((u - x)^2 + (v - y)^2)/\sigma^2\right]$$

(S21)

In this case, the centroid of the point source image is

$$\vec{C}_j = \frac{z_j^2 \tan^2 \theta_0}{z_j^2 \tan^2 \theta_0 + 2\ln(2)\sigma^2}[u_0, v_0]$$

(S22)

Note that the centroid shift goes to 0 as $z_j \rightarrow 0$ (as is the case for traditional LMI). This is because of the unphysical behavior of the purely geometric optics Gaussian PSF (PSF_j), whose width vanishes as $z_j \rightarrow 0$. In reality, the wave nature of light and discrete sampling of the fiber cores cause the PSF width to approach a finite value as the object nears the fiber facet. To account for these effects, we modify the Gaussian PSF to include the fitting parameter σ_0

$$PSF_j = \exp\left[-4\ln(2)|\vec{r}_j|^2/(z_j^2 \tan^2 \theta + \sigma_0^2)\right]\frac{\tan^2 \theta_0}{z_j^2 \tan^2 \theta_0 + \sigma_0^2}$$

(S23)

Which in turn yields a parallax-induced centroid-shift

$$\vec{C_j} = \frac{z_j^2 \tan^2 \theta_0 + \sigma_0^2}{z_j^2 \tan^2 \theta_0 + \sigma_0^2 + 2\ln(2)\sigma^2}[u_0, v_0] \tag{S24}$$

This equation describes how the apparent lateral shift of an object scales with viewpoint (u_0, v_0) and object depth z_j. In Eq. 5 of the main text we have rewritten the above equation in terms of a depth constant $h \overset{\text{def}}{=} \sigma/\tan \theta_0$

$$\vec{C_j} = \frac{z_j^2 \tan^2 \theta_0 + \sigma_0^2}{z_j^2 \tan^2 \theta_0 + \sigma_0^2 + 2\ln(2)h^2 \tan^2 \theta_0}[u_0, v_0] \tag{S25}$$

Evidently, we have at our disposal a tunable parameter in h^2, which we can tune to modify the shape of the centroid-shift curve $\vec{C_j}$. If $h^2 \ll z_j^2$, then $\vec{C_j} \approx [u_0, v_0]$, which is independent of depth and therefore precludes the appearance of parallax. If $h^2 \gg z_j^2$, then the centroid-shift behavior approaches a parabolic behaviour with object depth. In this instance, parallax will be observed. If h^2 is too large, however, the centroid-shift becomes increasingly small and cannot be visualized conveniently.

Section S3. Resilience to bending

Figures S3 and S4 demonstrate that a moderate fiber bundle bending displacement of ~25mm over a 120mm section of fiber has a negligible effect on our technique. This makes our light field imaging technique at least 2 orders of magnitude more robust than wavefront shaping techniques multimode fibers, where imaging quality degrades significantly after only a few hundred μm of fiber bending displacement (*16*). We use depths maps for analysis as they inherently rely on parallax from all viewpoints in uv-space. Therefore, any effect of fiber bending on our light field imaging technique will be apparent in the resulting depth maps.

In figs. S3A-C, we describe the fiber imaging geometry used to explore fiber-bending effects. A ~120mm section of a 1m long Fujikura FIGH-10-350S fiber is fixed to an optical table using tape. The remaining free arc (fig. S3B) is either left stationary or is manually bent by ~25mm repeatedly during a 10s acquisition (10 frames per second, 99ms integration time), as shown in fig. S3C. A flat sample of fluorescent beads is placed on the distal end at either $z = 11 \ \mu m$ or $z = 26 \ \mu m$ from the fiber facet. Depth maps are created using Adelson's least-squares method (*29*), with a window size of 9x9 pixels. Intensity and depth map images at $z = 11 \ \mu m$ are shown for reference in figs. S3D-E. Time traces of the mean depth map value within the field-of-view are shown in fig. S3F. At $z = 11 \ \mu m$, the standard deviation of the mean depth value time trace is slightly lower when actively bending the fiber compared to when left stationary (0.89 μm and 1.04 μm, respectively). The opposite is true at $z = 26 \ \mu m$, where we measure standard deviations of 2.56 μm and 2.22 μm, for actively bent and stationary fibers, respectively. Actively bending the fiber appears to have no systematic statistical effect on the temporal variation of mean depth map value.

While depth mapping appears to be unaffected by fiber bending, isolated core-to-core coupling can be observed in the raw camera images. In fig. S4, we follow the effects of an instance of large core-to-core coupling on depth mapping, when the sample is placed at $z = 11 \ \mu m$ from the fiber facet. At 3.5s (fig. S4A), the fiber is in its bent state (fig. S3C), whereas at t=4.2s (fig. S4B), the fiber is stationary (fig. S3B). Visual inspection shows that the depth maps at these two timepoints are nearly indistinguishable, with only very minor differences apparent upon magnification of a region of interest (ROI) in field-of-view (figs. S4A-B top insets). The raw core images indicate that core-to-core coupling is occurring in this ROI.

In particular the intensity of cores 1 (red) and 2 (blue) are strongly anti-correlated (Pearson correlation coefficient -0.80) as can be seen in their time traces (fig. S4C). These are the only two cores within this ROI that demonstrate significant intensity variation (>10%) during movie acquisition. Despite this behaviour, the standard deviation of the depth map at the location of core 1 is 0.73 μm, which is less than the standard deviation of the mean depth map value over the entire field-of-view, regardless of whether the fiber is actively bent (0.89 μm), or stationary (1.04 μm). We conclude that even instances of strong local core-to-core coupling are not resolvable above the noise floor in the depth map for typical acquisition parameters.

Section S4. Sampling considerations

This section is intended to provide a discussion of trade-offs related to sampling at the fiber bundle facets and the image sensor plane.

Distal fiber facet: Potential users of this technique will find that they are effectively limited to the fiber bundles used in this work (Fujikura) due to their market dominance in this core size regime. As such, there is freedom to choose between "*S*" and "*N*" series fibers, which have average fiber radii of ~1um and ~1.45um (ref. 16). We use the *S* series fibers due to their superior spatial sampling density. The *N* series fibers will support ~2x more modes, which in theory will increase the information content in the angular domain, though taking advantage of this extra information may require more sophisticated modal unmixing techniques. However, this extra information in the angular domain comes at the expense reduced information in the spatial domain. We note that this trade-off is ever-present in light field imaging, regardless of the imaging architecture.

Proximal fiber facet: Correct sampling at the proximal facet is crucial for high quality results. As with any fiber related technique, it is advisable to match the NA of the microscope objective to the NA of the fiber, in this case~0.4. 0.4NA microscope objectives are commonplace in either 10x or 20x magnifications. Using an objective with a lower NA is still possible, however, the collection efficiency of higher order modes by the microscope objective will suffer, leading to a decrease in SNR for large angles in the light field. Another view of the same effect is that a low-NA objective cannot spatially resolve the high-order modes due to the diffraction limit. The fundamental cause of this effect is the same: the photons carrying high spatial frequency information from high-order modes are not collected by a low NA microscope objective.

Maintaining a large field of view is also ideal, so using a low magnification (10x), 0.4NA microscope objective is preferable.

Image sensor: The image sensor must sample the image projected by the microscope objective (through the tube lens) with a high enough sampling density to resolve all of the modes present. Thus, one must satisfy the Nyquist criterion at the diffraction limit of the objective, resulting in a maximum pixel size of $\approx$ Mλ/4NA=3.125μm for 500nm light, NA=0.4 and magnification M=10x. This is the maximum pixel size permissible in order to avoid compromising the angular information of the light field. This matches our camera's pixel size of 3.1μm. Smaller pixel sizes or larger magnifications are possible , at the expense of SNR due to increased camera read noise for the same fiber area. Note that this sampling requirement necessitates more pixels per core than a standard microendoscope system - roughly 2x more in each dimension. As a result, read noise is increased by approximately a factor of 2, and therefore 4x more photons are needed to maintain the same SNR as in a system where the cores are just resolvable. This increased photon budget is a direct consequence of the enhanced information content delivered by the resolved core modes.

Section S5. Light field imaging considerations: Axial direction

There is no fundamental limit to the axial range of our fiber imaging technique aside from SNR and dynamic range considerations. If the entire PSF is contained within the fiber's FOV, our approach will be able to return a light field. The more stringent restriction is that of the SNR. If the sample is 3D with some parts close to the fiber facet and some parts far away, then parts close to the sample will appear to be much brighter. At some depth, the photons emitted by a distant object are so spread out over the fiber and camera that the signal will fall below the noise floor of the camera. One can remedy this by increasing the excitation intensity or integration time but close objects will eventually saturate the camera. Saturated pixels violate our imaging model, leading to erroneous results. Therefore, in practical circumstances, the axial FOV is restricted by the dynamic range of the camera and/or by the height and brightness variation of the sample itself. There is no hard and fast rule here, but our experience indicates that good results can be obtained between within $75\mu m$ of the fiber facet with 8-bit machine vision cameras. More expensive sCMOS cameras will certainly do better as they have larger electrons wells and lower read noise.

The ability to truly resolve overlapping objects in 3D is achieved using lightfield focal stack deconvolution. The resolution of this approach is dependent on the SNR and the inherent contrast and geometry of the sample, therefore no simple relationship can be given to quantify the axial resolution for all samples and cameras (see ref. 28 for further discussion). However, the optimal sampling approaches outlined above are essential for achieving maximum axial resolution. To reiterate, major factors that will decrease axial resolution are: a) microscope objective NA less than 0.4; b) oversampling at the image sensor. Lower wavelengths are also ideal from the perspective of information throughput as the more modes can propagate in cores at 500nm than 600nm (for example). However, this effect is minimal and in practice emission wavelength is dictated by clinical restrictions (FDA approved contrast agents) and the typical autofluorescence window in the green.
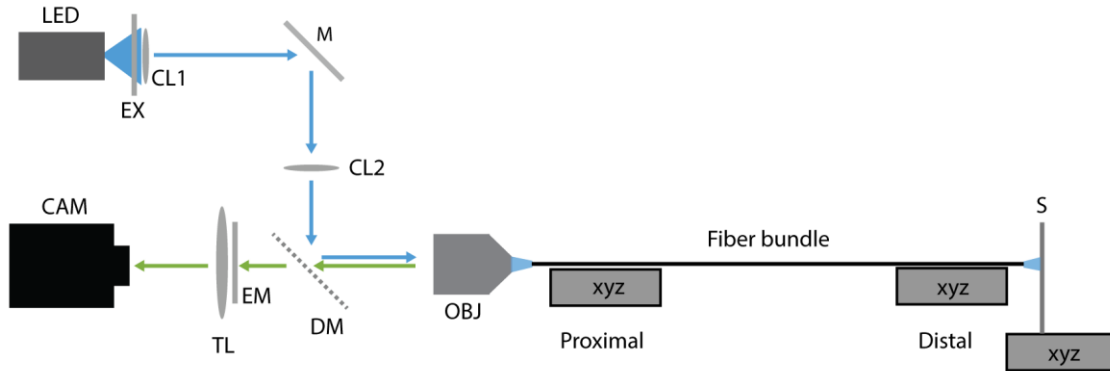
Supplementary Figures and Videos

Fig. S1. Optical setup for fluorescence imaging through optical fiber bundles. Components are labeled as follows: LED=Light emitting diode (Thorlabs, center wavelength 455nm); EX=Band pass excitation filter (Semrock, center wavelength 465nm, 40nm FWHM); CL1=Condenser lens 1 (Thorlabs, f=30mm); M=Mirror; CL2=Condenser lens 2 (Thorlabs, f=100mm); DM=Long pass dichroic mirror (Semrock, Cut-on wavelength 509nm); OBJ=Microscope objective (Olympus, 10x, 0.4NA); xyz=3-axis micrometer stages (Thorlabs) ; S=Sample; EM=Fluorescence emission filters (532nm long pass (Semrock) + 600nm short pass (Edmund Optics); TL=Tube lens (Thorlabs, f=200mm); CAM = Camera (Point Grey Grasshopper 3). Blue and green arrows represent excitation and fluorescence emission light, respectively.
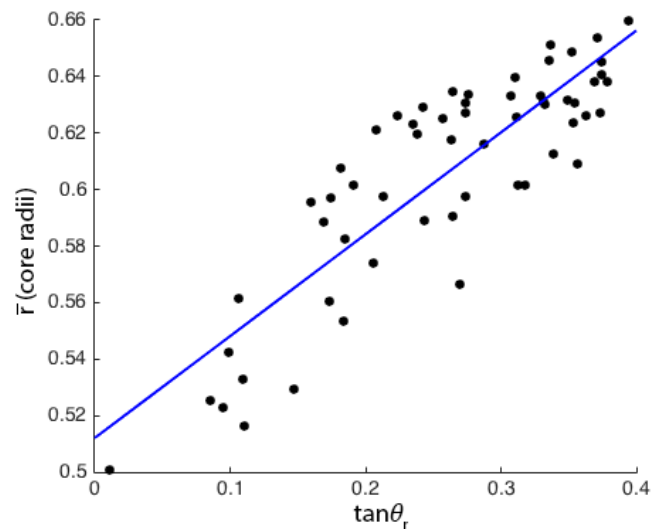

Fig. S2. Correlation between ray angle and intracore light distribution. Scatterplot of first radial moment $\bar{r}$ (defined as $\bar{r} = \sum |r| I_c / \sum I_c$, integral is taken over the image of the core I_c), against the tangent of the *input* ray angle ($\tan \theta_r = \sqrt{\tan^2 \theta_x + \tan^2 \theta_y}$). Each datapoint corresponds to the $\bar{r}$ value of an individual core from the image in Fig. 1d. The y-axis is displayed in units of core radii to account for the variable sizes of the cores. The blue line is a linear fit to the data, with $\bar{r} = 0.36 \times \tan \theta_r + 0.51$. The first radial moment $\bar{r}$ and the tangent of the ray angle $\tan \theta_r$ are strongly correlated, with a Pearson correlation coefficient of 0.86 (p-value 3.2×10^{-18}). The linear fit is not intended to convey a specific physical relationship between $\bar{r}$ and $\tan \theta_r$, but rather to demonstrate that the two quantities are strongly correlated.
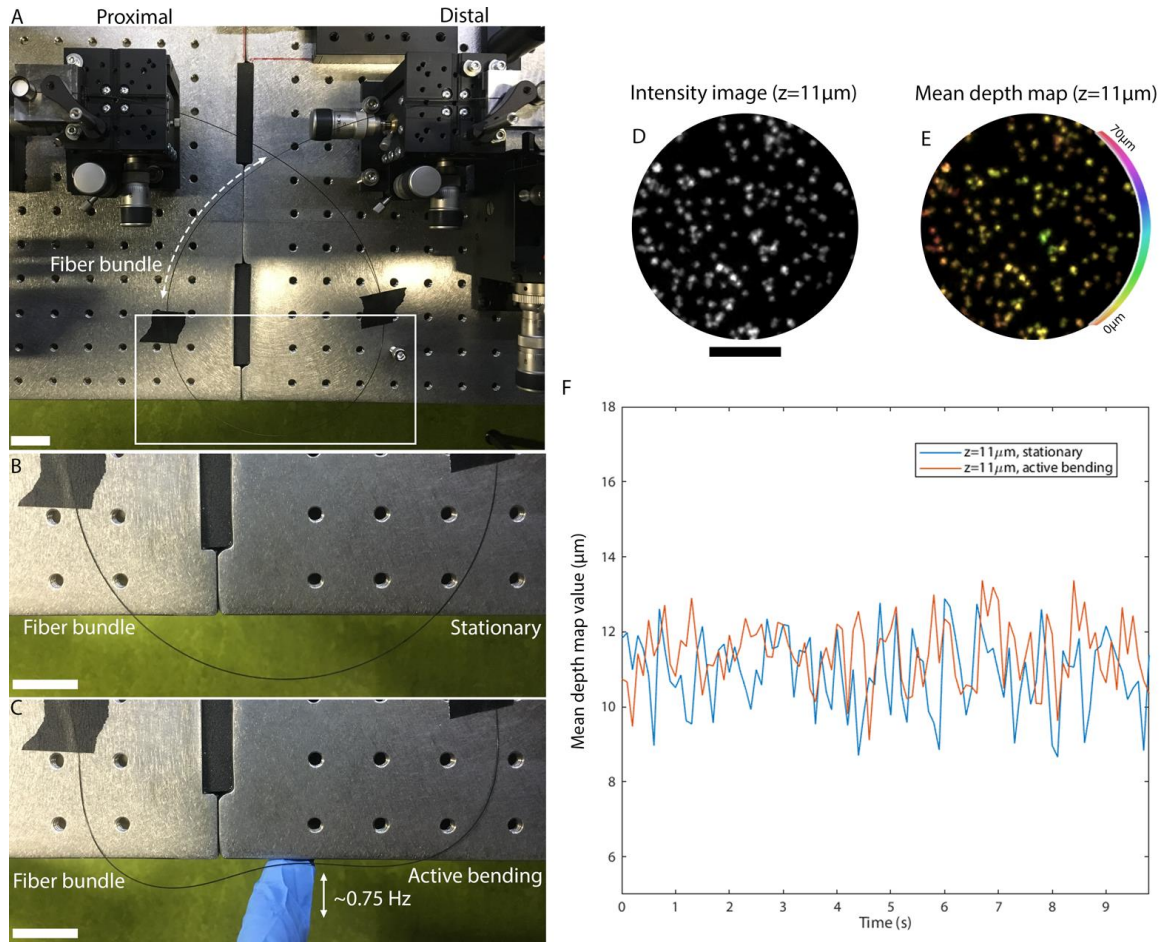
Fig. S3. Depth mapping is robust to dynamic fiber bending. (**A**) Overview of a 1m long fiber bundle (Fujikura FIGH-10-350S) in the optical system described in fig. S1. The white dashed arc indicates the location of the fiber bundle on the optical table for clarity. Scale bar 25mm. (**B**) Zoom-in of a portion of the looped fiber bundle when stationary. The fiber bundle is affixed to the optical table at either end of an arc by black tape. Scale bar 25mm. (**C**) As in (B), but with the fiber actively bent by a finger. The approximate displacement of the middle of the arc when bent is 25mm. This bending is repeated manually at a rate of approximately 0.75Hz over 10s for depth map bending analysis. (**D**) Intensity image of a field of view of fluorescent beads observed through the fiber bundle, after light field processing (central viewpoint). In this image, this flat sample is placed at $11\mu m$ from the fiber facet. Scale bar **$100\mu m$**. (**E**) Mean depth map of the field of view in (D). (**F**) Mean depth map value for sample-to-fiber facet distance of $z = 11\ \mu m$ for stationary and actively bent fiber bundles. (as pictured in (B) and (C)). The stationary dataset is taken of the same field of view immediately after the active bending dataset for both sample-to-fiber distances. When the fiber is actively bent (stationary), the time-averaged mean depth map value is 11.43 μm (10.92 μm), with standard deviation 0.89 μm (1.04 μm). There is no discernible difference in the magnitude of time trace noise for actively bent vs. stationary fiber bundles. The inherent depth mapping noise is larger than variation introduced by this bending.
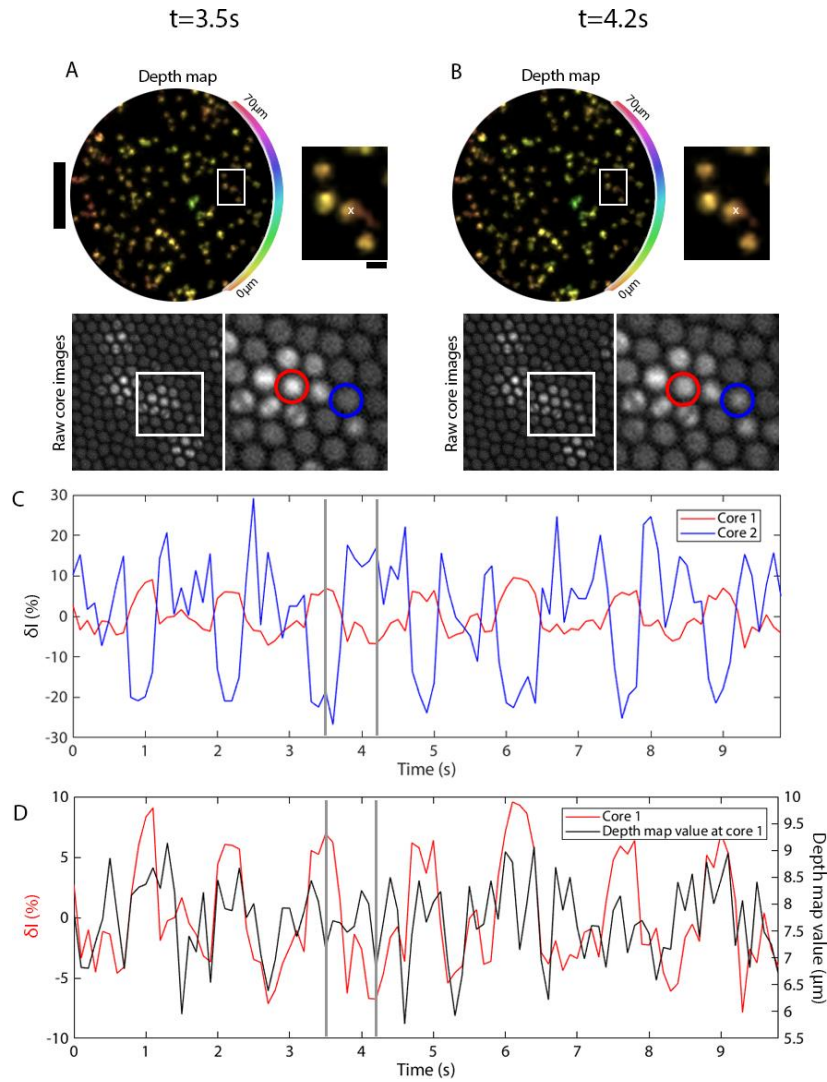
Fig. S4. Effect of dynamic core-to-core coupling during bending. (**A**) Top: Depth map of a field of view of fluorescent beads at t=3.5s during active bending (fig. S3). Fiber-to-sample distance $z = 11\mu m$. Scale bar $100\,\mu m$. Inset: magnified view of boxed region. Scale bar $10\,\mu m$. Bottom: Raw core images corresponding to the depth map inset region, and magnified view of the cores. Core 1 is circled in red, core 2 is circled in blue. (**B**) Top: Depth map of a field of view of fluorescent beads at t=4.2s during active bending (fig. S3). Inset: magnified view of white boxed region. Note the small difference between the insets of (A) and (B). Bottom: Raw core images corresponding to the depth map inset region, and magnified view of the cores. Core colours as in (A). Note the small difference between red and blue circled cores at t=3.5s (a) vs. t=4.2s (B). (**C**) Time trace of the intensity at the center of cores 1 and 2 (red and blue, respectively in (A) and (B)). Y-axis is plotted in terms of the % change in intensity from the mean intensity for a given core. Vertical lines indicate t=3.5, 4.2s. The intensity time trace of core 1 and cores 2 are strongly anti-correlated (Pearson correlation coefficient -0.80, p $= 8 \times 10^{-23}$). (**D**) Time trace of core 1 overlaid with the depth map value at the location of core 1 (white x) in the depth maps in (A) and (B). The intensity time trace of core 1 (red) is partially correlated to this local depth map value (Pearson correlation coefficient 0.41, p $= 3 \times 10^{-5}$). The standard deviation of the depth map value at this location is 0.73 μm, with mean 7.70 μm. This standard deviation is slightly less than the standard deviation of the mean depth map time trace for the entire field-of-view (0.89 μm, see fig. S3), indicating that even large core coupling effects are negligible compared to the inherent depth mapping noise of our technique.
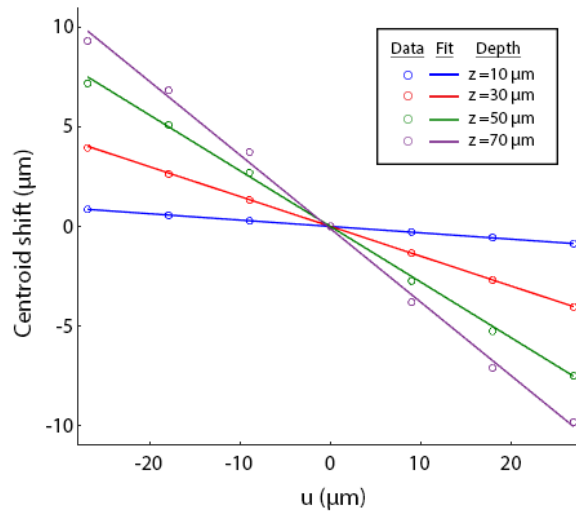
Fig. S5. Centroid shift of a fluorescent bead viewed from different u positions (viewpoints). Centroid shift vs. u data and linear fits are shown for depths of $z = 10$ (blue), 30 (red), 50 (green) and $70\mu m$ (purple). Datapoints correspond to experimentally measured centroid shifts, and solid lines are linear fits to the data. The R-squared values of the fits are 0.9998, 1.000, 0.9987, and 0.9968 for bead depths of $z = 10$, 30, 50, $70\mu m$, respectively, indicating that data match the expected theoretical linear trend in Eq. 5 of the text.
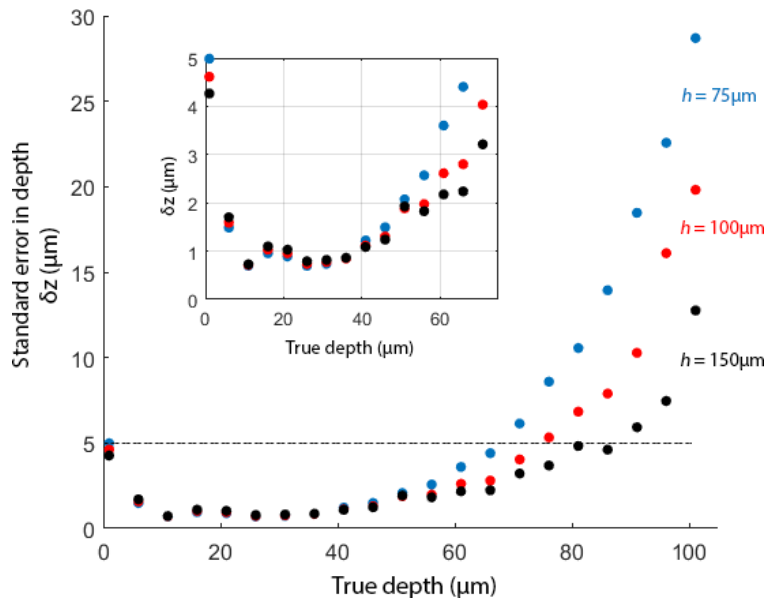

Fig. S6. SE in depth measurements from Fig. 3B. The raw standard deviations (given by the error bars in Fig. 3A) are converted to an uncertainty in depth (δz) by finding the depths corresponding to the centroid shifts at the true depth $\pm$ one standard deviation (of the centroid shift measurements). The average differences between these depths and the true depth is reported as the standard error in depth δz. Blue, red and black datapoints are for light field reconstructions with $h = 75, 100,$ and $150\mu m$, respectively. Inset: Zoomed-in plot of the depth uncertainty up to $71\mu m$. Uncertainty in measured depth is less than $5\mu m$ for depths less than $71\mu m$, and better than $2.1\mu m$ for depths between $6\mu m$ and $51\mu m$. The larger uncertainty at a depth of $1\mu m$ occurs because the PSF is almost entirely confined to a single pixel (core), yielding poor localization precision.
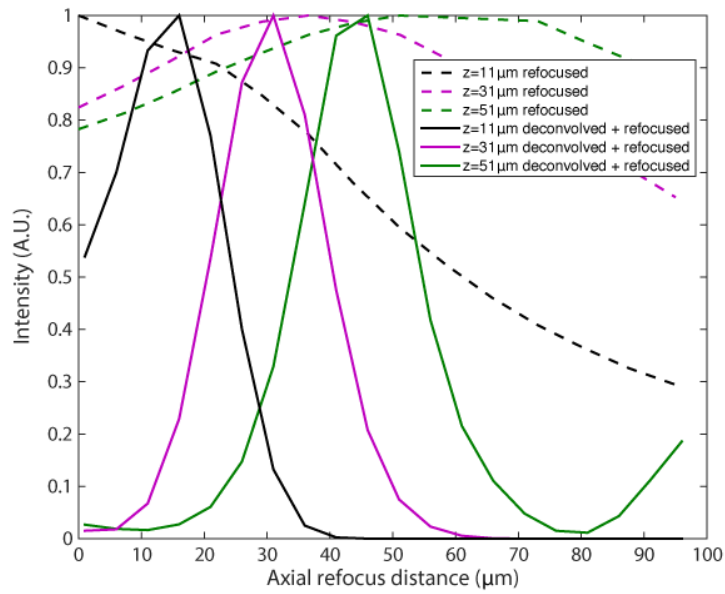
Fig. S7. Effect of light field focal stack deconvolution on axial localization. Mean axial intensity profile for refocused light fields (dotted curves) and deconvolved + refocused (solid curves) light fields for ground truth sample positions of $z = 11, 31,$ and $51\,\mu m$. Mean axial intensity profiles are computed by summing over the horizontal (x) direction in Figs. 4e,g,i for refocused light fields and in Figs. 4k,m,o for deconvolved + refocused light fields. Light field focal stack deconvolution + refocusing significantly sharpens the axial intensity profile compared to refocusing only. The deconvolved light field focal stack axial width ($\sim 20-30\,\mu m$) is comparable to the axial slice thickness of fixed-focus commercial fluorescence microendoscope systems (ref. 7).
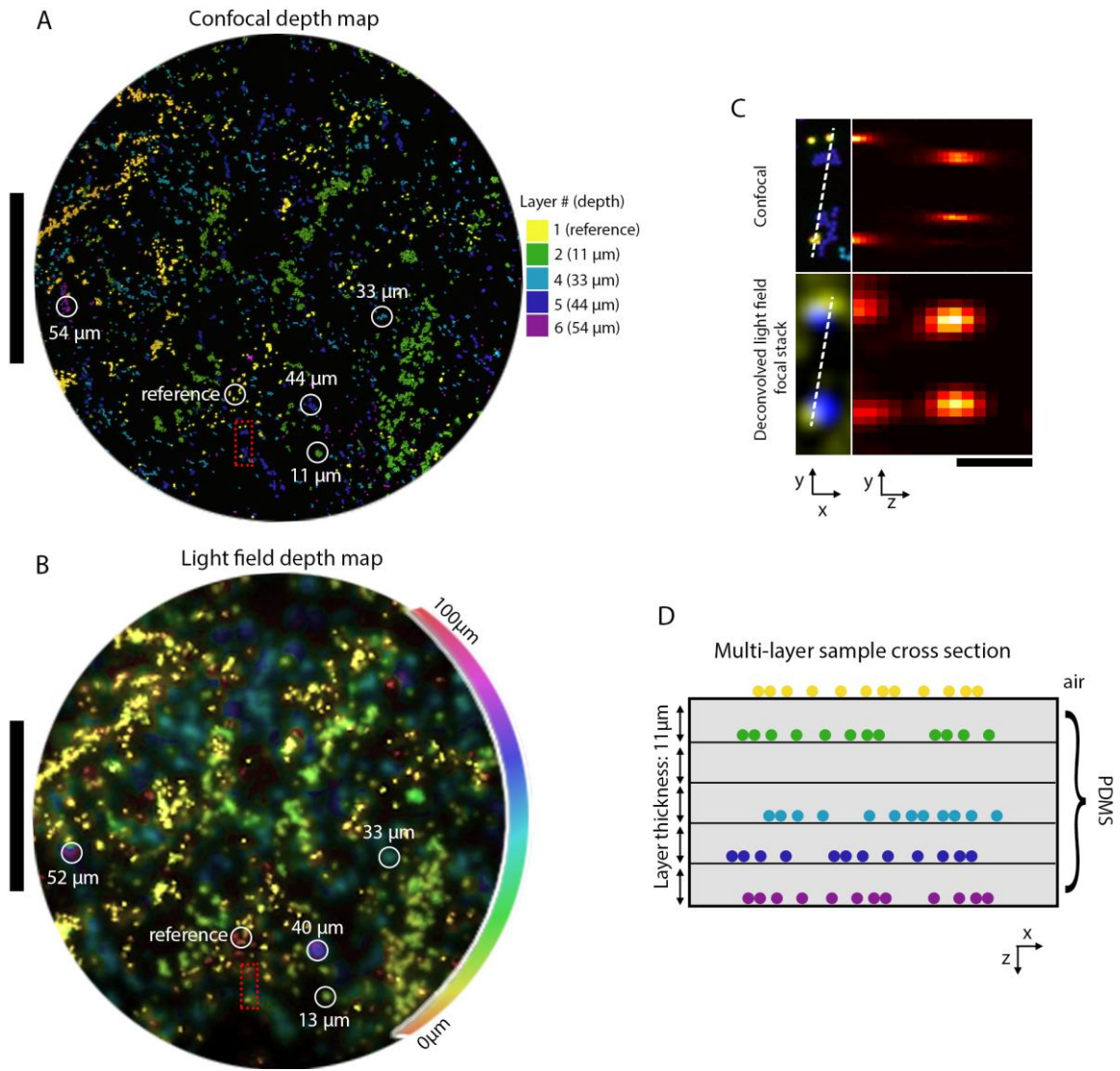
Fig. S8. Confocal versus light field depth maps. Comparison between ground truth confocal (**A**) and light field (**B**) depth maps for 3D fluorescent bead sample in Figs. 5b. The sample consists of 6 evenly spaced layers of beads in PDMS, however there are no beads from layer #3 in this field of view. Representative subregions containing beads from a single layer are circled and labelled with the layer depth. For (B) the reported depth is the average depth within a 3x3 pixel subregion centred on the circled region. The colour legend in (A) is the depth colour code for the confocal image. The depths in (A) are considered to be the ground truth; the depths in (B) are measured depths using the light field. For both datasets, we take the first layer as the reference layer and calculate subsequent depth with respect to this first layer. Both scale bars are $200\mu m$. (**C**) Comparison between confocal image stack and deconvolved light field focal stack, along xy (left) and yz (right) axes, for the area highlighted by the red boxes in (A) and (B). XY axis images are color-coded for depth: layer 1 is colored yellow and layer 5 is colored blue. The yz axis images are cross-sections along the dotted white line in xy axis images. Scale bar 25um. (**D**) Cross-sectional schematic of the multi-layered PDMS sample with fluorescent beads. Each layer has a thickness of $\approx 11\mu m$. All layers are incased in PDMS, except for the top layer (yellow beads) which are open to air.
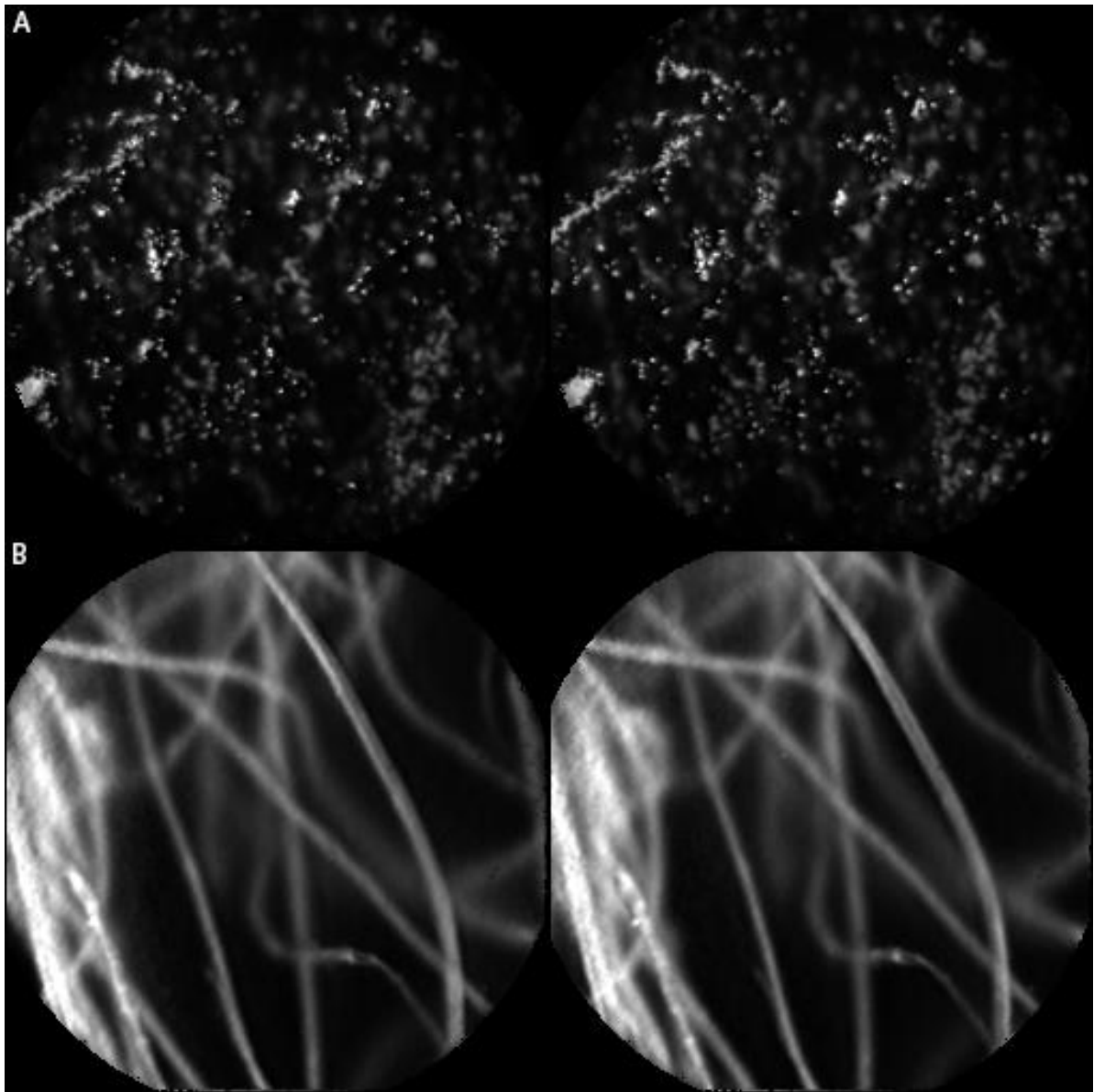
Fig. S9. Stereo pair for the 3D samples from Fig. 5. (**A**) 3D bead sample. (**B**) Fluorescent paper sample. These stereo pairs can be viewed on a mobile phone using virtual reality goggles such as Google Cardboard or similar.
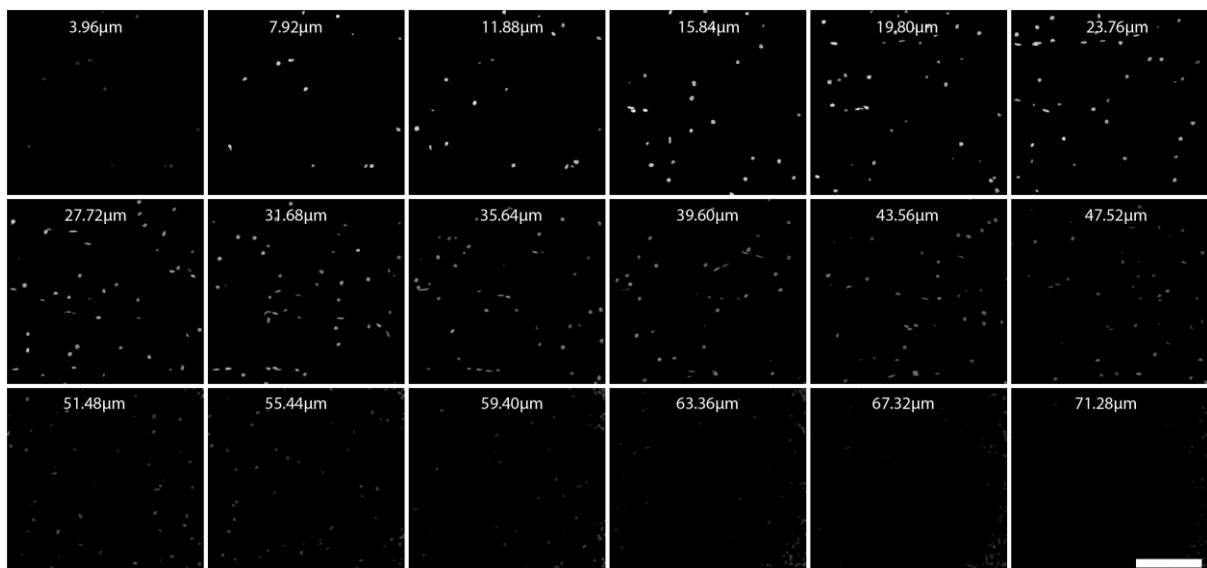

Fig. S10. Confocal stack of proflavine-stained mouse brain. Selected slices of the confocal stack used to calculate the dashed blue curve in Fig. 6C of the main text. Here, images are gamma-corrected for visibility. Scalebar $100\mu m$.

Supplementary Movies

For movies S1-S8, the point of fixation is moved slightly into the sample so that objects closest to the fiber facet appear to move in the opposite direction from those farthest from the fiber facet. This is achieved by applying a translation to each frame in the animation, proportional to the magnitude and direction of the viewpoint shift of that frame. Moving the point of fixation in this way enhances natural depth perception but does not affect the relative disparity between objects in the scene.

Movie S1. Viewpoint shifting animation of multilayered bead sample (Fig. 5, A and B) along the

x axis.

Movie S2. Viewpoint shifting animation of multilayered bead sample (Fig. 5, A and B) along the

y axis.

Movie S3. Viewpoint shifting animation of multilayered bead sample (Fig. 5, A and B) along a circular trajectory.

Movie S4. Viewpoint shifting animation of fluorescent paper sample (Fig. 5, C and D) along the

x axis.

Movie S5. Viewpoint shifting animation of fluorescent paper sample (Fig. 5, C and D) along the

y axis.

Movie S6. Viewpoint shifting animation of fluorescent paper sample (Fig. 5, C and D) along a circular trajectory.

Movie S7. Viewpoint shifting animation of proflavine-stained mouse brain sample (Fig. 6, A to C) along the x axis.

Movie S8. Viewpoint shifting animation of skin autofluorescence (Fig. 6, D to F) along the x axis.