Supplemental Document

Space-time Fourier ptychography for *in vivo* quantitative phase imaging: supplement

Ming Sun, Kunyi Wang, Yogeshwar Nath Mishra, Simeng Qiu, and Wolfgang Heidrich*

Visual Computing Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
wolfgang.heidrich@kaust.edu.sa

Supplement DOI: https://doi.org/10.6084/m9.figshare.26530786

Parent Article DOI: https://doi.org/10.1364/OPTICA.531646

Space-time Fourier Ptychography for *in vivo* Quantitative Phase Imaging: supplemental document

This document serves as a supplementary to "Space-time Fourier Ptychography for *in vivo* Quantitative Phase Imaging". We first present the imaging forward model. Then we detail the ST-FP optimization framework along with the solver algorithm. Quantitative validation is provided through simulation analysis. We show the adaptability of ST-FP across various scenarios such as moving targets and live microorganisms. Furthermore, an ablation study, results under different movement speeds, the static imaging study, and a comparison with another motion-corrected method are given, along with a comparison with the conclusion of detailed parameter settings for each case.

1. FORWARD MODEL

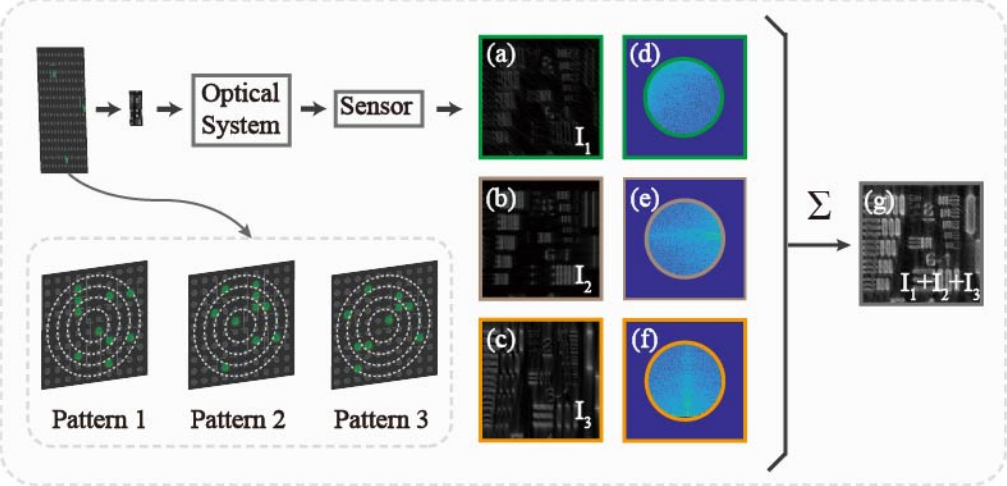

Fig. S1. Schematic of Fourier Ptychography (FP) with multiplexed illumination. (a)-(c) Sub-images corresponding to different spatial frequency regions obtained under varied LED illuminations. Each single low-resolution (LR) image captured under a specific LED, is indicative of the spatial frequency information of the object. (d)-(f) Resultant frequency domain representations, showing the spatial frequency shifts induced by different illumination angles. (g) Multiplexed illumination yields the sum of the intensities of the sub-images. LEDs are grouped in rings according to their illumination angle. For each illumination pattern, we randomly select LEDs from each ring, ensuring an equal number of LEDs are used.

The long acquisition time limits Fourier Ptychography (FP) imaging ability to real-time applications. Multiplexed illumination can be employed here to mitigate this problem by turning on multiple LEDs together [1–3], as shown in Fig. S1. It yields the sum of intensities of the sub-images from the LED set since their local plane waves are mutually incoherent with each other.

To avoid huge variations of intensity between different frames, we design a $T \times N_{LED}$ pattern matrix $\mathbf{P}$ after dividing LEDs into different regions. The pattern matrix can be expressed as:

$$\mathbf{P} = [\mathbf{P}_1^\top, \mathbf{P}_2^\top, \ldots, \mathbf{P}_T^\top]^\top. \tag{S1}$$

$\mathbf{P}$ is formulated as a binary matrix where each vector $\mathbf{P}_t$ represents a unique LED pattern. Each element within that vector is either 0 or 1, indicating whether the corresponding LED is turned off or on, respectively. The LEDs are arranged into distinct groups based on their illumination angles

to maintain the consistency of light brightness across successive frames. For each group, we select the same number of LEDs in each pattern, assuring the relative constancy of the intensity of the light source. We empirically determined 40 LEDs per pattern. The maximum illumination NA used is approximately 0.28, which closely matches the NA of the objective lens, ensuring sufficient information available for effective phase retrieval [4]. For this study, we utilized six different illumination patterns for each dataset during the acquisition process, which could have comprehensive data collection, facilitating detailed and accurate FP reconstructions. Note that we do not pursue scanning of all the LEDs since previous work [5, 6] has shown that capturing only part of the sub-images in FP will not deteriorate the performance compared with the case that captures all the data. Another reason is that if a large number of the LEDs turn on at once, it will be difficult to decompose since there will be little diversity for phase contrast [1].

2. SPACE-TIME FOURIER PTYCHOGRAPHY (ST-FP) FRAMEWORK

A. Optimization Model Construction

The optimization framework of ST-FP incorporates various regularization terms to address the ill-posed nature of the dynamic state reconstruction problem. Here, we elucidate the rationale behind the selection of these regularization strategies.

The least-squares error between the observed measurements and the predicted object leads to the first term of Eq. (6) in the main text. The second term comes from the optical flow constraint [7]. In the next line, we move from a coarsely discredited time axis in CS-FP to an essentially continuous time axis, where each captured projection image $\mathbf{b}_t$ owns its time stamp t, and warping is used to align the frames with the individual projections. It corresponds to two separate data terms that can be utilized in a reconstruction:

$$\|\mathcal{A}_{t+1}(\text{warp}(\mathbf{o}_t, \mathbf{v}_t)) - \mathbf{b}_{t+1}\|_2^2 + \|\mathcal{A}_{t-1}(\text{warp}(\mathbf{o}_t, -\mathbf{v}_{t-1})) - \mathbf{b}_{t-1}\|_2^2 . \tag{S2}$$

This term implements the warp and projection method: given the estimated frame $\mathbf{o}_t$ and the motion field $\mathbf{v}_{t-1}$, we can approximate the $\mathbf{o}_{t-1}$ by warping the frame $\mathbf{o}_t$ backward in time. Similarly, $\mathbf{o}_{t+1}$ can be calculated by warping $\mathbf{o}_t$ forward by $\mathbf{v}_t$.

The optical flow constraint constraint results in an under-determined system of equations, so a discontinuity preserving Total Variation (TV) regularization term based on robust Huber norm (controlled by H_{μ_1}) can be employed here [8]:

$$\mathcal{L}_{\mathbf{v}_t} = \|\mathbf{D}_s \mathbf{v}_{t,x}\|_{\mathrm{H}_{\mu_1}} + \left\|\mathbf{D}_s \mathbf{v}_{t,y}\right\|_{\mathrm{H}_{\mu_1}} , \tag{S3}$$

where $\mathbf{v}_{t,x}$ and $\mathbf{v}_{t,y}$ are x and y components of $\mathbf{v}_t$, respectively. $\mathbf{D}_s$ represents the spatial discrete gradient operator. This regularization constructs the fourth term in Eq. (6).

We also introduce smoothness constrain in our objective function. In this prior we incorporate a spatial and a temporal components with Huber norm parameterized by H_{μ_2} and H_{μ_3}:

$$\mathcal{L}_{s1} = \|\mathbf{D}_s \mathbf{o}_t\|_{\mathrm{H}_{\mu_2}} , \tag{S4}$$

$$\mathcal{L}_{s2} = \|\mathbf{D}_t \mathbf{o}_t\|_{\mathrm{H}_{\mu_3}} , \tag{S5}$$

where $\mathbf{D}_t$ is the matrix representing the first-order forward finite difference operator along the time direction. The Huber norm is beneficial for optimization constrain since it allows to adjustment the trade-off between the $l1$ and $l2$ norm properties, which provides a compromise between the sparsity and the smoothness [9].

B. Solver

The joint optimization problem in Eq. (6) is challenging to solve, so we address it in an alternating mode between the object set of different frame times $\{\mathbf{o}_t\}_t$ and the flows $\{\mathbf{v}_t\}_t$. We first initialize $\mathbf{o}_t$ in each time by $\mathcal{A}_t^{-1}(\mathbf{b}_t)$, where $\mathcal{A}_t^{-1}(\cdot)$ represents the inverse process of $\mathcal{A}_t(\cdot)$. Then the $\{\mathbf{o}_t\}_t$ and $\{\mathbf{v}_t\}_t$ of two spitted subproblems will be updated alternately.

The deformation field estimation task can be expressed as:

$$\min_{\{\mathbf{v}_t\}_t} \alpha \sum_{t=1}^{\mathrm{T}-1} \|\mathbf{D}_t|\mathbf{o}_t| + \mathbf{D}_s|\mathbf{o}_t|\cdot\mathbf{v}_t\|_1 + \beta \sum_{t=1}^{\mathrm{T}-1} \left(\|\mathbf{D}_s \mathbf{v}_{t,x}\|_{\mathrm{H}_{\mu_1}} + \left\|\mathbf{D}_s \mathbf{v}_{t,y}\right\|_{\mathrm{H}_{\mu_1}} \right) . \tag{S6}$$

The sub-problem for the object reconstruction is given by:

$$\min_{\{\mathbf{o}_t\}_t} \sum_{t=1}^{T} \|\mathcal{A}_t(\mathbf{o}_t) - \mathbf{b}_t\|_2^2 + \alpha \sum_{t=1}^{T-1} \|\mathbf{D}_t|\mathbf{o}_t| + \mathbf{D}_s|\mathbf{o}_t|\cdot\mathbf{v}_t\|_1 + \delta_1 \sum_{t=1}^{T} \|\mathbf{D}_s\mathbf{o}_t\|_{\mathrm{H}_{\mu_2}} + \delta_2 \sum_{t=1}^{T-1} \|\mathbf{D}_t\mathbf{o}_t\|_{\mathrm{H}_{\mu_3}} + \gamma \left(\sum_{t=1}^{T-1} \|\mathcal{A}_{t+1}(\mathrm{warp}(\mathbf{o}_t, \mathbf{v}_t)) - \mathbf{b}_{t+1}\|_2^2 + \sum_{t=2}^{T} \|\mathcal{A}_{t-1}(\mathrm{warp}(\mathbf{o}_t, -\mathbf{v}_{t-1})) - \mathbf{b}_{t-1}\|_2^2 \right). \tag{S7}$$

B.1. Flow Field Estimation

The first term in Eq. (S6) comes from the Horn-Schunck style brightness constancy term [7]. To estimate flow with large displacements, we use a multi-scale strategy [7, 10, 11], constructing a coarse-to-fine structure through a pyramid of down-sampled images. The number of pyramid levels S is adaptive and determined, ensuring that the top level attains a size of approximately 50 pixels along each axis, see Table S5 for the specific setting. To enhance robustness against large perturbations due to shadow and shading reflection artifacts, we employ a preprocessing step via ROF structure-texture decomposition method [11, 12] to the input frames. The resulting texture and structure components are linearly combined in a ratio of 10:1. The parameter settings align with their work [11].

B.2. Object Reconstruction

Assume $\mathbf{o}_t \in \mathbb{C}^{n_1}$, $\mathbf{v}_t \in \mathbb{C}^{n_1}$ and $\mathbf{b}_t, \mathbf{b}_{t+1}, \mathbf{b}_{t-1} \in \mathbb{C}^{n_2}$, we regroup all pixels for each variable into a column vector, resulting in $\mathbf{O} \in \mathbb{C}^{n_1 T}$, $\mathbf{V} \in \mathbb{C}^{n_1 T}$ and $\mathbf{b}, \mathbf{b}_-, \mathbf{b}_+ \in \mathbb{C}^{n_2 T}$. The equation above can be rewritten in block matrix form as:

$$\min_{\mathbf{O}} \|\mathcal{A}(\mathbf{O}) - \mathbf{b}\|_2^2 + \alpha \|\nabla_t|\mathbf{O}| + \nabla_s|\mathbf{O}|\cdot\mathbf{V}\|_1 + \delta_1 \|\nabla_s\mathbf{O}\|_{\mathrm{H}_{\mu_2}} + \delta_2 \|\nabla_t\mathbf{O}\|_{\mathrm{H}_{\mu_3}} + \gamma \left(\|\mathcal{A}(\mathcal{W}^+(\mathbf{O})) - \mathbf{b}_+\|_2^2 + \|\mathcal{A}(\mathcal{W}^-(\mathbf{O})) - \mathbf{b}_-\|_2^2 \right), \tag{S8}$$

where $\mathcal{A}(\cdot)$ the combination of all the $\mathcal{A}_t(\cdot)$ to form an aggregate forward model for the whole image sequence. ∇_s and ∇_t are similarly composed of elements $\mathbf{D}_s$ and $\mathbf{D}_t$. $\mathcal{W}^+(\cdot)$ and $\mathcal{W}^-(\cdot)$ are the forward and backward warping operators integrating for all frames. Then the problem above can be solved by ADMM optimization framework [13]. The optimization alternates between updating the variable $\mathbf{O}$ and the auxiliary variables that represent constraints (e.g., gradients and warps). For efficient solving, the $\mathbf{O}$-update problem is addressed using a splitting strategy [14], which separates the data term from the other terms. Recognizing the complex nature of the variable, we explicitly decompose it into the real and imaginary parts. This decomposition allows us to handle the complex domain intricacies directly within our optimization framework. Specifically, we employ a proximal operator that is adapted to the ℓ_1 norm and Huber norm [9] which operates on both the real and imaginary components of the complex vector.

3. VALIDATION VIA SIMULATION EXPERIMENTS

Figure 2 in the main text presents a comparative analysis of the conventional CS-FP technique and our proposed ST-FP method in a dynamic simulation case. The superior performance of ST-FP, especially against motion-induced artifacts, is also quantitatively substantiated by metrics such as Root Mean Square Error (RMSE), Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR), with our method surpassing CS-FP in all categories. Tables S1 and S2 underscore the accuracy of ST-FP, with notable improvements in SSIM, RMSE, and PSNR values.

We also test the performance of our method on another sample of Ascaris lumbricoides eggs. Figures S2(a) and S2(f) display the amplitude and phase ground truths. Figure S2(b) illustrates the central LR sub-image obtained under the central illumination of the FP system. Figures 3(c) and 3(g) are the HR amplitude and phase reconstruction results from the standard FP method to serve as a baseline. Figures S2(d) and S2(e) show the comparison of CS-FP and ST-FP amplitude outputs, respectively. Figures S2(h) to S2(i) are the results corresponding to phase. ST-FP demonstrates an enhanced ability to maintain phase estimation fidelity and deliver a more consistent, stable reconstruction, especially during object motion, where CS-FP suffers noticeable quality deterioration. Tables S3 and S4 give the quantitative metrics in SSIM, RMSE, and PSNR values, also visualized in Figs. S2(j) and S2(k)

Table S1. Details information of amplitude simulation results (USAF target)

	SSIM ↑		RMSE ↓		PSNR ↑	
Frame	ST-FP	CS-FP	ST-FP	CS-FP	ST-FP	CS-FP
1	**0.6822**	0.6087	**0.1055**	0.1288	**19.5330**	17.8000
2	**0.6831**	0.5837	**0.1065**	0.1439	**19.4540**	16.8406
3	**0.6773**	0.5337	**0.1093**	0.1580	**19.2310**	16.0250
4	**0.7205**	0.5735	**0.0942**	0.1679	**20.5146**	15.4997

Table S2. Details information of phase simulation results (USAF target)

	SSIM ↑		RMSE ↓		PSNR ↑	
Frame	ST-FP	CS-FP	ST-FP	CS-FP	ST-FP	CS-FP
1	**0.5655**	0.4581	**0.1952**	0.2242	**14.1926**	12.9873
2	**0.5667**	0.4792	**0.1951**	0.2225	**14.1961**	13.0537
3	**0.5708**	0.4076	**0.1936**	0.2373	**14.2620**	12.4933
4	**0.5521**	0.4251	**0.2136**	0.2475	**13.4087**	12.1296

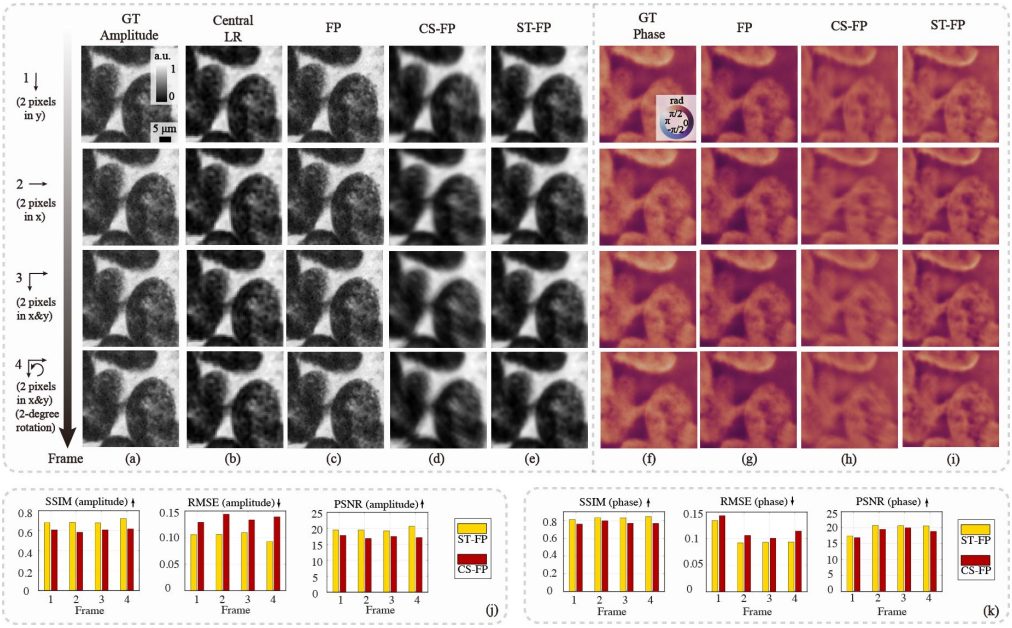

Fig. S2. Comparison of dynamic simulation across four frames using for a sample of Ascaris lumbricoides eggs. Frames simulate distinct motion types: translations along the y-axis, x-axis, combined x-y axes, and a composite translation-rotation motion. (a) GT for amplitude. (b) LR sub-image under central FP illumination. (c)-(e) Amplitude reconstruction for traditional FP method (144 sub-images used), CS-FP, and proposed ST-FP. (f) GT for phase. (g)-(i) are phase results of traditional FP, CS-FP, and our method. (j) and (k) are comparison of amplitude and phase reconstruction quality against the GT, using SSIM, RMSE, and PSNR metrics for both CS-FP and ST-FP.

Table S3. Details information of amplitude simulation results (Ascaris lumbricoides eggs)

	SSIM ↑		RMSE ↓		PSNR ↑	
Frame	ST-FP	CS-FP	ST-FP	CS-FP	ST-FP	CS-FP
1	**0.6959**	0.6491	**0.0912**	0.0966	**20.7989**	20.2970
2	**0.7045**	0.6458	**0.0896**	0.0978	**20.9491**	20.1931
3	**0.7050**	0.6460	**0.0895**	0.0947	**20.9669**	20.4708
4	**0.7111**	0.6534	**0.0917**	0.1051	**20.7528**	19.5678

Table S4. Details information of phase simulation results (Ascaris lumbricoides eggs)

	SSIM ↑		RMSE ↓		PSNR ↑	
Frame	ST-FP	CS-FP	ST-FP	CS-FP	ST-FP	CS-FP
1	**0.8128**	0.7604	**0.1340**	0.1431	**17.4605**	16.8861
2	**0.8339**	0.7969	**0.0921**	0.1061	**20.7156**	19.4862
3	**0.8313**	0.7702	**0.0932**	0.1006	**20.6072**	19.9457
4	**0.8460**	0.7687	**0.0937**	0.1142	**20.5686**	18.8500

4. EXPERIMENTS FOR MOVING TARGETS

A. ST-FP on a red blood cells (RBCs) target

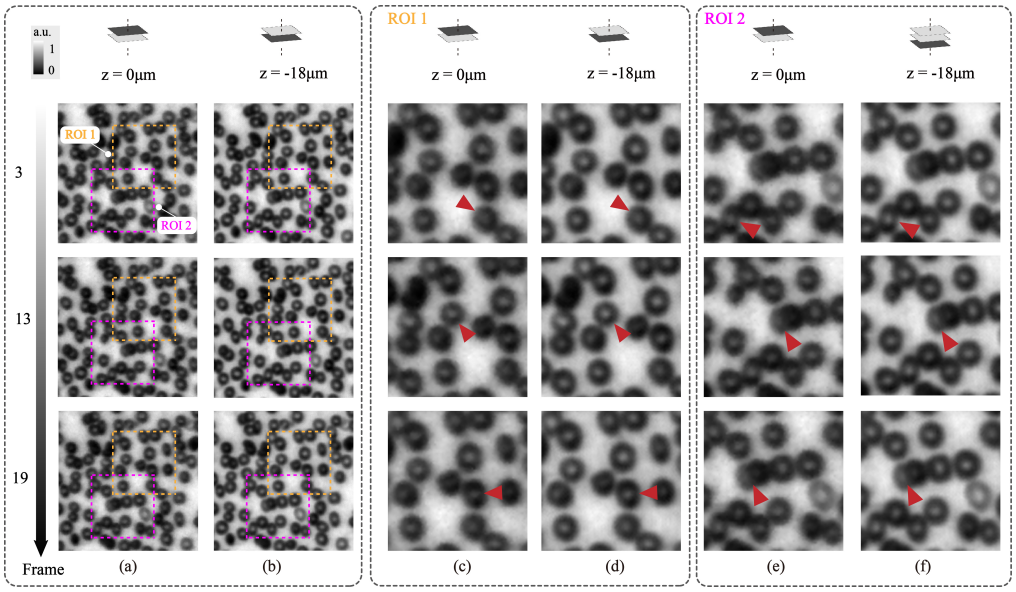

Fig. S3. Digital refocusing capability of ST-FP demonstrated on the RBCs sample across various frames. (a)-(b) Reconstructions at 0 μm (the reference plane) and -18 μm across different time frames, highlighting two regions of interest (ROI 1 and ROI 2). (c)-(d) The amplitude modulation within ROI 1 at selected frames. (e)-(f) The refocusing results for ROI 2. Red arrows highlight regions where digital refocusing has corrected the focus, resulting in sharper edges and clearer features compared to the original blurry areas.

We tested our method by moving static samples in the real world. First, red blood cells (RBCs) were used to assess the performance of ST-FP, as shown in Fig. 3 in the main text. Here we further demonstrate the digital refocusing capability of ST-FP. Note that FP has limitations in post-reconstruction digital refocusing with raw data, hence the integration of defocus correction within the phase-retrieval loop is crucial [15]. Figure S3 shows this capability where the RBCs,

previously analyzed, are digitally refocused across various defocusing distances. The rows in the figure correspond to specific frames in the temporal sequence. Figures S3(a)-(b) present the reconstructions at 0 µm (the reference plane) and -18 µm, where Regions of Interest (ROI 1 and ROI 2) are marked, across various time frames within the imaging sequence. Figures S3(c)-(d) show the digital refocusing results for ROI 1, demonstrating the ST-FP ability to clarify and sharpen the edges at different z-planes. Figures S3(e)-(f) perform a similar demonstration for ROI 2, with the red arrows highlighting the regions where ST-FP has significantly improved the edge definition and contrast of the RBCs.

B. ST-FP on a phase target

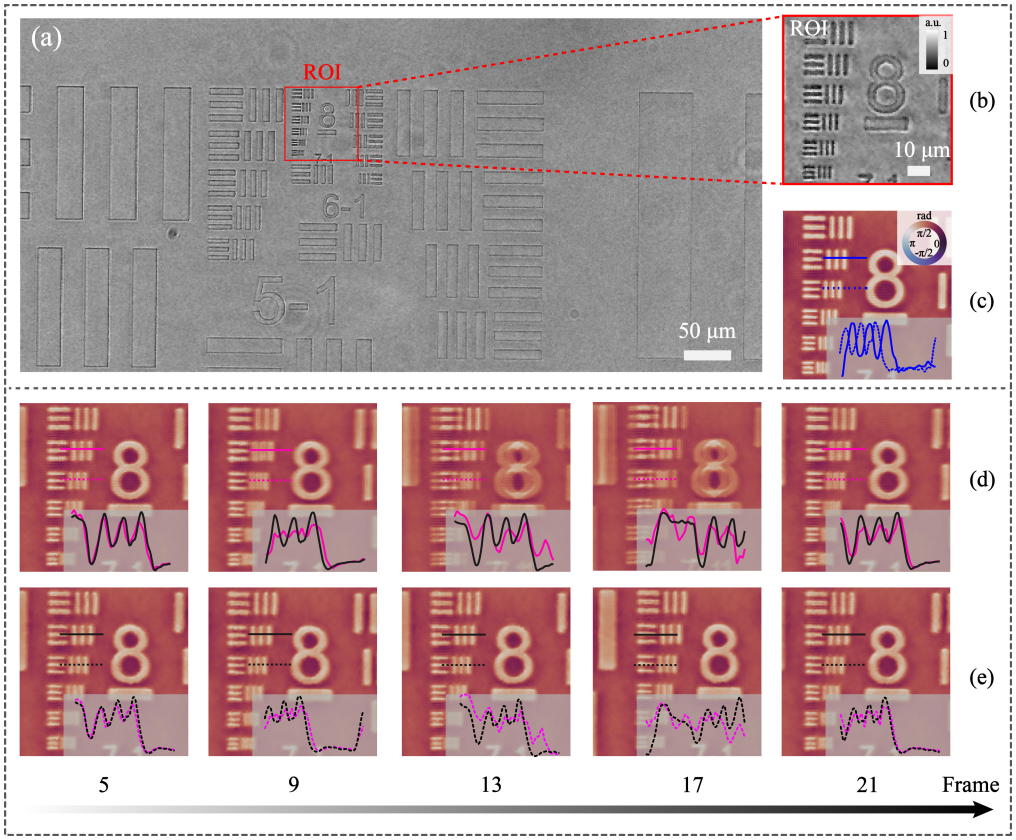

Fig. S4. Comparative analysis of quantitative phase imaging using moved pure phase USAF target. (a) Full FOV LR image of the target acquired with a single center LED illumination. (b) Zoomed view of the ROI. (c) Reconstructed phase using the standard FP method, with the phase profiles of Groups 9 along the selected lines (blue) indicating the resolution capability. (d) Sequence of phase images obtained under dynamic conditions using the conventional CS-FP method, with line plots (pink) showing variability due to motion artifacts. (e) ST-FP results demonstrate enhanced stability and resistance to motion artifacts, with line plots (black) evidencing a consistent and accurate phase estimation. Movement speeds were 1.07, 0.84, 6.39, 7.88, and 0.40 PPF, respectively. The corresponding variance in speed were 0.03, 0.03, 0.08, 0.06, and 0.03 PPF, respectively.

We utilized a quantitative United States Air Force (USAF) phase target to evaluate the resolution capability of our ST-FP method. Initially, employing a conventional FP approach coupled with EPRY [16] for aberration correction, the static sample was imaged to establish a reference for quantitative phase reconstruction, utilizing 144 sub-images. Figures S4(a) and S4(b) present the full FOV original LR image captured under a centered LED and the ROI, respectively. Figure S4(c) shows the HR reconstruction result from the standard FP method. Subsequently, we induced sample motion by manually translating the stage along one axis to simulate dynamic conditions.

We leveraged a Fast Fourier Transform (FFT) based image registration algorithm [17] to estimate the global shift, initializing the optical flow computations. In our analysis, we focused on the second and third elements of group 9, corresponding to spatial resolutions of 574.7 and 645.1 line pairs per millimeter (lp/mm), respectively. Under dynamic imaging scenarios, CS-FP generated phase images that were significantly impacted by movement, resulting in a noticeable degradation of smoothness and clarity, as shown in Fig. S4(d). In contrast, ST-FP exhibited remarkable stability to motion-induced artifacts, thereby maintaining high fidelity and uniformity across the imaging sequence, as shown in Fig. S4(e). This performance enhancement is quantitatively validated by the line plots extracted from the group 9 elements of the target using both imaging methods. Movement speeds were calculated in pixels per frame (PPF) in the raw data dimension for each set of frames and were estimated as 1.07, 0.84, 6.39, 7.88, and 0.40 PPF, respectively. The corresponding variance in speed were 0.03, 0.03, 0.08, 0.06, and 0.03 PPF, respectively. The CS-FP-derived plots show significant variability, indicative of susceptibility to motion, with a result that reflects the struggle to compensate for the target translational disturbances. Conversely, the ST-FP line plots display remarkable consistency and flatness, indicating a more precise and stable phase estimation.

5. EXPERIMENTS FOR LIVE MICROORGANISMS

A. ST-FP on live rotifers

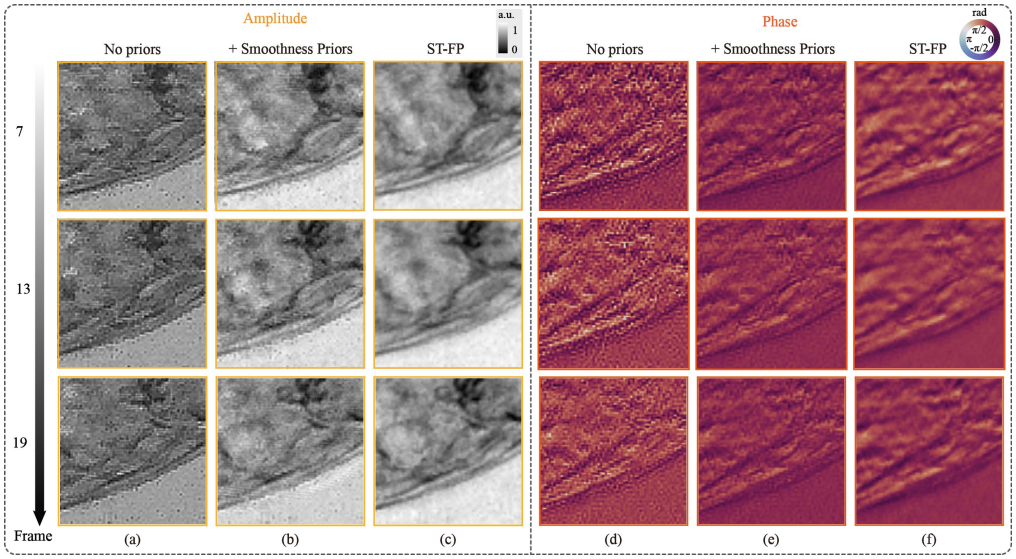

Fig. S5. Comparative analysis of reconstructions for a rotifer dataset (Figure 5 in the main text). (a) Reconstruction by only the data term. (b) Reconstruction incorporating smoothness priors, intended to improve image continuity. (c) Reconstruction with both smoothness and warping priors (ST-FP), is hypothesized to enhance fidelity by aligning with physical and dynamic properties. (d)-(f) The corresponding phase reconstructions for each case.

Figures 5 and 6 in the main text have demonstrated the performance of ST-FP in both detailed preservation and digital refocusing capability. Here we contrast the reconstruction outcomes under three different scenarios to demonstrate the impact of various priors. Figures S5(a) and S5(d) exhibit results relying solely on the data fidelity term, which is marred by noise and a lack of recovery of high-frequency information due to insufficient priors. Incorporating smoothness priors for space and time results in improved continuity, although some areas exhibit blurring effects, as shown in Figs. S5(b) and S5(e). Our methodology, which also considers dynamic uniformity between successive frames, reveals more distinct details of the inner structure and boundaries, as shown in Figs. S5(c) and S5(f).

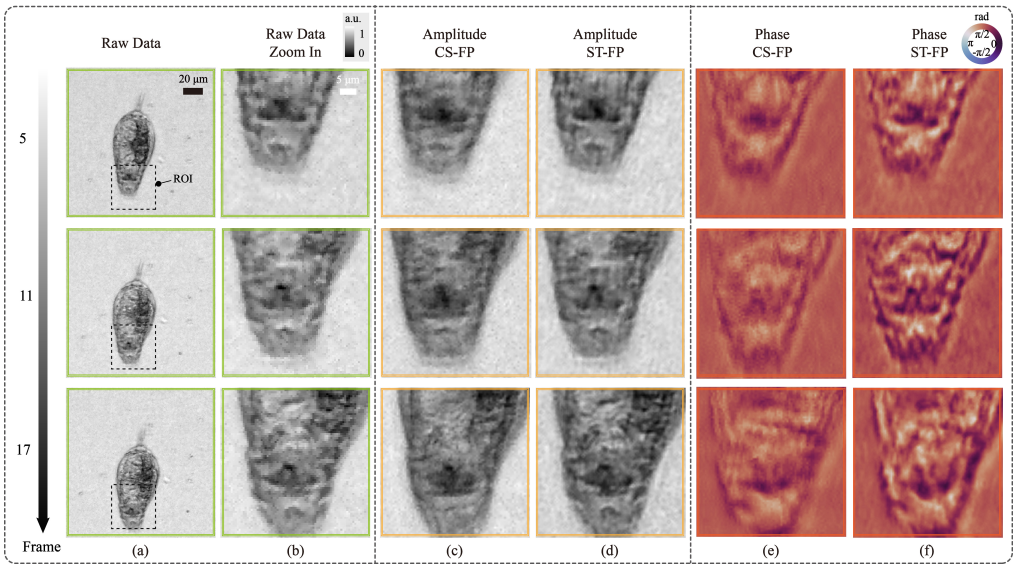

Fig. S6. Results comparison of a live rotifer with faster movement at different times. (a) LR raw data. (b) Region of interest (ROI) from (a). (c) and (d) are amplitude results of CS-FP and ST-FP. (e) and (f) are phase results of CS-FP and ST-FP.

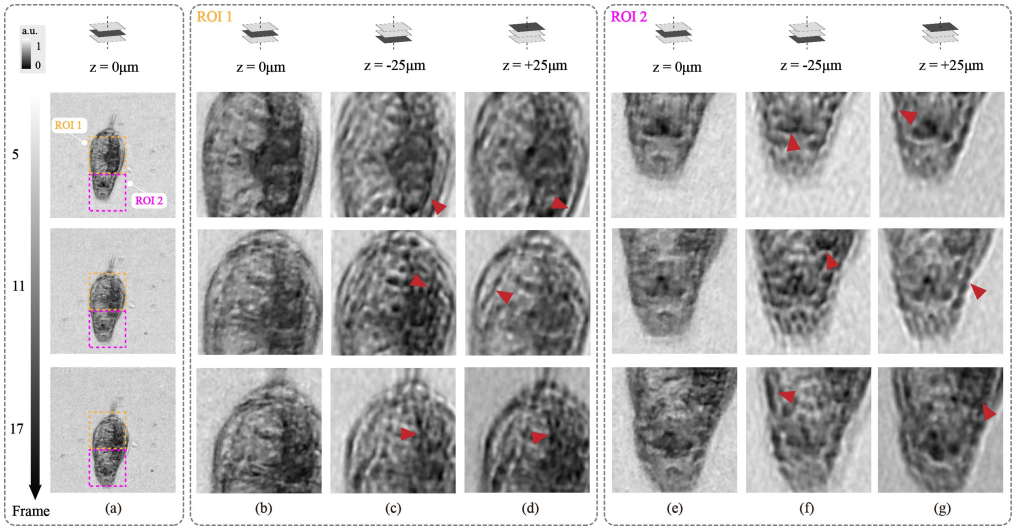

Fig. S7. Temporal digital refocusing on the rotifer dataset (Figure S6) across different propagation distances. (a) Reconstruction of ST-FP in the zero plane across different frame times with two ROIs. (b)-(d) The amplitude profiles of ROI 1 at some specific frames, showing the refocusing effects at distances of 0 μm (the reference plane), -25 μm, and 25 μm, respectively. (e)-(g) The refocusing results for ROI 2. Red arrows highlight areas where edge clarity and definition are significantly improved through digital refocusing.

Next, we present an evaluation of ST-FP on a live rotifer exhibiting fast movement, aiming to demonstrate the robustness of our method. Figure S6 compares the results of the CS-FP and ST-FP at different time frames. Figures S6(a) and S6(b) show the LR raw data and a zoomed-in ROI, respectively. The rapid movement poses a challenge for the CS-FP reconstruction, as the assumption of minimal movement between frames is more likely to be violated, potentially introducing motion artifacts into the reconstructed images. It can be observed in the amplitude and phase results of CS-FP, depicted in Figs. S6(c) and S6(e). The fine details are obscured, and the phase contrast is diminished due to temporal blurring. Conversely, ours is better suited to handle

the quick movements of the rotifer. The amplitude and phase reconstructions produced by ST-FP, as shown in Figs. S6(d) and S6(f), retain sharper details and more consistent phase information across the frames. The results underline the capability of ST-FP to capture the intricate details of the structure and the robustness of the method against rapid movements.

The digital refocusing ability of ST-FP is crucial for observing the three-dimensional dynamics of the rotifer. Figure S7 demonstrates this ability, where we digitally refocus the reconstructed object through various defocusing distances. Each row in the figure corresponds to a particular frame time. The red arrows highlight the enhanced definition and contrast in the rotifer features, affirming the adaptability of ST-FP to fast-moving biological samples.

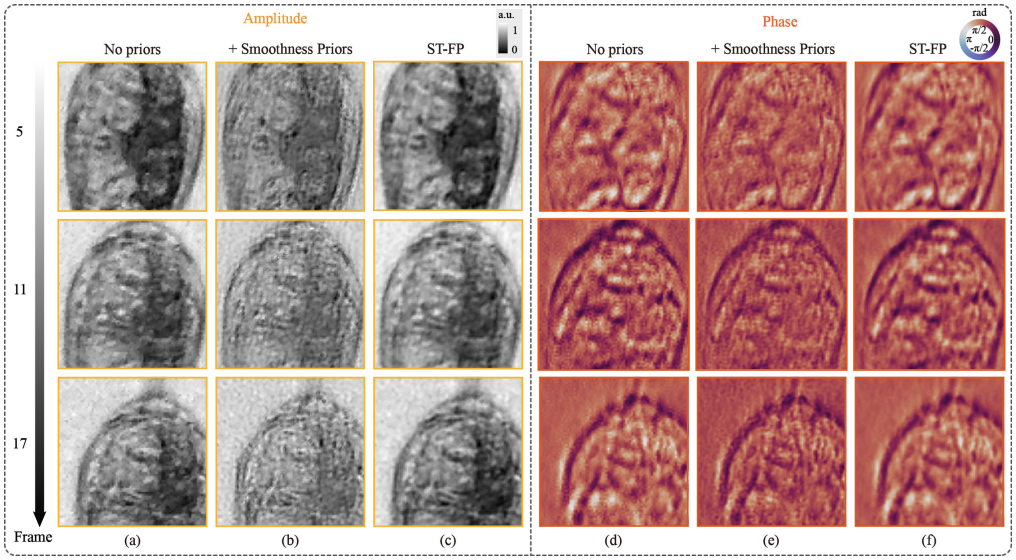

Fig. S8. Comparative analysis of reconstruction techniques for a rotifer dataset (Figure S6). (a) Reconstruction by only the data term. (b) Reconstruction incorporating smoothness priors, intended to improve image continuity. (c) Reconstruction with both smoothness and warping priors (ST-FP). (d)-(f) The corresponding phase reconstructions for each case.

Figure S8 provides the reconstruction outcomes when employing different priors of fast biological motion. Each row corresponds to a different time frame, showcasing the temporal evolution of the imaging process. Figures S8(a) and S8(d) display the amplitude and phase recovery results without any priors. They are based solely on the fidelity to the LR data, which unfortunately includes noise and lacks high-frequency details, leading to noisy results with unresolved internal structures of the rotifer. Figures S8(b) and S8(e) show the results when smoothness priors for space and time are added. However, this can result in a blurring effect, particularly when the subject's motion is rapid. The temporal smoothness constraint may inadvertently blend details from consecutive frames, diminishing temporal resolution and introducing motion artifacts. Figures S8(c) and S8(f) are the reconstructions using ST-FP, which incorporates dynamic uniformity constraints between successive frames. By balancing temporal consistency with spatial detail retention, our approach enforces dynamic uniformity constraints between successive frames, revealing more intricate internal details and defined edges.

B. ST-FP on a live tardigrade

Building upon our analyses before, we extend our examination to a tardigrade sample. This examination aims to further validate the efficacy of the method in high-resolution details of microscopic entities in complex motion. Figure S9 shows the raw data against the reconstructed images, providing a stark contrast in clarity and detail. The raw data and the ROIs of the tardigrade suffer from limitations in resolution and contrast, as shown in Figs. S9(a) and S9(b). Figures S9(c) and S9(d) show the amplitude reconstructions of CS-FP and ST-FP, displaying a progressive enhancement in image quality, with the latter achieving superior resolution and contrast, imperative for detailed biological studies. Figures S9(e) and S9(f) show the phase reconstructions of CS-FP and ST-FP, further exhibiting the advantages of ST-FP, with the phase

images revealing intricate details that are otherwise obscured in CS-FP.

Figure S10 shows the digital refocusing capabilities of our method on the tardigrade sample. The refocused amplitude profiles of ROI 1 and ROI 2 across various distances from the reference plane underline the proficiency in adjusting focus, an essential feature for in-depth biological analysis. The red arrows direct attention to the regions where the ST-FP method distinctly enhances edge clarity and definition.

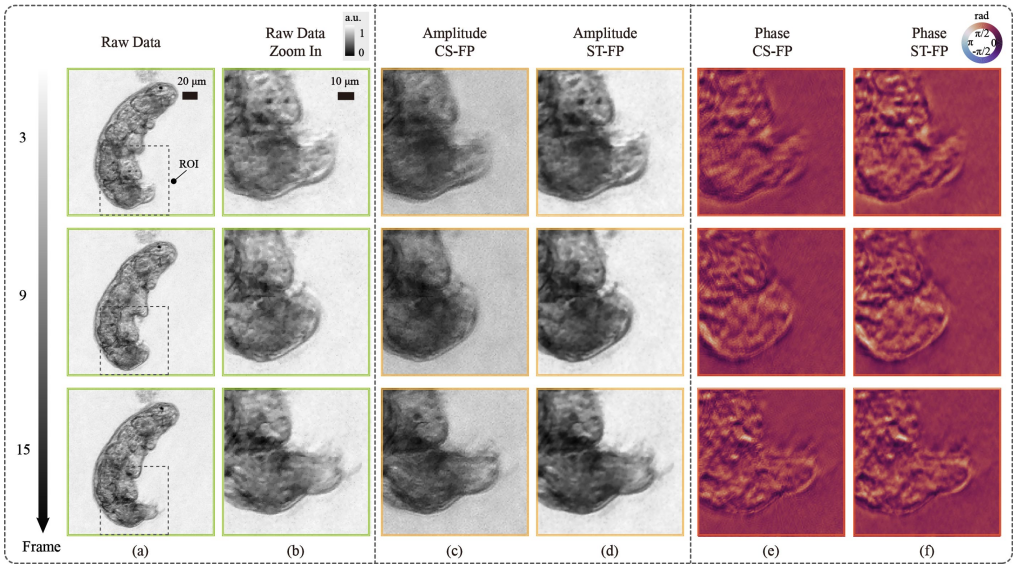

Fig. S9. Results comparison of a live tardigrade at different times. (a) LR raw data. (b) ROI from (a). (c) and (d) Amplitude results of CS-FP and ST-FP. (e) and (f) Phase results of CS-FP and ST-FP.

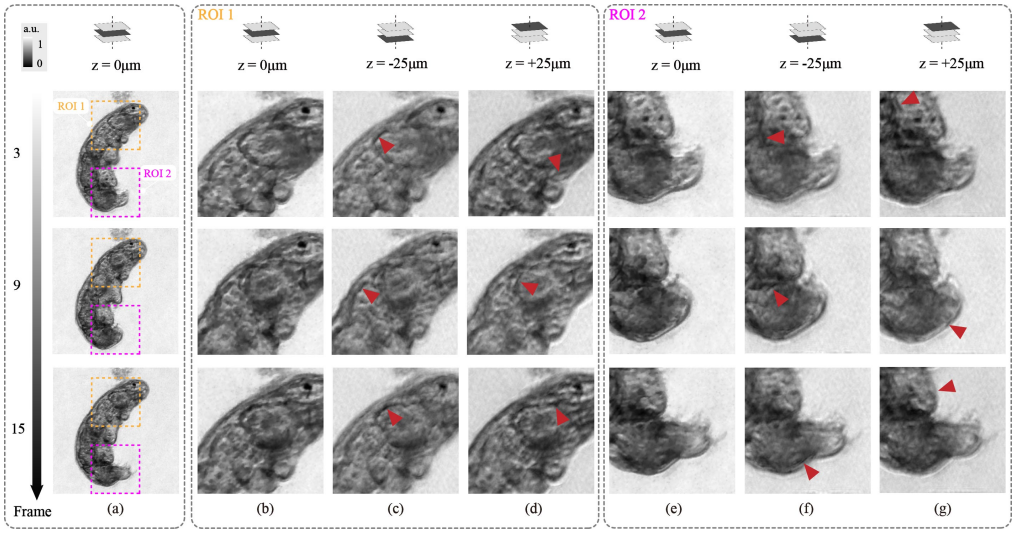

Fig. S10. Temporal digital refocusing on the tardigrade dataset (Figure S9) across different propagation distances. (a) Reconstruction of ST-FP in the zero plane across different times with two ROIs. (b)-(d) The amplitude of ROI 1 at some specific frames, showing the refocusing results at distances of 0 μm (the reference plane), -25 μm, and 25 μm, respectively. (e)-(g) The refocusing results for ROI 2. Red arrows highlight areas where edge clarity is improved through digital refocusing.

Figure S11 compares the reconstruction outcomes under three different cases employing various

priors in the reconstruction process. Figures S11(a) and S11(d) present the amplitude and phase reconstruction without the use of priors. It relies solely on data fidelity, resulting in difficulty in resolving subtle details. Figures S11(b) and S11(e) show the reconstruction when smoothness priors are applied. The images reconstructed with ST-FP reveal an improvement, presenting clearer and more detailed visualizations, as shown in Figs. S11(c) and S11(f).

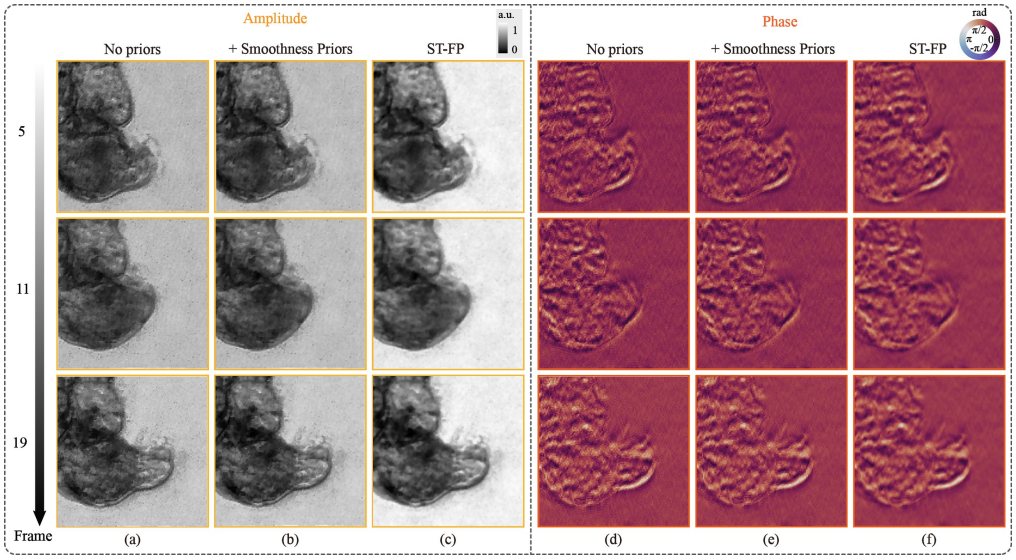

Fig. S11. Comparative analysis of reconstruction techniques for a tardigrade dataset (Figure S9). (a) Reconstruction by only the data term. (b) Reconstruction incorporating smoothness priors, intended to improve image continuity. (c) Reconstruction with both smoothness and warping priors (ST-FP). (d)-(f) The corresponding phase reconstructions for each case.

6. ABLATION STUDY OF CS-FP

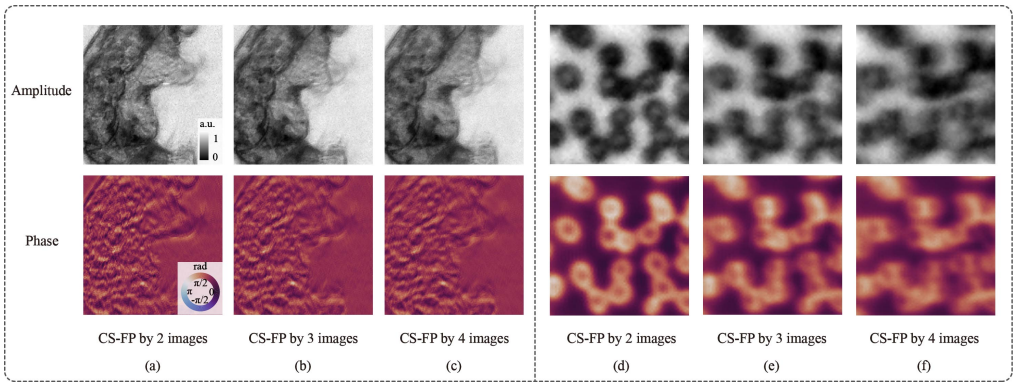

Fig. S12. Comparative Analysis of CS-FP by different image quantities for object reconstruction. (a)-(c) Results of a tardigrade reconstructed using 2, 3, and 4 of CS-FP respectively. (d)-(f) Results of moving RBCs reconstructed with different numbers of images. The two-image approach is shown to significantly reduce the blurring effect, yielding the clearest results.

In our study, we evaluated the effectiveness of CS-FP for reconstructing images of quasi-static objects with a limited number of frames, considering the hardware sampling speed and application context. Traditional CS-FP reconstruction techniques rely on the assumption that the object under observation exhibits negligible motion within the exposure of each frame time, allowing for the use of back-projection techniques with multiple captured images under multiplexed illumination.

To validate the optimal number of images required for accurate reconstruction while minimizing motion artifacts, we carried out an ablation study featuring two distinct samples: a tardigrade and moving RBCs. The reconstruction accuracy typically improves with more images, as they provide additional information. However, it also raises the probability of capturing object movement, which can introduce inaccuracies.

Figures S12(a)-(c) show the amplitude and phase reconstructions of a tardigrade using 2, 3, and 4 images, respectively. Similarly, Figures S12(d)-(f) depict the ablation study for the RBC sample. Our findings indicate that a reconstruction strategy employing two images ($N_{CS} = 2$) achieves an optimal compromise between reconstruction precision and the reduction of motion blur.

7. ST-FP ACROSS DIFFERENT MOVEMENT SCALES

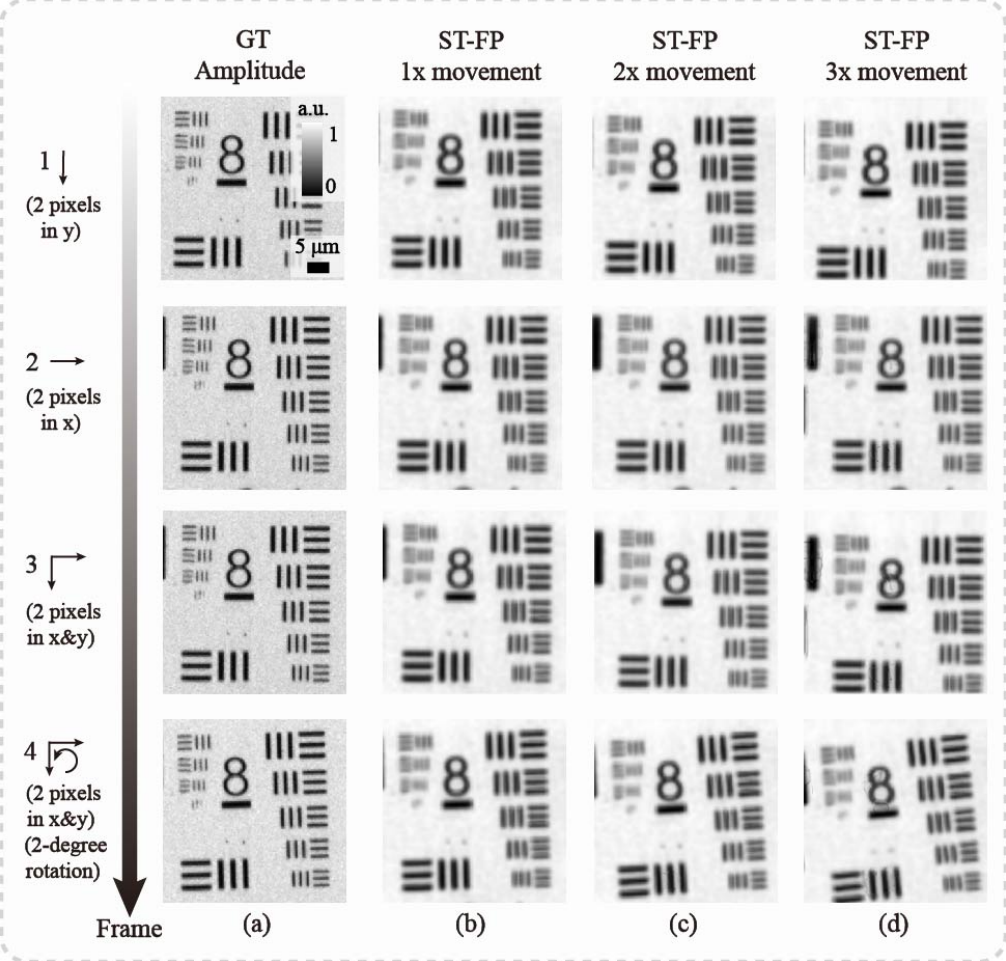

Fig. S13. Simulation results (amplitude) of the complex-valued USAF target in various movement distances. (a) Ground truth (GT) for amplitude under 1× movement, serving for reference. (b)-(d) Amplitude reconstruction of ST-FP under 1×, 2×, and 3× movement. The baseline movement (1×) corresponds to a 2-pixel shift and a 2-degree rotation.

To explore the robustness and performance of our ST-FP method under varying dynamic conditions, we conducted extended simulations to include movements at different scales. The baseline shift is 2 pixels and the baseline rotation is 2 degrees, which we refer to as 1× movement. Figures S13(a) and S14(a) present the ground truth images for amplitude and phase under the baseline movement conditions, serving as a reference for subsequent comparisons. Figures S13(b) and S14(b) display the ST-FP reconstruction results under baseline (1×) movement conditions. These results maintain high fidelity, with clear and accurate reconstructions across all tested

frames. After doubling the movement parameters, the amplitude and phase results shown in Figs. S13(c) and S14(c) continue to demonstrate robust performance by ST-FP, preserving most of the structural details with only slight degradations noticeable in the finer features. At the 3× movement scale, Figures S13(d) and S14(d) show the reconstructions. While ST-FP handles increased dynamics effectively, noticeable artifacts begin to emerge, particularly around finer details such as those in groups 8 and 9. For future improvements, refining the motion estimation algorithms or integrating more sophisticated motion compensation techniques could extend these boundaries, allowing for accurate reconstructions even at higher dynamics.

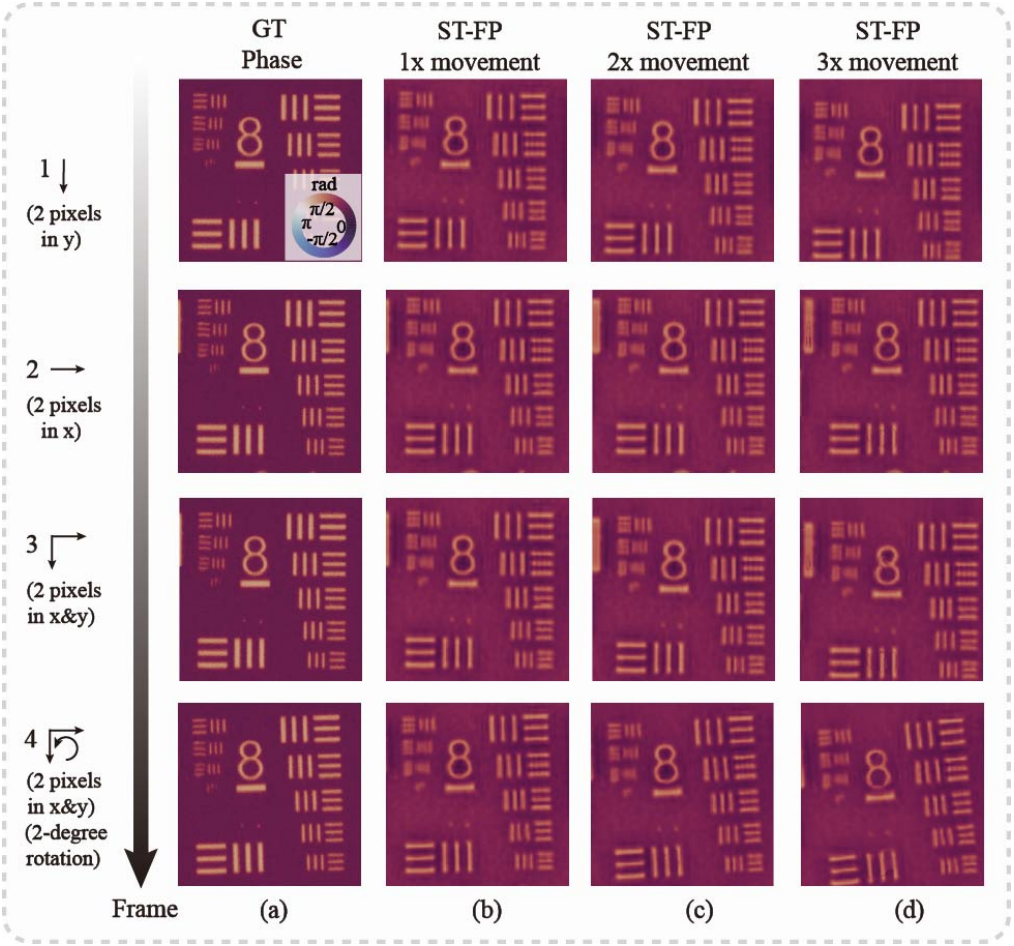

Fig. S14. Simulation results (phase) of the complex-valued USAF target in various movement distances. (a) Ground truth (GT) for phase under 1× movement, serving for reference. (b)-(d) Phase reconstruction of ST-FP under 1×, 2×, and 3× movement. The baseline movement (1×) corresponds to a 2-pixel shift and a 2-degree rotation.

8. ST-FP ON STATIC TARGETS

Figure S15 illustrates the simulation results of moving and static samples. Figures S15(a) and S15(e) show the ground truth for amplitude and phase. Figure S15(b) illustrates the LR image obtained under the central illumination of the FP system. The amplitude reconstruction for the dynamic sample is shown in Fig. S15(c). Here, ST-FP effectively captures the motion-induced variations. Figure S15(d) presents the amplitude reconstruction for a static sample. Notably, the ST-FP method significantly enhances the resolution and clarity even without actual sample motion. For example, observe the top left three line pairs (elements of group 9) in this figure. In the LR image, these lines are blurred, whereas in the static ST-FP reconstruction, they are sharply defined and easily distinguishable. This improvement is indicative of the ability to extract and utilize high-frequency information from the static sample. It shows that the ST-FP approach not

only excels in dynamic imaging but also shows improvement in static imaging settings. The phase reconstructions are shown in Figs. S15(f) and S15(g). In both cases, ST-FP provides a clearer and more detailed phase image compared to the LR image.

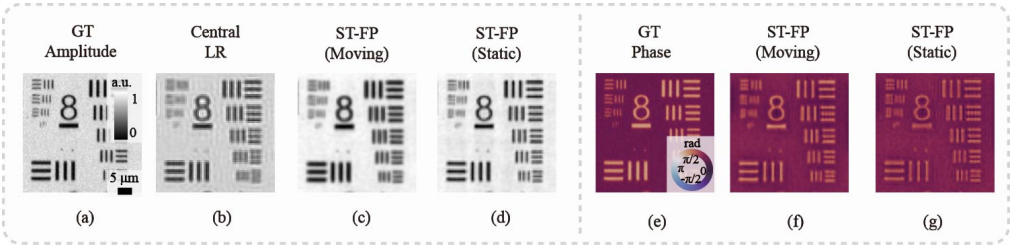

Fig. S15. Simulation results of complex-valued USAF target in moving and static cases. (a) Ground truth (GT) for amplitude. (b) LR sub-image under central FP illumination. (c)-(d) Amplitude reconstruction of ST-FP in moving and static cases. (e) GT for phase. (f)-(g) Phase results of ST-FP in moving and static cases.

9. COMPARISON TO MOTION-CORRECTED FOURIER PTYCHOGRAPHY

Motion-corrected Fourier Ptychography (mcFP) [18] enhances FP by correcting for sample motions during imaging. It starts with a high-resolution spatial spectrum guess and iteratively refines this through motion shift estimations in each sub-spectrum, aligning the reconstructed image with observed data. Our method extends the capabilities of mcFP by using a comprehensive optical flow estimation framework that addresses not only translational but also rotational and non-rigid movements. This broader motion correction capability makes ST-FP particularly effective in dynamic imaging scenarios, surpassing the primarily translational focus of mcFP. Below, we present comparative results of mcFP and ST-FP.

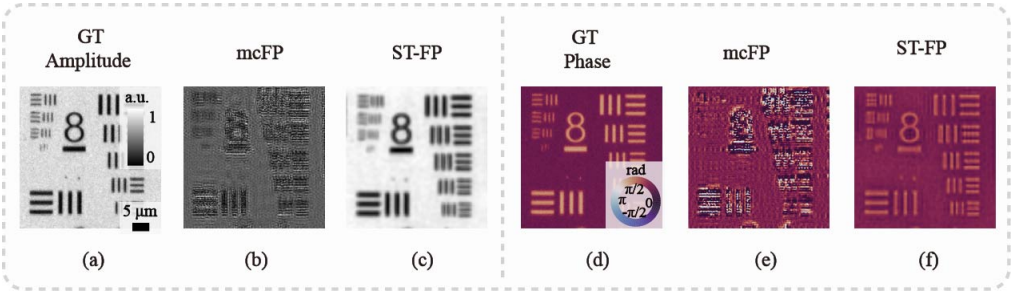

Fig. S16. Simulation results of complex-valued USAF target of mcFP and ST-FP. (a) Ground truth (GT) for amplitude. (b)-(c) Amplitude reconstruction of mcFP and ST-FP. (d) GT for phase. (e)-(f) Phase results of mcFP and ST-FP.

Figures S16(a) and S16(d) show the GT in amplitude and phase. Artifacts apparently appear in the mcFP results, as shown in Figs. S16(b) and S16(e). Our ST-FP method successfully reconstructs complex motions with high fidelity, as shown in Figs. S16(c) and S16(f). While mcFP has shown promising results in scenarios with predictable and translational motions, our approach excels in more dynamically challenging applications in live-cell imaging and other biological settings where sample motion can be unpredictable.

10. PARAMETER SETTINGS

In our framework, meticulous attention was given to the parameter settings across various simulations and experimental scenarios. Certain parameters hold constant throughout all experiments to provide a stable foundation for our comparisons and analyses. Specifically, the weight for forward and backward warping γ is set to 0.001, the parameter controlling the strength of spatial regularization δ_1 is 0.5, and the Huber parameter μ_3 is maintained at 0.001. The down-sampling factor is fixed to 0.6, and the counts of optical flow iterations is set to 10. These parameters were

Table S5. Parameter settings for the simulations and experiments.

Parameter	Dataset							
	Fig. 2	Fig. S2	Fig. 3	Fig. S4	Fig. 4	Fig. 5	Fig. S6	Fig. S9
α	0.04	0.05	0.0002	0.003	0.01	0.02	0.004	0.0015
β	1	1	0.01	0.1	0.1	0.1	0.1	0.01
δ_2	0.001	0.001	0.001	0.001	0.001	0.1	0.001	0.001
μ_1	0.01	0.1	0.1	0.01	0.01	0.04	0.01	0.03
μ_2	0.5	0.5	0.05	0.05	0.05	0.05	0.05	0.05
S	5	5	6	6	6	7	7	7

chosen based on preliminary studies to optimize the balance between accuracy and computational efficiency. Table S5 enumerates the variable parameters adjusted across different datasets. These adjustments were critical for fine-tuning the performance in diverse scenarios, reflecting the adaptability and robustness of our method.

REFERENCES

1. L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for Fourier ptychography with an LED array microscope," Biomed. Opt. Express **5**, 2376–2389 (2014).
2. L. Tian, Z. Liu, L.-H. Yeh, *et al.*, "Computational illumination for high-speed in vitro Fourier ptychographic microscopy," Optica **2**, 904–911 (2015).
3. X. Li, L. Li, X. Liu, *et al.*, "Dictionary-based compressive Fourier ptychography," Opt. Lett. **47**, 2314–2317 (2022).
4. J. Sun, C. Zuo, J. Zhang, *et al.*, "High-speed Fourier ptychographic microscopy based on programmable annular illuminations," Sci. Rep. **8**, 7669 (2018).
5. M. Sun, Y. Li, G. Huang, *et al.*, "Self-adapting search algorithm for Fourier ptychographic microscopy," Opt. Quant. Electron **53**, 704 (2021).
6. A. Zhou, N. Chen, H. Wang, and G. Situ, "Analysis of Fourier ptychographic microscopy with half of the captured images," J. Opt. **20**, 095701 (2018).
7. E. Meinhardt-Llopis and J. Sánchez, "Horn-Schunck optical flow with a multi-scale strategy," Image Process. Line **3**, 151–172 (2013).
8. M. Werlberger, W. Trobin, T. Pock, *et al.*, "Anisotropic Huber-l1 optical flow," in *Proc. Brit. Mach. Vis. Conf.*, (2009), pp. 1–11.
9. A. Beck, *First-Order Methods in Optimization* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017).
10. J. Sánchez Pérez, E. Meinhardt-Llopis, and G. Facciolo, "TV-l1 optical flow estimation," Image Process. Line **3**, 137–150 (2013).
11. A. Wedel, T. Pock, C. Zach, *et al.*, *An Improved Algorithm for TV-L1 Optical Flow* (Springer-Verlag, Berlin, Heidelberg, 2009), pp. 23–45.
12. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Phys. D: Nonlinear Phenom. **60**, 259–268 (1992).
13. L. Li, X. Wang, and G. Wang, "Alternating direction method of multipliers for separable convex optimization of real functions in complex variables," Math. Probl. Eng. **2015** (2015).
14. A. Beck, "On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes," SIAM J. Optim. **25**, 185–209 (2015).
15. H. Zhou, C. Shen, M. Liang, and C. Yang, "Analysis of postreconstruction digital refocusing in Fourier ptychographic microscopy," Opt. Eng. **61**, 073102 (2022).
16. X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," Opt. Express **22**, 4960–4972 (2014).
17. B. Reddy and B. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," IEEE Trans. Image Process. **5**, 1266–1271 (1996).

18. L. Bian, G. Zheng, K. Guo, *et al.*, "Motion-corrected Fourier ptychography," Biomed. Opt. Express **7**, 4543–4553 (2016).