TransiT: Transient Transformer for Non-line-of-sight Videography

Supplementary Material

1. Distortion Model under Fast Scanning

1.1. Data Acquisition

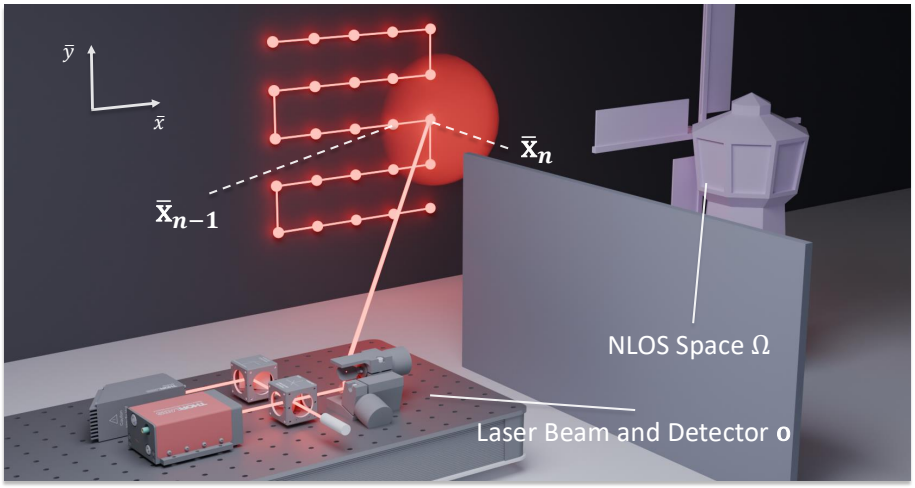

Figure 1. Confocal Non-Line-of-Sight imaging setup.

Fig. 1 illustrates a conceptual diagram of the confocal Non-Line-of-Sight (NLOS) imaging system. The illumination and the detection points are at the same location $\bar{\mathbf{x}}_n = (\bar{x}_n, \bar{y}_n, \bar{z}_n = 0)$ on the relay wall. In our NLOS imaging system, as the pulsed laser scans across the relay wall, we continuously record the arrival times of individual photons detected by a single photon avalanche diode (SPAD) located at $\mathbf{o} = (x_o, y_o, z_o)$. The moment $t = 0$ is when the laser emits each pulse at the same location $\mathbf{o}$, meaning all recorded arrival times t are measured relative to the laser's emission time.

The total time capturing a single frame is then divided uniformly into a grid. Each grid cell represents a specific illumination and detection point $\bar{\mathbf{x}}_n$ on the wall in the confocal setup. The real-measured data for a single grid cell $\tau_*(\bar{\mathbf{x}}_n, t)$ then can be formulated as:

$$\tau_*(\bar{\mathbf{x}}_n, t) = \sum_{i=1}^{N} \delta(t - t_i), \tag{1}$$

where t_i is the arrival time of the i-th photon detected during the time allocated to grid cell $\bar{\mathbf{x}}_n$, and N is the total number of photons detected for that grid cell.

1.2. Confocal NLOS Imaging Forward Model

Existing NLOS imaging algorithms require input data that exclude the initial photon path from the laser source to the relay wall point ($\mathbf{o} \rightarrow \bar{\mathbf{x}}_n$) and the final path from the relay wall back to the detector ($\bar{\mathbf{x}}_n \rightarrow \mathbf{o}$). Specifically, under the confocal setting, we define $t = 0$ as the moment when the laser beam illuminates a point on the relay wall at $\bar{\mathbf{x}}_n$ rather than when it is emitted from the laser source at $\mathbf{o}$. From this point, photons scatter outward as a spherical wave toward the hidden scene. The arrival times that form the ideal transients are recorded when the photons return to the same point $\bar{\mathbf{x}}_n$ on the relay wall, instead of back to the detector at $\mathbf{o}$. We refer to such input data as **ideal transients** $\tau(\bar{\mathbf{x}}_n, t)$. Following O'Toole et al. [3], the forward model under the confocal setting can be formulated as:

$$\tau(\bar{\mathbf{x}}_n, t) = \frac{1}{r^4} \iiint_{\Omega} \rho(\mathbf{x}) \cdot \delta(2\|\bar{\mathbf{x}}_n - \mathbf{x}\| - tc)\, d\mathbf{x}, \tag{2}$$

where Ω represents the NLOS space and ρ the albedo of the hidden scene at any point $\mathbf{x} = (x, y, z)$. The Dirac delta function δ converts time t to the distance $r = 2\|\bar{\mathbf{x}}_n - \mathbf{x}\|/c$ from the illumination point to the hidden scene and back to the detection point, with c the speed of light.

1.3. Distortion from Fast Scanning

The direction of the laser beam is guided by a scanning galvanometer system, which has a minimum response time of ∼0.4 ms from when it receives a signal to when it completes the movement. For typical static NLOS imaging, the scanning time per point is relatively long, ranging from several to hundreds of milliseconds. In these cases, we can assume that the laser beam moves instantly from one point to the next within the scanning grid, allowing us to neglect the time it takes to travel between scanning points.

However, in our fast-scanning scenario, in order to scan a 16×16 grid at 10 FPS, the scanning time per grid point is also ∼0.4 ms. This means that as soon as the beam reaches one point, it immediately receives the signal for the next target location and begins moving toward it. Unlike the ideal per-point scanning assumed in static NLOS imaging, this creates a continuous scanning path rather than discrete jumps between points. **Distorted ideal transients** under fast-scan scenario $\hat{\tau}(\bar{\mathbf{x}}_n, t)$ is therefore an integral of photon events illuminated along the path between points:

$$\hat{\tau}(\bar{\mathbf{x}}_n, t) = \frac{1}{\|S\|} \int_S \tau\left(\bar{\mathbf{x}}_n - \mathbf{s}, t\right) \, d\mathbf{s}, \tag{3}$$

where $S = \bar{\mathbf{x}}_n - \bar{\mathbf{x}}_{n-1}$ represents the one-dimensional path between adjacent scanning points, $\frac{1}{\|S\|}$ the normalization term. Under fast-scan scenario, $\|S\| \neq 0$, otherwise $\hat{\tau}(\bar{\mathbf{x}}_n, t) = \tau(\bar{\mathbf{x}}_n, t)$ becomes slow-scan scenario. The serpentine scanning pattern restricts adjacent points to change along either $\bar{x}$- or $\bar{y}$-axis on the relay wall, allowing us to simplify this path integral to a single dimension.

1.4. Distortion from Real Measurement

Real measurement $\tau_*(\bar{\mathbf{x}}_n, t)$ records two additional photon propagation paths compared to the ideal transients in Eq. 2: one from the laser to the relay wall and one from the relay wall to the detector:

$$\tau_*(\bar{\mathbf{x}}_n, t) = \frac{1}{\|\bar{\mathbf{x}}_n - \mathbf{o}\|^2 r^4} \cdot \iiint_\Omega \rho(\mathbf{x}) \cdot \delta\left(2(\|\bar{\mathbf{x}}_n - \mathbf{x}\| + \|\bar{\mathbf{x}}_n - \mathbf{o}\|) - tc\right) \, d\mathbf{x}, \tag{4}$$

where $1/\|\bar{\mathbf{x}}_n - \mathbf{o}\|^2$ is the attenuation from a diffuse reflection after photons arrive at the detection point. From Eq. 2 and Eq. 4, we have the relationship between the real measurement $\tau_*(\bar{\mathbf{x}}_n, t)$ and the ideal transients $\tau(\bar{\mathbf{x}}_n, t)$:

$$\tau_*(\bar{\mathbf{x}}_n, t) = \frac{1}{\|\bar{\mathbf{x}}_n - \mathbf{o}\|^2} \cdot \tau\left(\bar{\mathbf{x}}_n, t - \frac{2\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{c}\right) \tag{5}$$

$$\tau(\bar{\mathbf{x}}_n, t) = \|\bar{\mathbf{x}}_n - \mathbf{o}\|^2 \cdot \tau_*\left(\bar{\mathbf{x}}_n, t + \frac{2\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{c}\right) \tag{6}$$

From Eq. 6, one can notice that converting a piece of real measured data τ_* to the ideal transient τ involves two operations: a temporal shift by $\frac{2\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{c}$ and a scaling by $\|\bar{\mathbf{x}}_n - \mathbf{o}\|^2$is a temporal shifting and a scaling. Combining Eq. 3, Eq. 5, and Eq. 6, the **distorted real measurement** under fast-scan scenario $\hat{\tau}_*(\bar{\mathbf{x}}_n, t)$ can be derived as:

$$\hat{\tau}_*(\bar{\mathbf{x}}_n, t) = \frac{1}{\|\bar{\mathbf{x}}_n - \mathbf{o}\|^2} \cdot \hat{\tau}\left(\bar{\mathbf{x}}_n, t - \frac{2\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{c}\right) \tag{7}$$

$$= \frac{1}{\|S\|} \int_S \frac{1}{\|\bar{\mathbf{x}}_n - \mathbf{o}\|^2} \cdot \tau\left(\bar{\mathbf{x}}_n - \mathbf{s}, t - \frac{2\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{c}\right) \, d\mathbf{s} \tag{8}$$

$$= \frac{1}{\|S\|} \int_S \left(\frac{\|(\bar{\mathbf{x}}_n - \mathbf{s}) - \mathbf{o}\|}{\|\bar{\mathbf{x}}_n - \mathbf{o}\|}\right)^2 \cdot \tau_*\left(\bar{\mathbf{x}}_n - \mathbf{s}, t + \frac{2(\|\,(\bar{\mathbf{x}}_n - \mathbf{s}) - \mathbf{o}\| - \|\bar{\mathbf{x}}_n - \mathbf{o}\|)}{c}\right) \, d\mathbf{s}, \tag{9}$$

where $\hat{\tau}$ in Eq. 7 is the distorted ideal transients converted from distorted real measurement $\hat{\tau}_*$ by applying Eq. 6, $\bar{\mathbf{x}}_n - \mathbf{s}$ represents a specific point on the path between $\bar{\mathbf{x}}_n$ and $\bar{\mathbf{x}}_{n-1}$, while $(\bar{\mathbf{x}}_n - \mathbf{s}) - \mathbf{o}$ the directed line between that point and the detector.

Comparing Eq. 3 and Eq. 9, we observe that the distortion in the real measurement arises because the shifting and scaling operations we perform to convert the real measurement into the ideal transient do not align with the scaling and shifting that should be applied to each individual photon.

In other words, during the fast-scan scenario, the laser beam moves continuously between scanning points, and photons are emitted and detected while the beam is in motion. Each photon effectively has a unique path and timing based on the exact position of the laser beam when that photon was emitted. Ideally, to accurately convert the real measurement to the ideal transient, we would need to apply a specific temporal shift and scaling factor to each photon individually, accounting for its precise emission time and path length.

However, in practice, we apply uniform shifting and scaling based on the nominal positions of the scanning points $\bar{\mathbf{x}}_n$ and the assumption of instantaneous beam movement. This uniform approach does not capture the variations introduced by the continuous movement of the laser during fast scanning. The discrepancy between the actual photon paths and the assumed paths in the uniform scaling and shifting leads to distortions.

1.5. Overall Distortion Model

Overall, we present the overall distortion model that combines both fast-scan distortion (Sec. 1.3) and real measurement distortion (Sec. 1.4). Our objective is to transform the undistorted ideal transients into distorted transients that closely simulate the real measurements obtained during fast scanning.

Rewriting Eq. 3 using Eq. 5, the relation between distorted ideal transients converted by real measurement, $\hat{\tau}$, and undistorted ideal transients τ becomes:

$$\hat{\tau}(\bar{\mathbf{x}}_n, t) = \|\bar{\mathbf{x}}_n - \mathbf{o}\|^2 \cdot \hat{\tau}_* \left(\bar{\mathbf{x}}_n, t + \frac{2\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{c}\right) \tag{10}$$

$$= \frac{1}{\|S\|} \int_S \|\bar{\mathbf{x}}_n - \mathbf{o}\|^2 \cdot \tau_* \left(\bar{\mathbf{x}}_n - \mathbf{s}, t + \frac{2\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{c}\right) d\mathbf{s} \tag{11}$$

$$= \frac{1}{\|S\|} \int_S \left(\frac{\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{\|(\bar{\mathbf{x}}_n - \mathbf{s}) - \mathbf{o}\|}\right)^2 \cdot \tau \left(\bar{\mathbf{x}}_n - \mathbf{s}, t + \frac{2(\|\bar{\mathbf{x}}_n - \mathbf{o}\| - \|(\bar{\mathbf{x}}_n - \mathbf{s}) - \mathbf{o}\|)}{c}\right) d\mathbf{s}, \tag{12}$$

This integral can then be discretized to:

$$\hat{\tau}(\bar{\mathbf{x}}_n, t) \approx \frac{1}{\|S\|} \sum_{i=1}^{M} \left(\frac{\|\bar{\mathbf{x}}_n - \mathbf{o}\|}{\|(\bar{\mathbf{x}}_n - i\Delta\mathbf{s}) - \mathbf{o}\|}\right)^2 \cdot \tau \left(\bar{\mathbf{x}}_n - i\Delta\mathbf{s}, t + \frac{2(\|\bar{\mathbf{x}}_n - \mathbf{o}\| - \|(\bar{\mathbf{x}}_n - i\Delta\mathbf{s}) - \mathbf{o}\|)}{c}\right) \Delta\mathbf{s}, \tag{13}$$

where M represents the number of sampled points between two adjacent scanning points, and $\Delta\mathbf{s}$ the step size between sample points along the path S. By applying the formula Eq. 13, we can introduce distortion to the noiseless grid simulation data, generating a large amount of data that considers high-speed scanning distortion.

2. More Comparisons

We additionally compared our method with f-k [2] and PnP [4] on more synthetic cases. From the Fig. 2 and Fig. 3, it is evident that our method still achieves more clear and accurate results.

We also compare our method with USM [1]. The USM method is designed primarily for static NLOS reconstruction, and shows limited performance in high-speed dynamic scenarios. In Fig. 4, we provide additional results of three synthetic scenes by USM for dynamic NLOS reconstruction. It is evident that these results are blurry and inferior to ours, especially for the fast-moving propeller.

3. More Ablation Study

We additionally assess performance of TransiT on the real-measured data from f-k [2]. Fig. 5 shows that it is evident that our method still achieves more clear and accurate results on real-measured data.

In Table 1, we also provide additional ablation results to verify our algorithm and show the effectiveness of several steps.

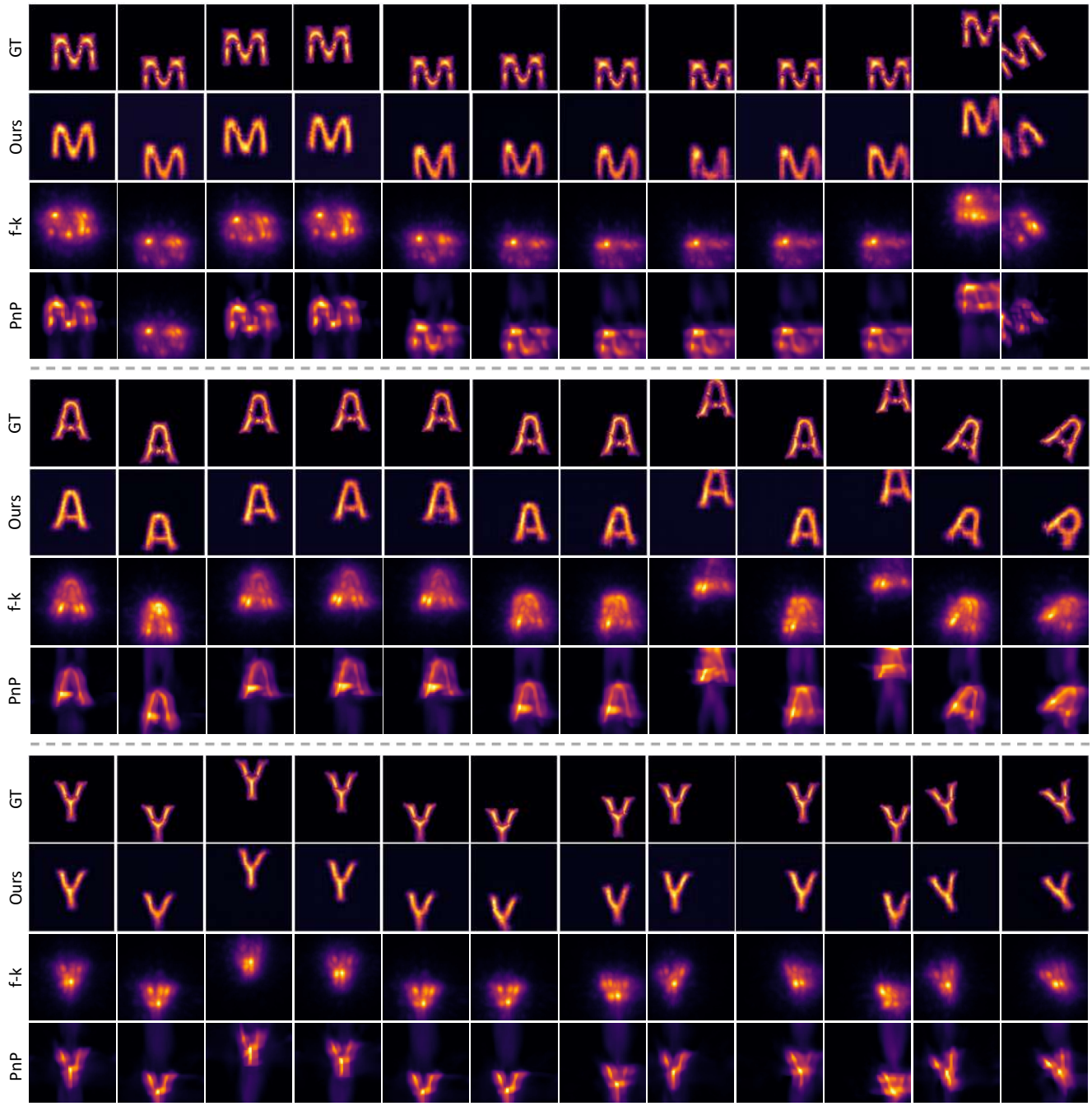

Figure 2. **More comparison of synthetic results.** From top to bottom: Ground truth, ours, f-k, and PnP. The results are reconstructed across multiple frames of 16×16 of noisy synthetic data for different objects — a character 'M', a character 'A', and a character 'Y'.

Table 1. Additional results of ablation.

Method	ED↓	CS↑	SSIM↑	PSNR↑
w/o transients compression	0.0817	0.8465	0.6065	16.78
w/o feature fusion	0.0698	0.9370	0.8314	23.78
w/o space attention	0.0813	0.8875	0.8165	21.13
w/o temporal attention	0.0703	0.9107	0.8234	23.73
full model	**0.0613**	**0.9490**	**0.8493**	**25.41**

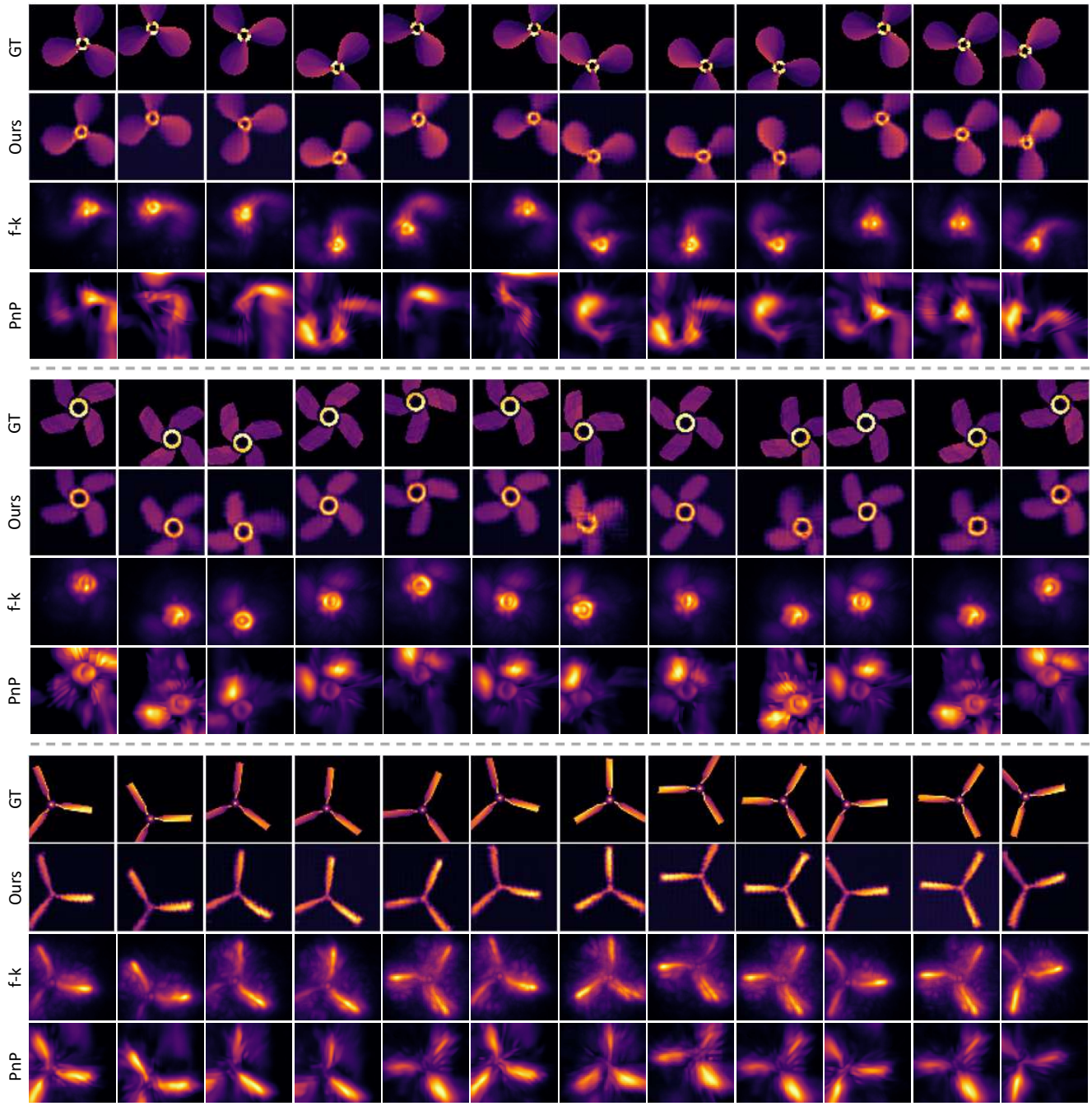

Figure 3. **More comparison of synthetic results.** From top to bottom: Ground truth, ours, f-k, and PnP. The results are reconstructed across multiple frames of 16×16 of noisy synthetic data for different objects — a propeller, a propeller, and a windmill.

References

[1] Yue Li, Yueyi Zhang, Juntian Ye, Feihu Xu, and Zhiwei Xiong. Deep non-line-of-sight imaging from under-scanning measurements. In Advances in Neural Information Processing Systems, pages 1–12, 2023. 3

[2] David B Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast fk migration. ACM Transactions on Graphics (ToG), 38(4):1–13, 2019. 3

[3] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. Nature, 555(7696):338–341, 2018. 1

[4] Juntian Ye, Yu Hong, Xiongfei Su, Xin Yuan, and Feihu Xu. Plug-and-play algorithms for dynamic non-line-of-sight imaging. ACM Transactions on Graphics, 43(5):1–12, 2024. 3

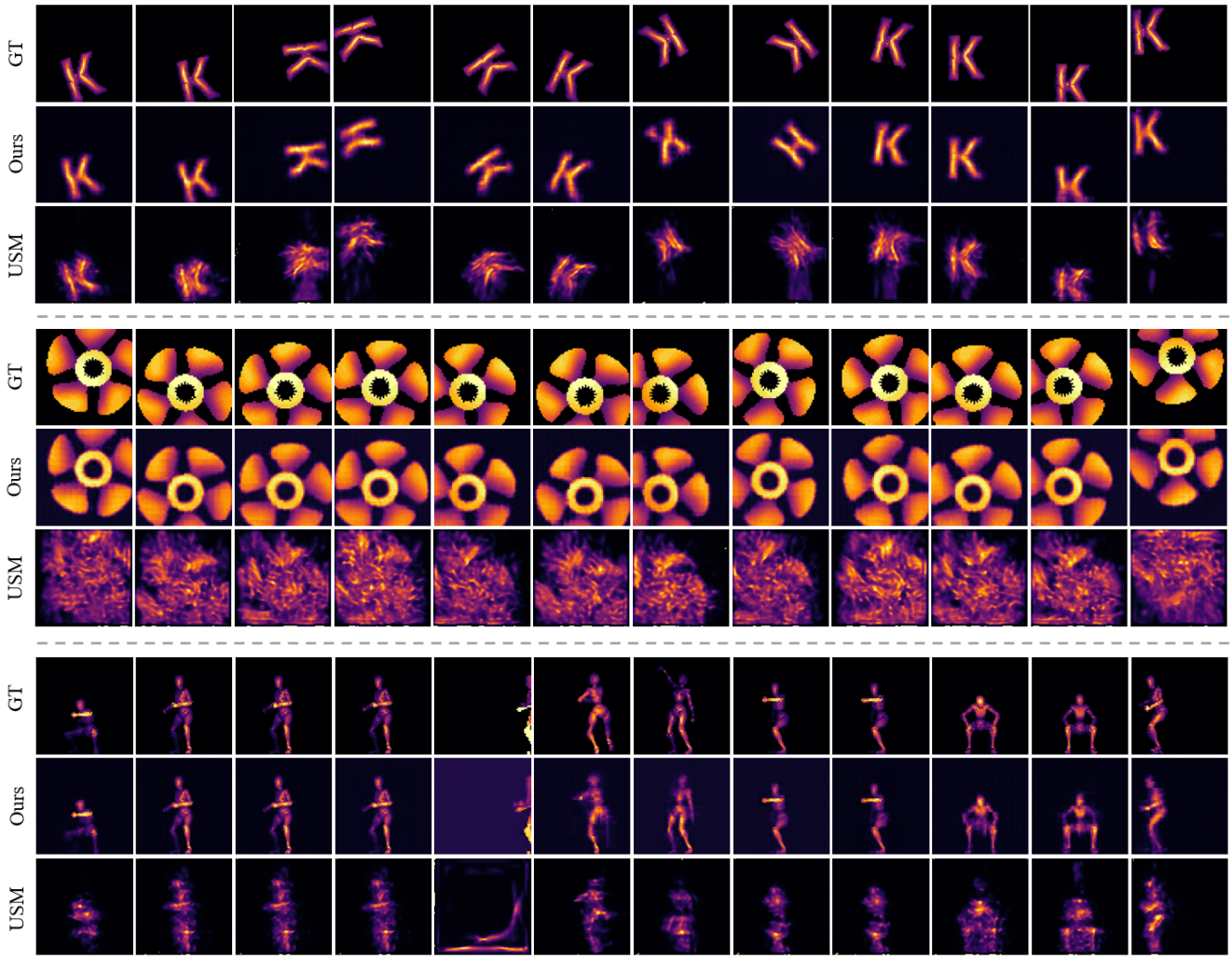

Figure 4. **More comparison of different methods.** From top to bottom: Ground truth, ours and USM. The results are reconstructed across multiple frames of 16×16 of noisy synthetic data for different objects — a character 'K', a propeller, and a human.

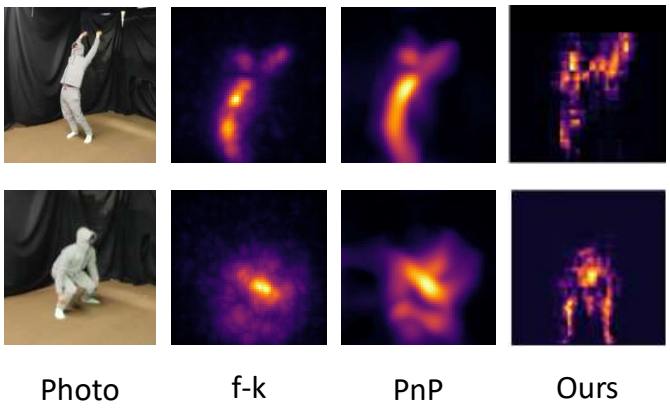

Figure 5. More ablation study. The results are reconstructed from transients of human on real-measured data from f-k.