# Fisher Information Guidance for Learned Time-of-Flight Imaging (Supplementary Material)

Jiaqu Li, Tao Yue, Sijie Zhao, Xuemei Hu
School of Electronic Science and Engineering, Nanjing University, Nanjing, China
jqli@smail.nju.edu.cn, yuetao@nju.edu.cn, sjzhao@smail.nju.edu.cn, xuemeihu@nju.edu.cn

The supplementary material is composed of the following parts: a) more evaluation metrics of depth reconstruction, b) generalization ability on other depth datasets. Since we currently working on the following up research works based upon the proposed method in this paper, we promise to open source the code soon.

**More evaluation metrics of depth reconstruction.** Here, we show comparisons in other metrics for depth performance evaluation [2]. These metrics are defined as:

Root Mean Squared Error (RMSE) : $\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(\widehat{y_i}-y_i)^2}$,

Absolute Relative Difference (Abs. Rel) : $\frac{1}{n}\sum\limits_{i=1}^{n}\frac{|\widehat{y_i}-y_i|}{\widehat{y_i}}$,

Squared Relative Difference (Sq. Rel) : $\frac{1}{n}\sum\limits_{i=1}^{n}\frac{(\widehat{y_i}-y_i)^2}{\widehat{y_i}}$,

Threshold Accuracy ($\delta$) : $\max(\frac{\widehat{y_i}}{y_i},\frac{y_i}{\widehat{y_i}}) = \delta < thr$ for $thr$ = 1.04.

Tab. 1 lists the average performance metrics of the three noise levels on the NYU-V2 test dataset. As shown, the proposed method elegantly outperforms the existing iToF imaging methods, coding functions, and reconstruction methods, in these metrics.

| (a) Overall Performance | RMSE (mm) ↓ | Abs. Rel ↓ | Sq. Rel ↓ | $\delta$ ↑ |
|---|---|---|---|---|
| Sinusoid + PS [4] | 385.652 | 0.170 | 31.975 | 0.188 |
| Square + PS [4] | 267.312 | 0.109 | 15.228 | 0.286 |
| Hamilton [3] | 320.479 | 0.095 | 20.977 | 0.421 |
| DeepToF [8] | 92.765 | 0.053 | 2.482 | 0.527 |
| (b) Coding Function | RMSE (mm) ↓ | Abs. Rel ↓ | Sq. Rel ↓ | $\delta$ ↑ |
| Sinusoid | 66.641 | 0.041 | 1.437 | 0.640 |
| Dual-freq Sinusoid | 49.330 | 0.018 | 0.790 | 0.920 |
| Square | 56.717 | 0.038 | 0.989 | 0.646 |
| Hamiltonian [3] | 30.466 | 0.016 | 0.235 | 0.926 |
| (c) Recovery Method | RMSE (mm) ↓ | Abs. Rel ↓ | Sq. Rel ↓ | $\delta$ ↑ |
| DeepToF [8] | 73.334 | 0.153 | 1.607 | 0.703 |
| MaskToF [1] | 40.468 | 0.019 | 0.350 | 0.893 |
| Our method | 23.513 | 0.011 | 0.129 | 0.972 |

Table 1. Comparison in other metrics of the overall performance, coding functions, and reconstruction methods. Note that ↓ denotes that the smaller value the better performance, and ↑ denotes that the bigger value the better performance.

**Generalization ability on other datasets.** To demonstrate the generalization ability of the proposed iToF imaging method, we perform a cross-dataset evaluation by training our network on the NYU-V2 dataset [6] and evaluating it on the other depth datasets, i.e. SUN RGB-D dataset [7] and 4D Light Field dataset [5] without any finetuning. We select 14 scenes of the 4D

Light Field Dataset and 200 scenes of the SUN RGB-D dataset as the test datasets. As shown in Tab. 2, our method shows good generalization capability on the two datasets and still achieves the best depth reconstruction fidelity. The qualitative depth reconstruction results are shown in Figs. 1-6. Compared with the other methods, our method can reconstruct the depth details of the scenes and maintain a high depth accuracy even at high noise levels on the two test datasets.

| (a) SUN RGB-D | MAE (mm)/RMSE (mm) | | |
|---|---|---|---|
| Sinusoid + PS [4] | 198.30/330.28 | 264.28/417.36 | 335.20/502.36 |
| Square + PS [4] | 123.86/212.12 | 173.63/293.48 | 232.11/376.42 |
| Hamilton [3] | 111.95/269.03 | 167.77/357.11 | 230.85/438.57 |
| DeepToF [8] | 47.10/78.70 | 65.51/98.10 | 103.40/141.56 |
| Ours | 14.11/27.36 | 15.09/28.44 | 20.39/34.55 |
| (b) 4D Light Field | MAE (mm)/RMSE (mm) | | |
| Sinusoid + PS [4] | 251.41/422.65 | 320.19/505.67 | 389.86/581.65 |
| Square + PS [4] | 161.45/287.52 | 221.26/380.14 | 284.72/463.68 |
| Hamilton [3] | 161.30/364.15 | 225.74/453.58 | 290.72/528.47 |
| DeepToF [8] | 50.27/85.42 | 68.94/102.45 | 100.94/133.56 |
| Ours | 16.84/37.20 | 17.03/36.91 | 21.66/41.06 |

Table 2. Depth reconstruction performance of different iToF methods on other two datasets [5, 7] with respect to three different noise settings from the 2nd to 4th column.

## References


[1] Ilya Chugunov, Seung-Hwan Baek, Qiang Fu, Wolfgang Heidrich, and Felix Heide. Mask-tof: Learning microlens masks for flying pixel correction in time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9116–9126, 2021. 1

[2] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Proceedings of the International Conference on Neural Information Processing Systems*, page 2366–2374, 2014. 1

[3] Felipe Gutierrez-Barragan, Syed Azer Reza, Andreas Velten, and Mohit Gupta. Practical coding function design for time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1566–1574, 2019. 1, 2

[4] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Horaud. Time of flight cameras: Principles, methods, and applications, 2012. 1, 2

[5] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Proceedings of the Asian Conference on Computer Vision*, pages 19–34. Springer, 2016. 1, 2, 3, 4

[6] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Proceedings of the European Conference on Computer Vision*. Springer, 2012. 1

[7] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 567–576, 2015. 1, 2, 4, 5

[8] Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6383–6392, 2018. 1, 2

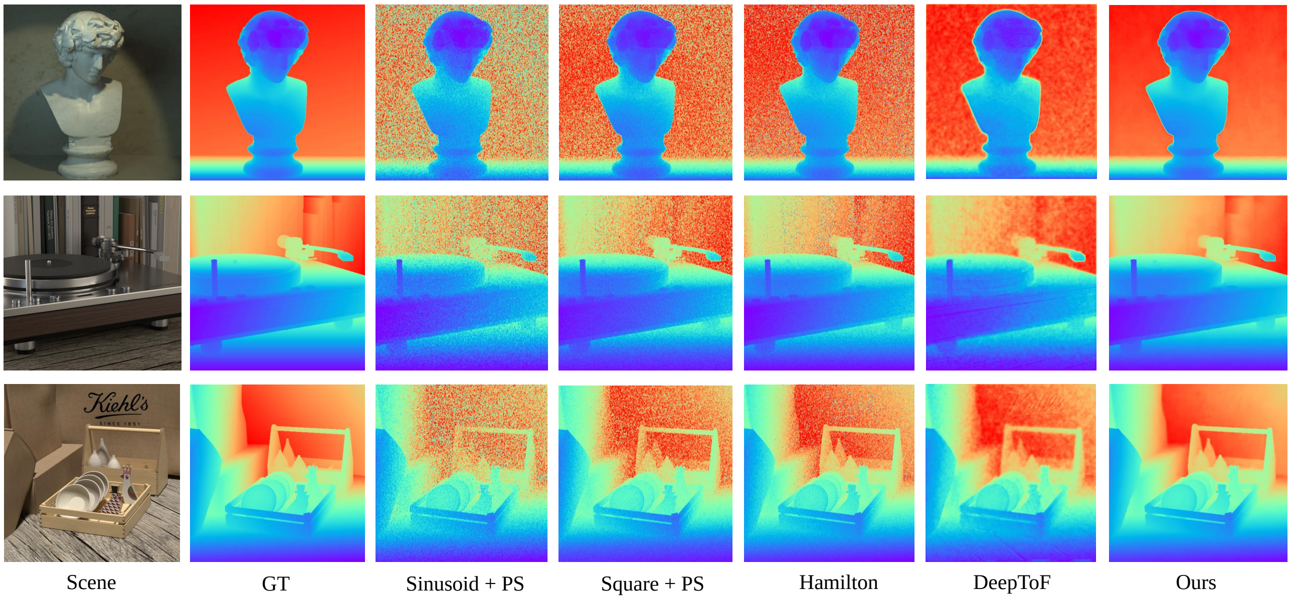


Figure 1. Depth reconstruction results on 4D Light Field Dataset [5] under small noise level, i.e. $(E, \beta) = (20000, 6000)$.

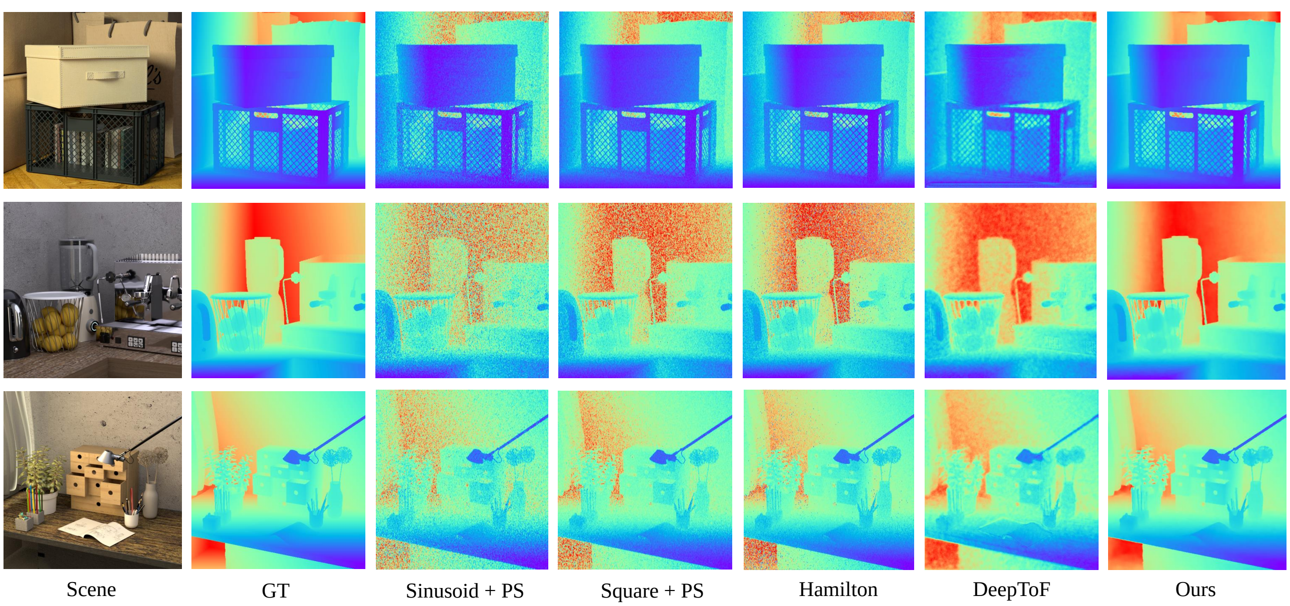


Figure 2. Depth reconstruction results on 4D Light Field Dataset [5] under middle noise level, i.e. $(E, \beta) = (14000, 6000)$.

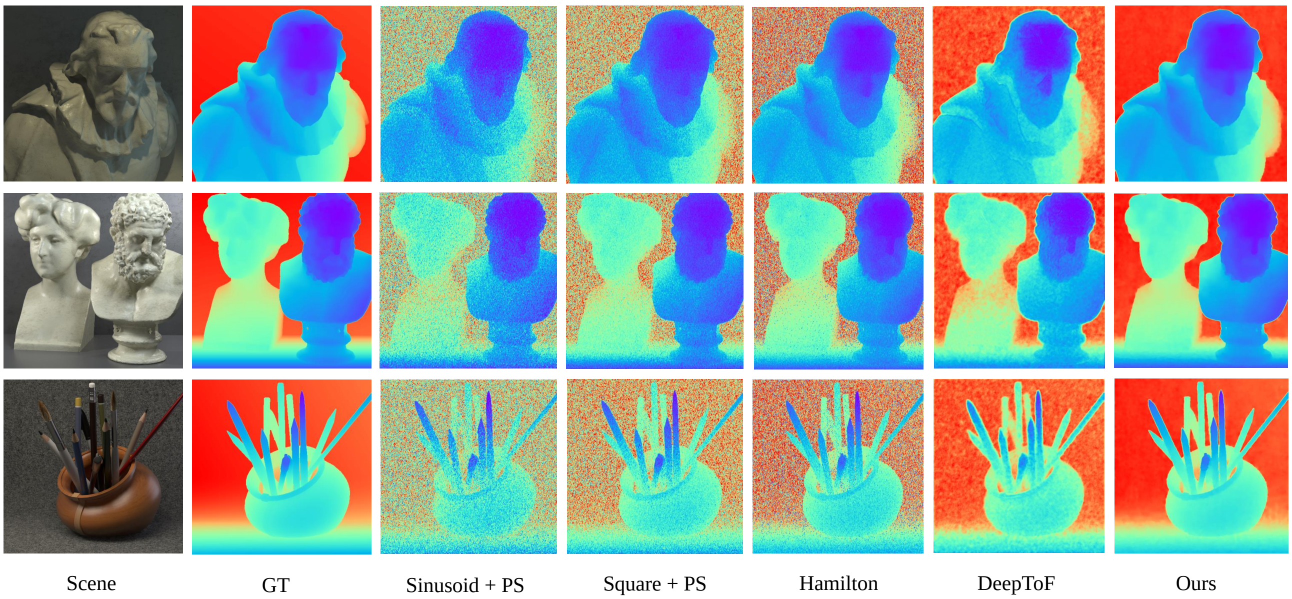


Figure 3. Depth reconstruction results on 4D Light Field Dataset [5] under large noise level, i.e. $(E, \beta) = (10000, 6000)$.

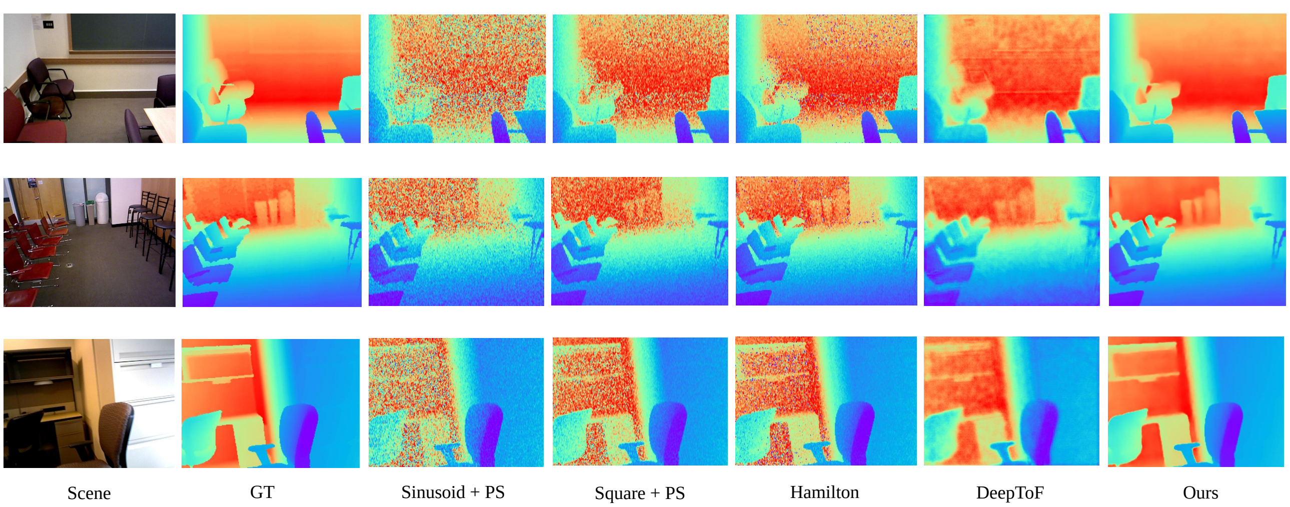


Figure 4. Depth reconstruction results on SUN RGB-D dataset [7] under small noise level, i.e. $(E, \beta) = (20000, 6000)$.

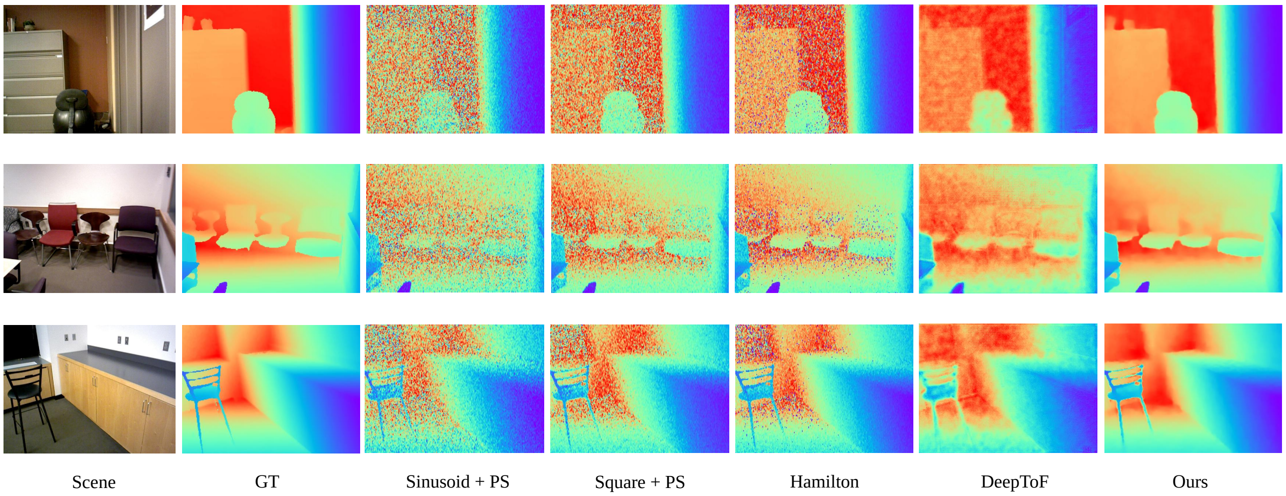


Figure 5. Depth reconstruction results on SUN RGB-D dataset [7] under middle noise level, i.e. $(E, \beta) = (14000, 6000)$.

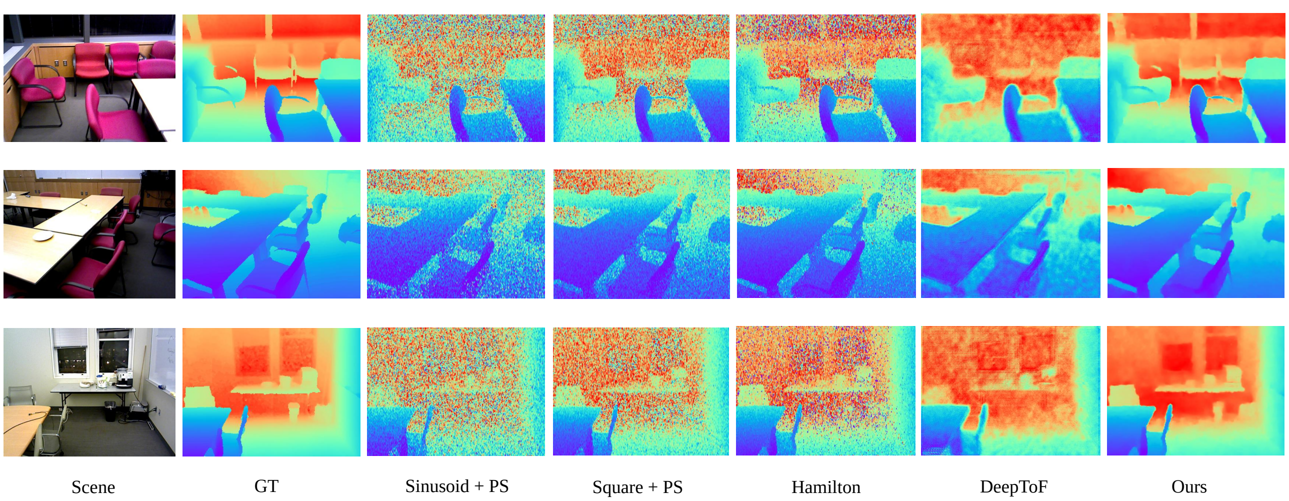


Figure 6. Depth reconstruction results on SUN RGB-D dataset [7] under large noise level, i.e. $(E, \beta) = (10000, 6000)$.