Snapshot Lidar: Fourier embedding of amplitude and phase for single-image depth reconstruction

Supplementary Material

Sarah Friday[1]*, Yunzi Shi[1]*, Yaswanth Cherivirala, Vishwanath Saragadam[2], Adithya Pediredla[1]
[1]Dartmouth College, [2]University of California, Riverside

Along with this supplementary pdf, we share a short video summarizing the paper's contributions and code and datasets for generating all the results.

Contents

1. Derivation for phase noise

In the main manuscript, we claimed that the variance of the phase noise for the snapshot technique is

$$\sigma_\phi^2 \approx \frac{\partial \phi}{\partial m_{kx}} m_{kx},$$

where ϕ is explicitly expressed as

$$\phi(x,y) = \arctan\left(\frac{m_{kx}(x,y)\sin(kx) \circledast \mathrm{sinc}(kx)}{m_{kx}(x,y)\cos(kx) \circledast \mathrm{sinc}(kx)}\right). \tag{1}$$

The variance of the measured image can be calibrated using standard noise calibration techniques [1, 4]. We show the proof of the above expression next.

Let $\mathcal{M}_{kx}(\omega_x,\omega_y) = \mathcal{F}(m_{kx}(x,y))$ be the Fourier transform of the snapshot measurement m_{kx}. Fourier transform after filtering the twin and shifting the twin-filtered image is given by

$$\hat{\mathcal{I}}(\omega_x,\omega_y) = \mathcal{M}_{kx}(\omega_x + k, \omega_y) \cdot B(|\omega_x| \leq k, \omega_y), \tag{2}$$

*The two authors contributed equally to this paper.

where B is a bandpass filter. The estimated phase of the scene $\phi(x, y)$ is the phase of the inverse Fourier transform of Eq. (2). Hence,

$$\begin{aligned}\arg \mathcal{F}^{-1}(\hat{\mathcal{I}}) &= \arg\left(\mathcal{M}_{kx}(\omega_x + k, \omega_y) \cdot B(|\omega_x| \leq k, \omega_y)\right), \\ \hat{\phi} &= \arg\left(m_{kx}(x, y)e^{jkx} \circledast k\,\mathrm{sinc}(kx)\right), \\ &= \arg\left(m_{kx}(x, y)(\cos kx + j \sin kx) \circledast k\,\mathrm{sinc}(kx)\right),\end{aligned}$$

which is same as Eq. (1). The amplitude noise can also be estimated similar as

$$\sigma_A^2 \approx \frac{\partial A}{\partial m_{kx}} m_{kx},$$

where A is explicitly expressed as

$$A(x, y) = \left(m_{kx}(x, y) \sin(kx) \circledast \mathrm{sinc}(kx)\right)^2 + \left(m_{kx}(x, y) \cos(kx) \circledast \mathrm{sinc}(kx)\right)^2. \tag{3}$$

However, time-of-flight cameras are often not used for estimating intensity images; instead, an inexpensive and high-resolution camera is often collocated with the ToF camera to measure intensity images.

2. Hardware prototype

In Supplementary Fig. 1, we show the hardware setup from several angles. We use the Melexis 75027 ToF camera, having replaced the manufacturer standard wide-angle lens with a 16mm Edmund optics lens. Then we scan one row at a time using the hardware region-of-interest (ROI) support. We steer this row with the help of the galvo system and phase shift the exposure code of the row linearly to capture the snapshot image in a single exposure. Using a NI-DAQ USB6363, we synchronize the camera and galvos on the external trigger generated by the Melexis camera. By capturing all the rows within one exposure duration, we get a snapshot capture. Unfortunately, current Melexis ToF cameras have a hardware lock on how many rows can be read in a second (100 rows per second) and hence, to capture a full frame (480 rows), we require around five seconds. Once Melexis removes this lock in future versions, we can capture snapshot images at 100 fps.

We found it more convenient to emulate the snapshot imaging technique by capturing multiple phase measurements and creating a composite image that emulates the rolling shutter effect. We show the emulation technique in Supplementary Fig. 2. Additionally, we must crop the composite image to ensure it contains an integer multiple of R lines, or the average number of rows/columns required for θ to span $[0, 2\pi]$. We only considered the region after cropping in comparative analysis against N-bucket and conventional methods.

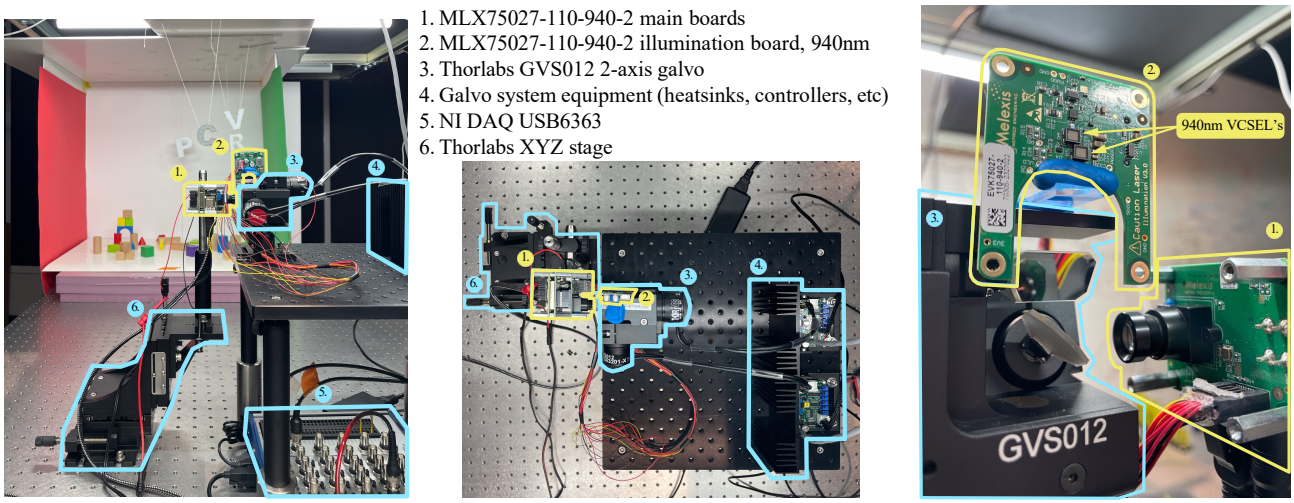
3. Robustness to local errors

As our reconstruction method works in the Fourier domain and the Fourier coefficients depend on all the pixels, it appears that local errors will affect overall reconstruction. However, this is not the case, as the reconstructed phase is still in the primal domain. To experimentally demonstrate the same, we have built a scene with specular and refractive objects that result in oversaturated pixels. In Supplementary Fig. 3, we show the reconstruction quality of a scene with specular and refractive objects. We could notice that our technique is robust to local saturation.

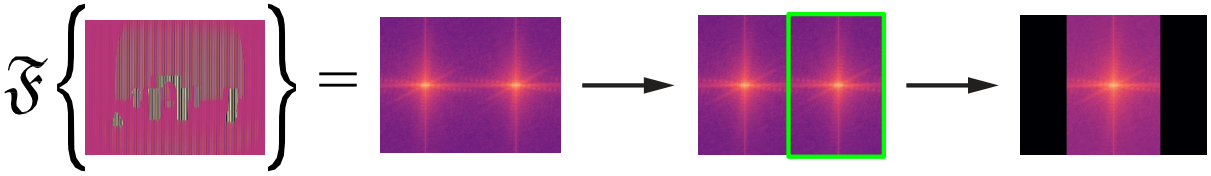
4. Comparison with epc660 camera

The epc660 camera and development software by ESPROS can reconstruct the phase using one or two measurements apart from the standard quadrature technique. Specifically, the epc660 CWToF sensor has a dual phase mode in which each row alternates as 0 phase shift and $\pi/2$ phase shift [3]. The sensor then combines the pairs of rows to create a single depth row, thus calculating the phase in a single capture. In the development software that comes with the epc660, "Dual MGX Mode" enables a feature that calculates the phase with 2 frames captured using the epc660's dual phase mode [2]. The first frame's rows alternate between the $\pi/2$ and π phase shifts, while the second frame's rows alternate between the 0 and $3\pi/2$ phase shifts.

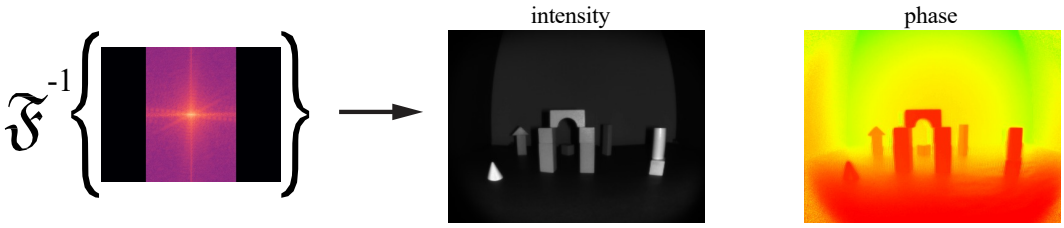
In Supplementary Fig. 4, we compare our method's phase reconstruction to the epc660 dual phase mode and Dual MGX mode methods. We create emulated measurements for both methods by stitching rows from quad images with the appropriate phases. To get full vertical resolution from the epc660 and Dual MGX methods, we iterate through all of the rows in the image, grouping with the row above to calculate phase.



a. Hardware setup with Melexis ToF camera with one row ROI and scanning with Galvos



b. Filtering ToF twin and frequency shifting hologram



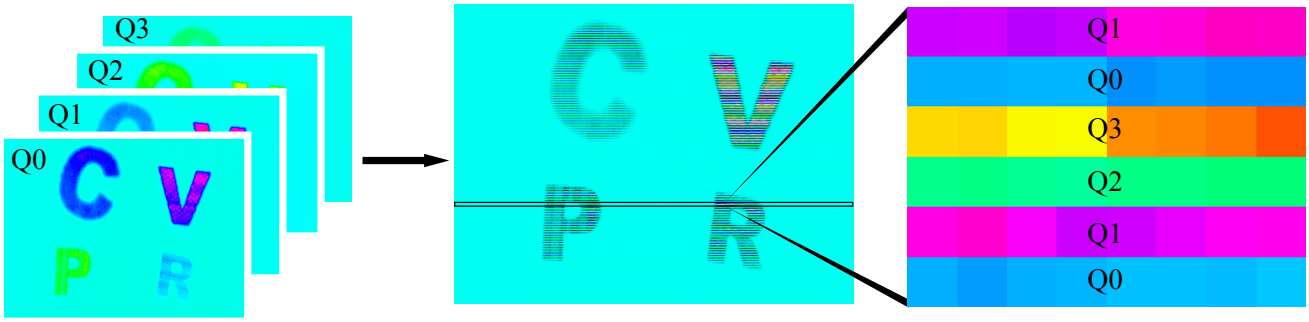
c. Reconstruction

Supplementary Figure 1. Snapshot hardware setup: (a) We use a Melexis time-of-flight camera and use hardware ROI support to scan only one row at a time. We scan other rows by steering the imaging beam using a galvo system. For every scanline, the phase shift (θ) of the camera is changed linearly. We achieve the synchronization between galvo mirrors and the camera with the help of NI-DAQ USB6363. Theoretically, the snapshot hardware method can capture all the rows within one frame capture duration. However, the Melexis device has a frame lock on the firmware, preventing us from capturing the full-frame. (b) The Fourier transform of the captured snapshot image contains both the ToF hologram and its twin. We filter the twin and frequency shift the ToF hologram. (c) The amplitude and phase of the inverse Fourier transform of the resultant ToF hologram gives the intensity and depth of the scene.

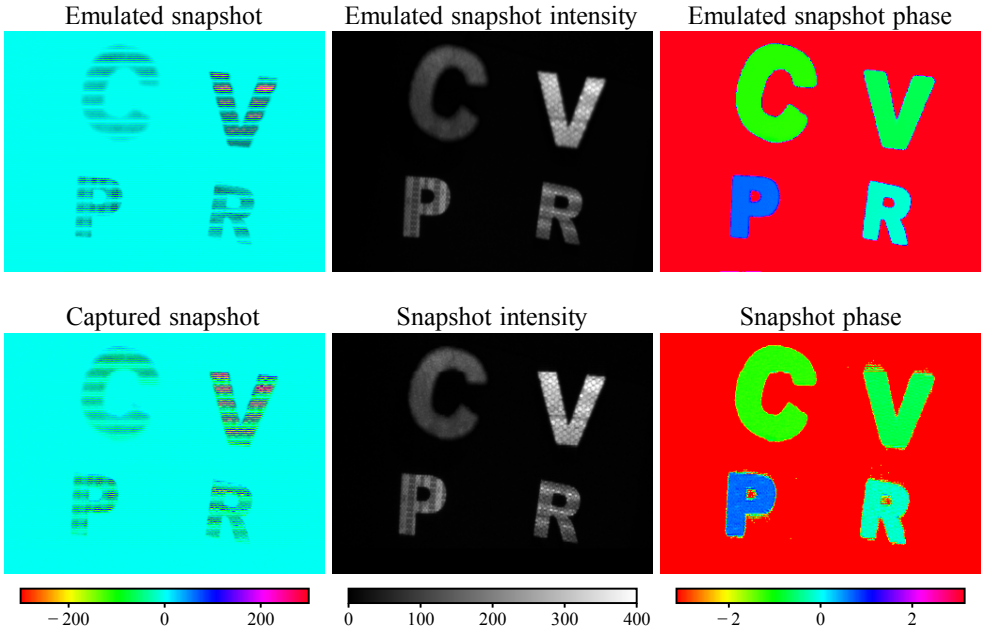
While the epc660 calculates phase in either single or two captures, the reconstruction error is consistently higher than our method, even with prefiltering. This trend is similar to how N-bucket reconstructions performed poorly compared to the Fourier reconstruction method. Note that, for both dual phase and dual MGX modes, the Fourier reconstruction method is not applicable as the phase variation rate is not linear.

5. Practicality of the Hardware Implementation of Phase-Shift

In an on-chip implementation, snapshot can be implemented using a rolling shutter and changing the phase shift per line (of either illumination or sensor); or by using a global shutter with a different phase shift per line of the sensor. The epc660 camera's dual phase mode (see Supplementary Fig. 4) already implements two distinct phases for two lines. The best



a. Compositing quad images to create the emulated snapshot image



b. Comparing the emulated rolling shutter with the hardware implemented rolling shutter reconstructions

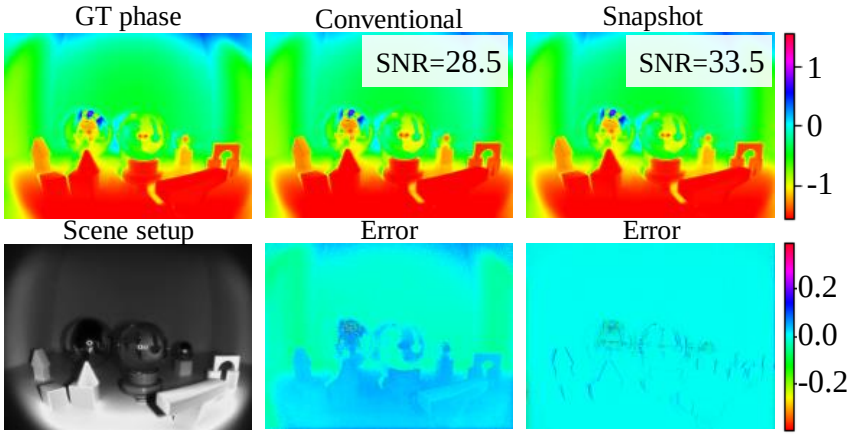
Supplementary Figure 2. Snapshot emulation. (a) We stitch a row/column from each phase measurement (quadrature in case $k = 2\pi/4$) to create a composite image that emulates the rolling shutter effect. While the snapshot hardware setup described in the previous figure can capture full-frame with the exposure duration, due to the firmware lock, it only captures 100 rows per second. Therefore, the emulation technique is faster and more convenient to show the proof-of-concept results and evaluate the proposed technique's advantages and disadvantages. (b) We show that the actual hardware setup and the emulation technique provide similar results.

snapshot implementation (i.e., $k = \pi/2$ or $R = 4$) requires only four phases for four lines; hence, an on-chip snapshot design is realizable.

The global shutter variant of our setup will not suffer from moving scene artifacts. The rolling shutter variant of our setup suffers from intraframe motion but is around 10dB better than both rolling and global shutter variants of the conventional technique, which suffer from interframe motion (See Supplementary Fig. 5).

The phase locked loops (PLLs) synchronizing the illumination and sensor are reliable at GHz range, and we are only operating in MHz. Hence, a precise linear phase shift is practical. Please note that our algorithm is robust to phase shift deviations (see Supplementary Fig. 6).

We acknowledge reviewer three (yNUC) for the insightful questions.



Supplementary Figure 3. Oversaturated pixels from specular and refractive objects do not affect the reconstruction quality of the neighboring pixels.

6. Additional figures

In this section, we provide extended visualizations for main document figures.

1. Supplementary Fig. 7 has additional visualizations of Fig. 5 in the main document.
2. Supplementary Fig. 8 and Fig. 9 have additional visualizations of Fig. 7 in the main document.
3. Supplementary Fig. 10 has additional visualizations of Fig. 8 in the main document.
4. Supplementary Fig. 11 has additional visualizations of Fig. 11 in the main document.

7. Code and Notebooks

We provide the code and notebooks for reproducing our method. The repository is hosted on GitHub.

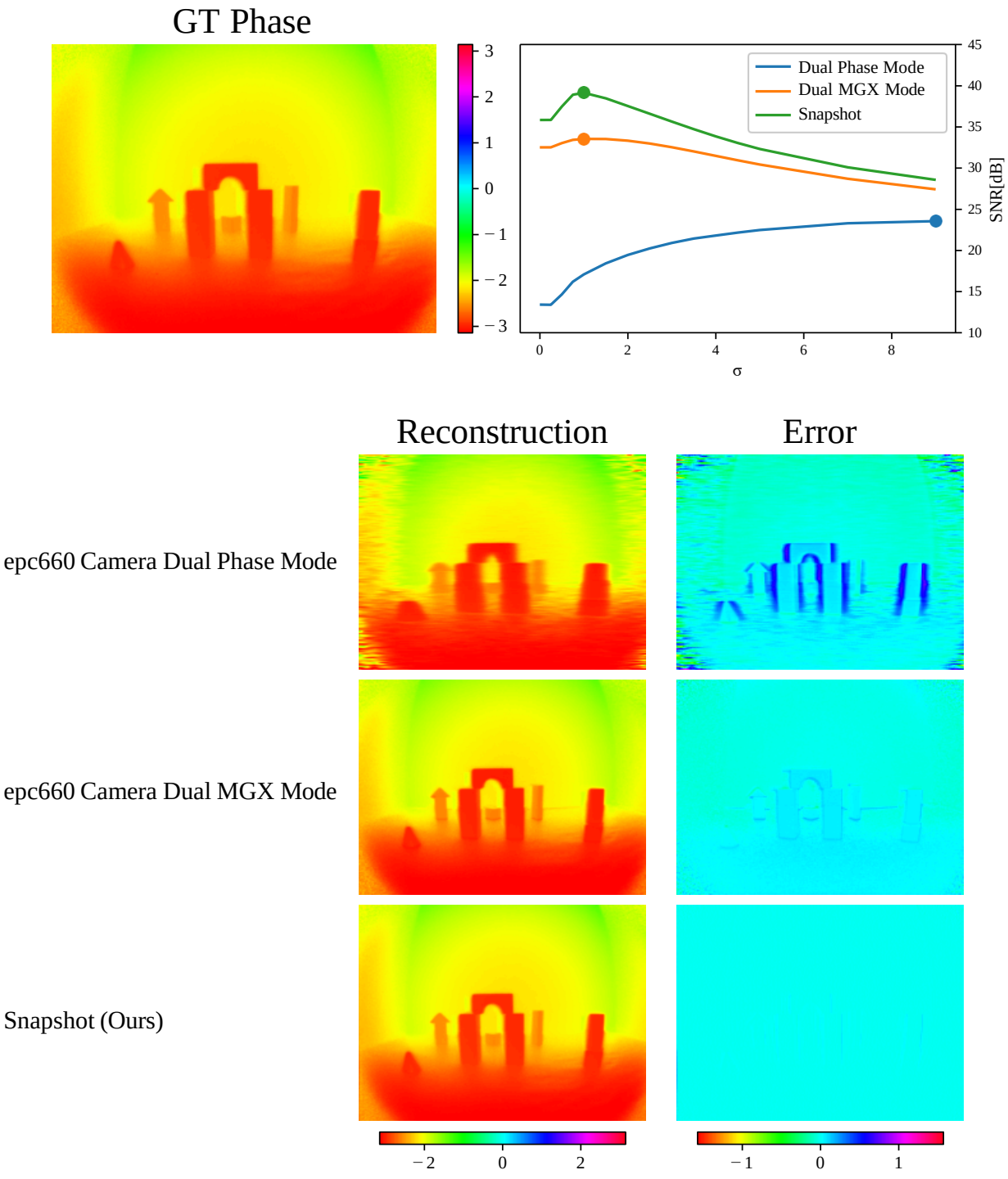
References

[1] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 643–652. 2023. 1

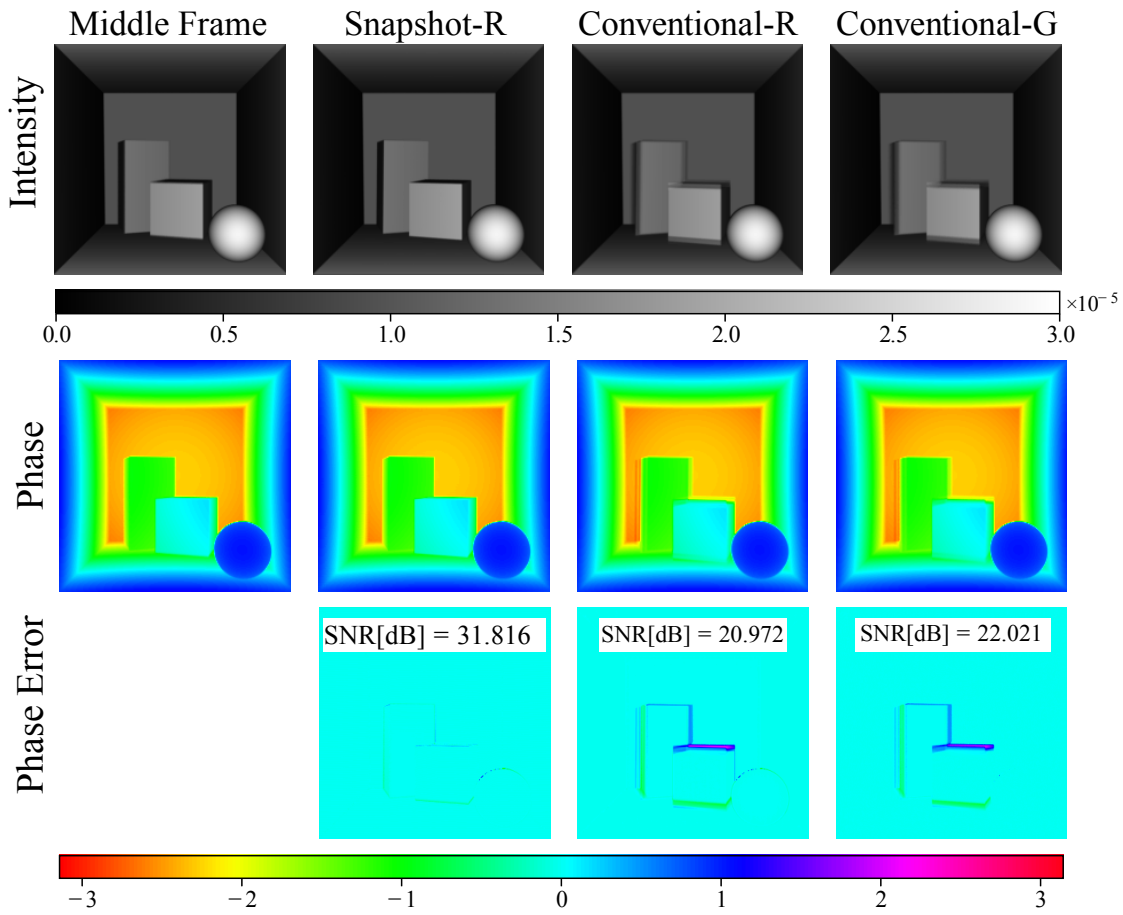
[2] ESPROS. Manual epc660 eval kit v1.09, 2018. 2

[3] ESPROS. epc660 datasheet v2.20, 2023. 2

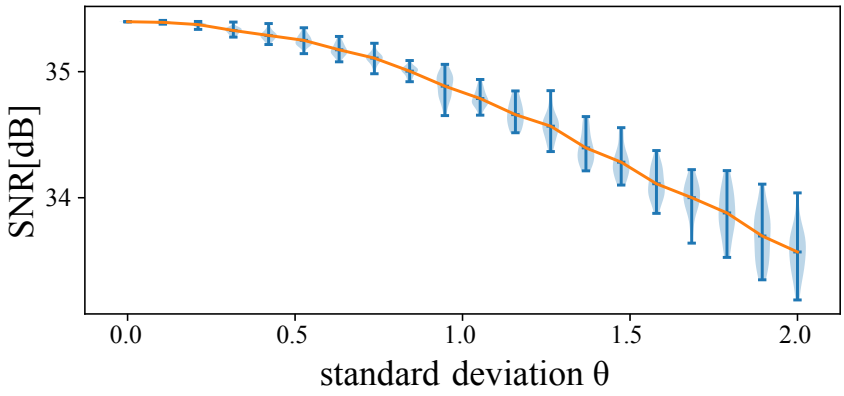
[4] Kristian Kirk and Hans Jørgen Andersen. Noise characterization of weighting schemes for combination of multiple exposures. In *BMVC*, pages 1129–1138. Citeseer, 2006. 1



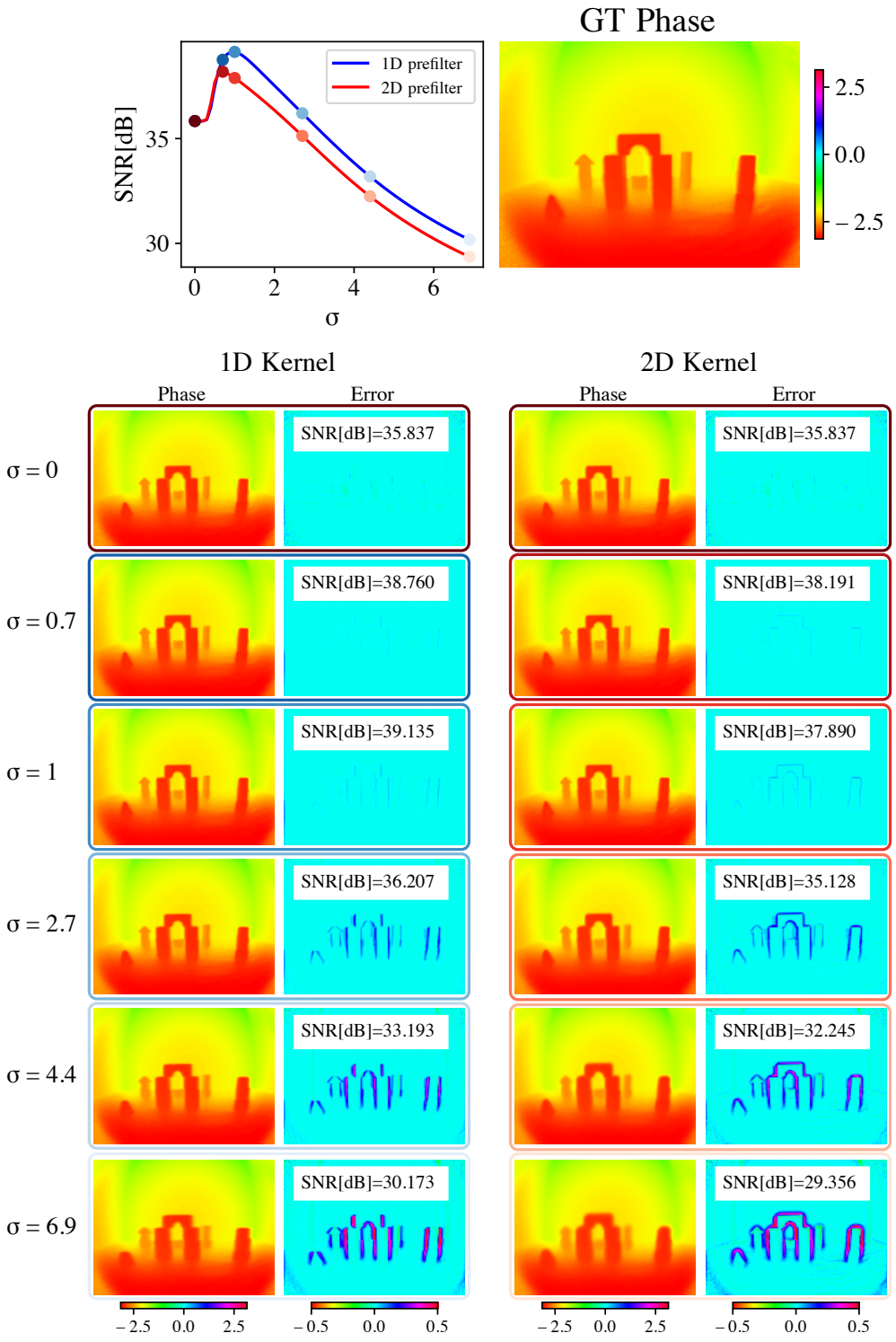
Supplementary Figure 4. Comparison between our method and ESPROS reconstruction methods at optimal σ's, plotted as points on the SNR graph. Our method consistently performs better than the ESPROS methods. The epc660 camera's dual phase mode calculates phase in a single frame with each row alternating between 0 and $\pi/2$ phase shifted signals. Dual MGX Mode requires 2 frames to calculate phase, making it still susceptible to motion artifacts but improving the reconstruction quality over a single dual phase mode frame. Both the epc660 dual phase mode and Dual MGX Mode techniques result in half vertical phase resolution.



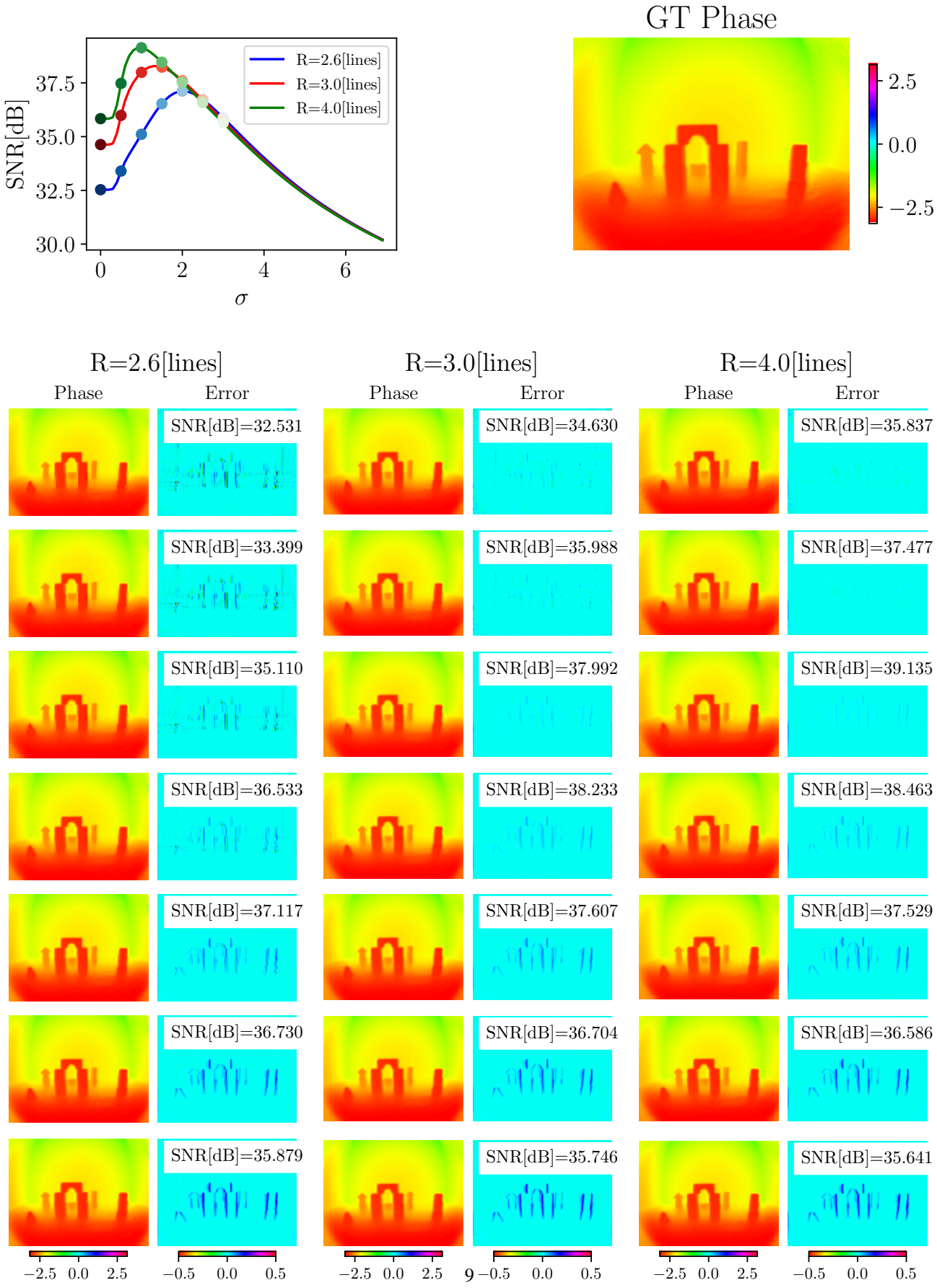
Supplementary Figure 5. Reconstructions of moving objects captured with a rolling shutter (-R) and global shutter (-G). The small box drops at 56 kmph, the large box drops at 16 kmph and moves from right to left at 56 kmph, while the sphere is still. The integration time for a full-sized phase frame is 500 μs.



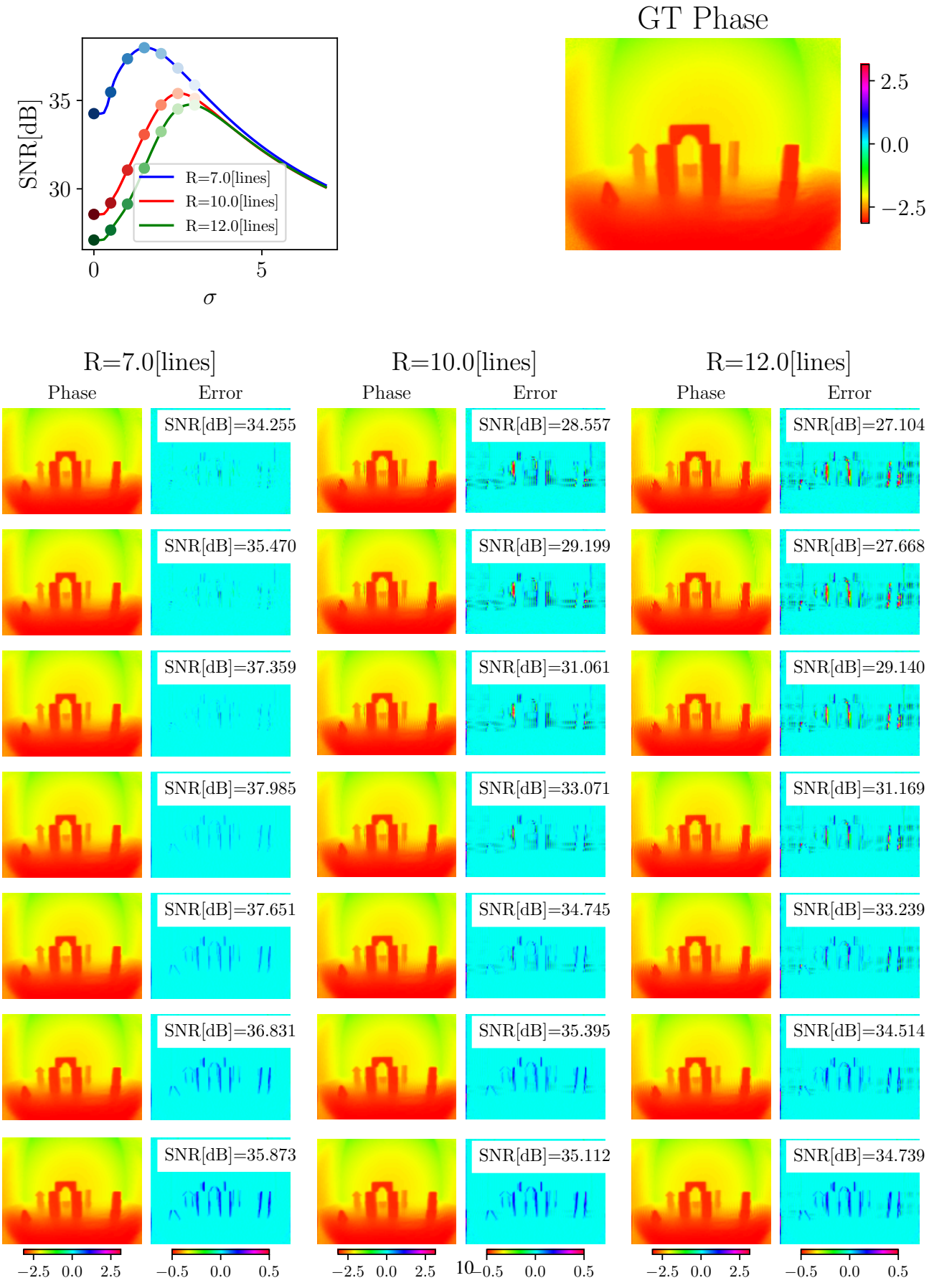
Supplementary Figure 6. PLLs are quite reliable and even if the phase shift has large random deviations, our technique is robust to them.



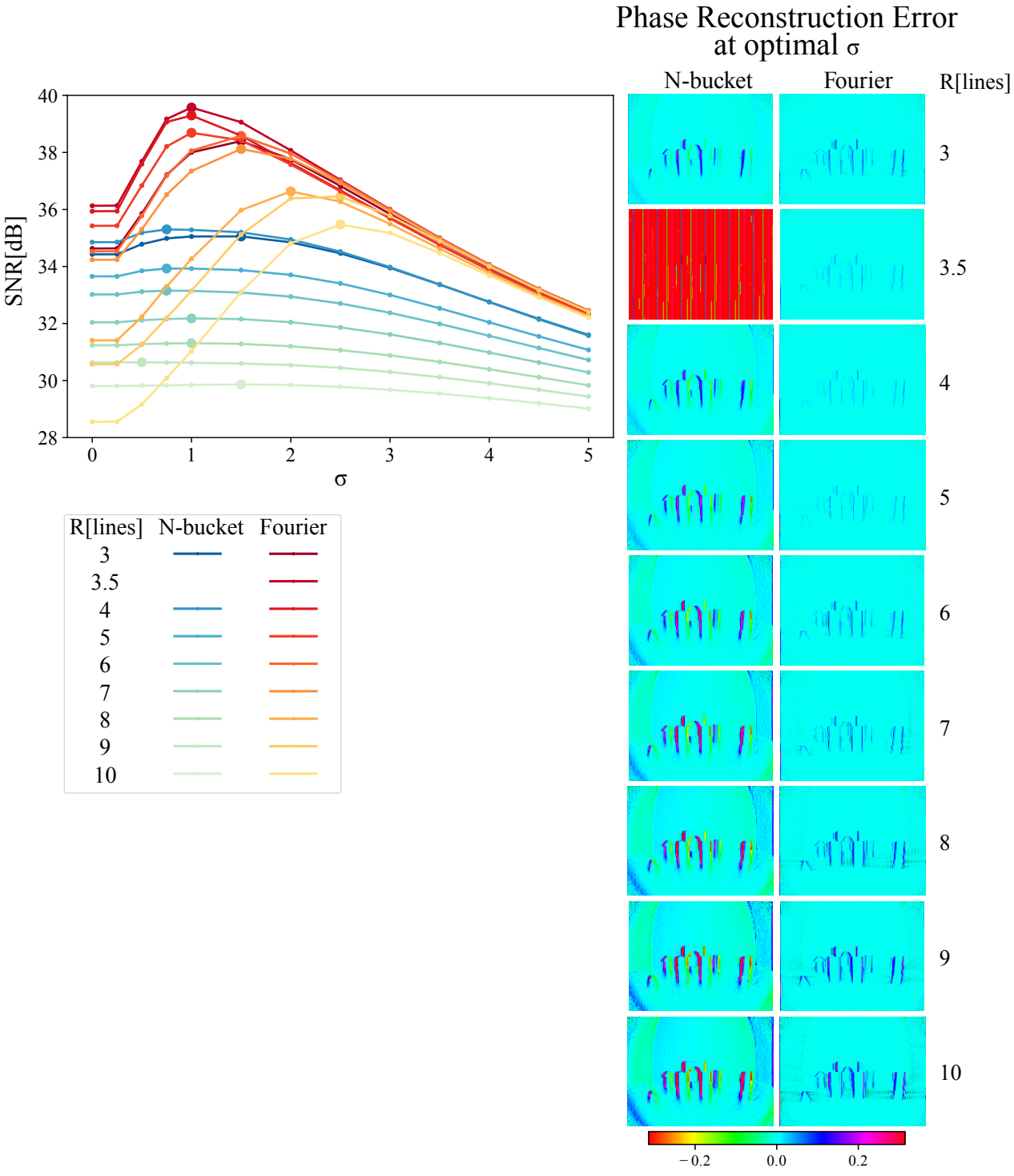
Supplementary Figure 7. SNR[db] as a function of varying prefilter kernel size σ for both a 1D and 2D Gaussian filter kernel on a snapshot phase reconstruction with $k = 90^{\circ}$. The 1D kernel consistently performs better than the 2D kernel as it preserves more edges. For the 1D kernel, the optimal σ_{1D} is 1 with an SNR[db] of 39.135. The 2D kernel's optimal σ_{2D} is 0.7, with an SNR[db] of 38.191.



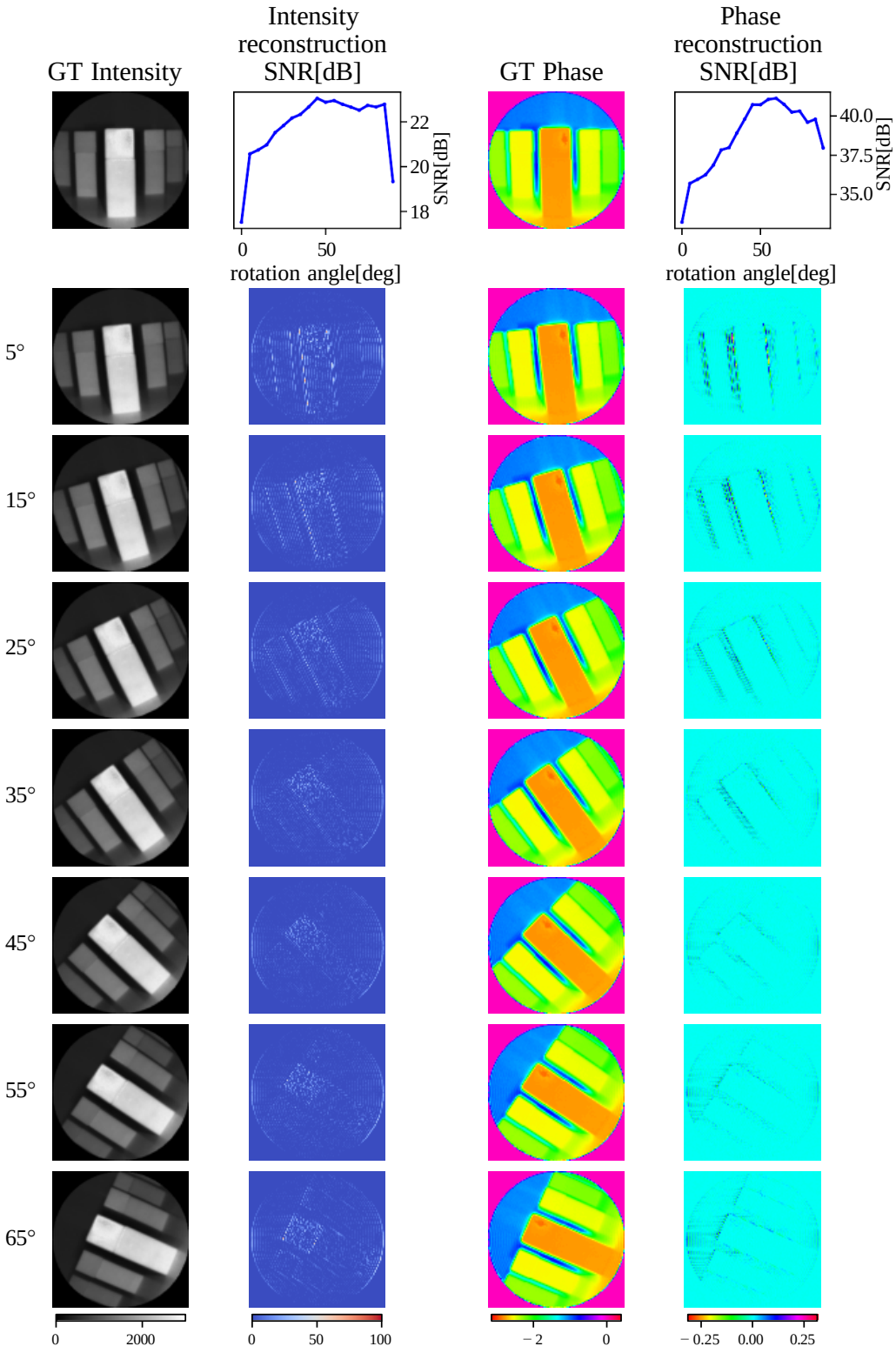
Supplementary Figure 8. Phase reconstruction quality for a given scene depends on R and σ. The overall best quality occurs at around $R = 4.0$ as the overlap between the hologram and twin is minimal at this value. Our method handles integer as well as fractional R.



Supplementary Figure 9. (Continued) More visualizations for the previous figure (Fig. 8).



Supplementary Figure 10. Comparison between the N-bucket and Fourier reconstruction techniques. The graph on the left shows SNR for various values of σ and R for both N-bucket and Fourier reconstruction techniques. The proposed Fourier reconstruction technique consistently outperforms the N-bucket technique for any R value. We visualize the phase reconstruction error on the right and highlight them as points on the left graph. Notice the N-bucket technique completely fails on non-integer values of R



Supplementary Figure 11. The effects of changing rolling shutter direction (obtained by rotating the camera) on intensity and phase reconstructions with no prefilter and $k = 90°$. When the edges are not aligned with the rolling shutter direction, the reconstruction suffers from less aliasing and noise folding artifacts.