Supplementary Materials for

First-Photon Imaging

Scene Depth and Reflectance Acquisition from One Detected Photon per Pixel

Ahmed Kirmani*, Dongeek Shin, Dheera Venkatraman, Franco N. C. Wong, and Vivek K Goyal*
Massachusetts Institute of Technology, Cambridge, MA 02139

*Corresponding authors. email: {akirmani | vgoyal}@mit.edu

Supplementary material for this paper includes:

A single PDF file named `1696-kirmani-sup.pdf` containing

- Proofs for the theory and algorithms presented in the main paper
- Additional figures detailing experimental setup and calibration
- Additional figures demonstrating the experimental results and comparisons with other denoising techniques
- Additional figures and plots for error analysis

Movie file named `1696-kirmani-sup-movie1.mpg`

Movie file named `1696-kirmani-sup-movie2.mpg`

Movie file named `1696-kirmani-sup-movie3.mpg`

List of Symbols

We first introduce symbols that are used frequently in this supplementary document.

1. $s(t)$: illumination waveform function
2. $r_{ij}(t)$: returning waveform function at pixel (i, j)
3. Δ : sampling time of single photon detector
4. T_r : pulse repetition interval
5. T_p : root mean square (RMS) pulse duration
6. Z_{max} : maximum scene depth
7. Z_{ij} : depth value at pixel (i, j)
8. α_{ij} : reflectance value at pixel (i, j)
9. η : detector quantum efficiency
10. b_λ : background light flux at operating optical wavelength λ
11. d : detector dark count rate
12. c : speed of light
13. S : total photon count contained in incident signal, $S = \int_0^{T_r} \eta s(t)\,\mathrm{d}t$.
14. B : total photon count contained in incident background, $B = \int_0^{T_r} \eta b_\lambda + d$.

Proofs for the theory and algorithms presented in main paper

Derivation of signal photon time-of-arrival likelihood, $f_{t_{ij}|\text{signal}}(\tau)$: As discussed in the paper, $\lambda_{ij}(t) = \eta\,\alpha_{ij} s(t - 2Z_{ij}/c) + (\eta b_\lambda + d)$ is the rate function of time-inhomogeneous Poisson process observed by the single photon avalanche diode (SPAD) at pixel (i, j). This Poisson processes observed at the detector is a merged stochastic process, containing signal and background components. Therefore, to derive the arrival time statistics of only signal photons, we set the background and detector dark count noise components to zero, i.e., $b_\lambda = d = 0$.

The time-correlated single-photon counting detection is capable of precisely timing the single photon arrivals within an accuracy interval of, Δ seconds, starting at time instant, τ, where, $\tau \in [0, T_r]$. Typically, Δ measures a few picoseconds and is much smaller than pulse duration, T_p, and pulse repetition interval, T_r. Using time-inhomogenous Poisson photon counting statistics, we obtain the following statistics for first signal photon detection's arrival time,

$$\begin{aligned}\Pr[\text{detecting first signal photon during } t \in (\tau, \tau+\Delta)] &= \Pr[\text{no signal photon detection in } t \in (0, \tau)] \\ &\quad \times \Pr[\text{one or more signal detections in } t \in (\tau, \tau+\Delta)] \\ &= \exp\left[-\alpha_{ij} \int_0^{\tau} s(t - 2Z_{ij}/c)\,\mathrm{d}t\right] - \exp\left[-\alpha_{ij} \int_0^{\tau+\Delta} s(t - 2Z_{ij}/c)\,\mathrm{d}t\right]\end{aligned}$$

Because Δ is relatively very small, it is a good approximation of an infinitesimal time-interval. Using this fact, we effectively treat the photon arrival time at pixel (i, j) as a continuous random variable with probability density function (pdf), $f_{t_{ij}|\text{signal}}(\tau)$. To derive this pdf we note that

$$f_{t_{ij}|\text{signal}}(\tau) = \lim_{\Delta \to 0} \frac{\Pr[\text{detecting first signal photon during } t \in (\tau, \tau+\Delta)]}{\Delta},$$

and arrive at the expression for pdf expression

$$f_{t_{ij}|\text{signal}}(\tau) \;\propto s(t - 2Z_{ij}/c)\; e^{-\alpha_{ij}\int_0^{\tau} s(t-2Z_{ij}/c)\,\mathrm{d}t} \approx\; s(t - 2Z_{ij}/c) \tag{1}$$

Effect of Low rate approximation: The latter approximation in Equation 2 is only valid at low-light levels, when the optical flux is low, i.e., $(S + B \ll 1)$.
After normalization, we obtain the final expression,

$$f_{t_{ij}|\text{signal}}(\tau) = s(t - 2Z_{ij}/c)\,\eta/S$$

Derivation of background noise photon time-of-arrival likelihood, $f_{t_{ij}|\text{background}}(\tau)$: As before, to derive the arrival time statistics of only background photons, we set the signal components to zero, i.e., $\alpha_{ij} = 0$. Then the incident flux at the detector is only due to background noise and detector dark counts, i.e., $\lambda_{ij}(t) = \eta b_\lambda + d$. This rate function is time-homogeneous as opposed to the Poisson process generated by signal photons. Using time-homogenous Poisson photon counting statistics, we obtain the following statistics for first background photon detection's arrival time,

$$\begin{aligned}\Pr[\text{detecting first background photon during } t \in (\tau, \tau+\Delta)] = \\ \Pr[\text{no signal photon detection in } t \in (0,\tau)] \\ \times \Pr[\text{one or more signal detections in } t \in (\tau, \tau+\Delta)] \\ = \exp\left[-(\eta b_\lambda + d)\,\tau\right] - \exp\left[-(\eta b_\lambda + d)\,(\tau+\Delta)\right]\end{aligned}$$

Again, because Δ is relatively very small, we effectively treat the photon arrival time at pixel (i,j) as a continuous random variable with probability density function (pdf), $f_{t_{ij}|\text{background}}(\tau)$. To derive this pdf we note that

$$f_{t_{ij}|\text{background}}(\tau) = \lim_{\Delta\to 0} \frac{\Pr[\text{detecting first background photon during } t \in (\tau, \tau+\Delta)]}{\Delta},$$

and arrive at the final expression for pdf expression

$$f_{t_{ij}|\text{signal}}(\tau) \;\propto (\eta b_\lambda + d)\;\; e^{-\int_0^\tau (\eta b_\lambda + d)\,\mathrm{d}t} \;\approx\; (\eta b_\lambda + d) \tag{2}$$

Effect of Low rate approximation: The latter approximation in Equation 2 is only valid at low-light levels, when the optical flux is low, i.e., $(S + B \ll 1)$.
After normalization, we obtain the final expression,

$$f_{t_{ij}|\text{background}}(\tau) = \frac{1}{T_r}$$

Derivation of probability $P_0(i,j)$**:** The total mean photon count at the detector is equal to $(\alpha_{ij}\, S + B)$. The observation time associated with a single pulsed illumination is equal to the pulse repetition periond, T_r. Denote with N the total number of photons measured by the detector in the time-interval, $[0, T_r]$. Then using Poisson photon counting statistics the

distribution of N is given as

$$\Pr[N = k] = \frac{e^{-(\alpha_{ij}\, S+B)}(\alpha_{ij}\, S + B)^k}{k!}$$

The probability, $P_0(i, j)$ of *not* detecting a photon in response to a single pulsed illumination is obtained as follows

$$P_0(i, j) = \Pr[N = 0] = e^{-(\alpha_{ij}\, S+B)}$$

Novel Image Formation: Formulations and Algorithms

Step 1: Reflectivity estimation: Let $\{n_{ij}\}$ be the dataset for the number of elapsed pulses until first photon detection obtained by raster scanning through all the pixels. Then, we can write the negative log-likelihood of observing the number of pulses until one detection as

$$\begin{aligned}\mathcal{L}_\alpha(\alpha_{ij}; n_{ij}) &= -\log\left(\left(e^{-(\alpha_{ij}S+B)}\right)^{n_{ij}-1}\left(1-e^{-(\alpha_{ij}S+B)}\right)\right)\\ &\equiv (n_{ij}-1)S\alpha_{ij} - \log\left(1-e^{-(\alpha_{ij}S+B)}\right),\end{aligned}$$

where $\equiv$ denotes equality up to a constant. We note that the strict concavity of $e^{-\alpha_{ij}S}$ over α_{ij} implies the strict concavity of $1-e^{-(\alpha_{ij}S+B)}$, and the strict convexity of $-\log\left(1-e^{-(\alpha_{ij}S+B)}\right)$. Thus, $\mathcal{L}_\alpha(\alpha_{ij}; n_{ij})$ is a strictly convex function of α_{ij}, since it is a sum of a convex and strictly convex function. Figure 1 shows how the negative log-likelihood function changes when the background illumination power B is varied.

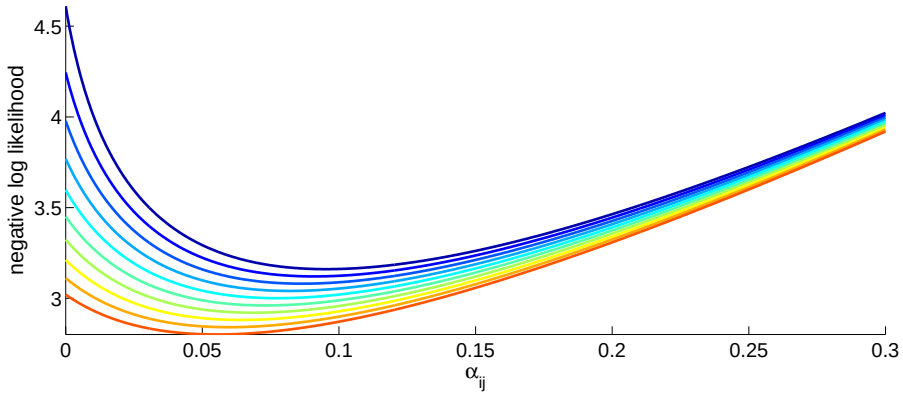


Figure 1: $\mathcal{L}_\alpha(\alpha_{ij}; n_{ij})$ vs. α_{ij} when $n_{ij} = 10$ and $S = 1$ for several values of B. B ranges from 0.01 (blue) to 0.05 (red). Note the global minimum of negative log-likelihood shifts as B changes.

Incorporating the prior knowledge that the reflectivity image of a natural scene is smooth, our regularized maximum likelihood estimate of scene reflectance is

$$\underset{\substack{\mathbf{A}=\{a_{ij}\}:\\ a_{ij}\geq 0}}{\arg\min} \sum_i\sum_j (n_{ij}-1)S\alpha_{ij} - \log\left(1-e^{-(\alpha_{ij}S+B)}\right) + \beta\|\Phi_\alpha \mathbf{A}\|_1,$$

where Φ_α is a sparsifying transform (e.g. discrete wavelet transform), $\|\cdot\|_1$ is the l_1-norm, and β is the variational parameter controls the degree of regularization.

Step 2: Background noise rejection: At pixel (i, j), the observed inhomogeneous Poisson process has rate $\lambda_{ij}(t) = \eta\alpha_{ij}s(t - 2Z_{ij}/c) + (\eta b_\lambda + d)$. Let "signal" and "background" be the events that the first detected photon is coming from pulse waveform (the first term in $\lambda_{ij}(t)$) and noise (the second term in $\lambda_{ij}(t)$), respectively. Because the photon detection can only come from either the pulse or background, we see that Pr[signal] $= \alpha_{ij}S/(\alpha_{ij}S + B)$, Pr[background] $= B/(\alpha_{ij}S + B)$. so that Pr[signal] + Pr[background] $= 1$. The likelihood function can thus be written as

$$\begin{aligned} p_{t_{ij}|\text{signal}}(\tau) &= (\Pr[\text{signal}] \times p_{t_{ij}|\text{signal}}(\tau)) + (\Pr[\text{background}] \times p_{t_{ij}|\text{background}}(\tau)) \\ &= \left(\frac{\alpha_{ij}S}{\alpha_{ij}S + B}\right) \frac{s(\tau - 2Z_{ij}/c)}{\int_0^{T_r} s(t - 2Z_{ij}/c)\, dt} + \left(\frac{B}{\alpha_{ij}S + B}\right) \frac{\mathbb{1}_{[0,T_r]}(\tau)}{B}, \end{aligned}$$

where $\mathbb{1}_A(x)$ is the indicator function of element x in set A. Thus, the time-of-arrival of the first detected photon has a probability density function that is a mixture of the pulse distribution and a high-variance uniform distribution modeling background light. Such signal noise model is known as the impulse noise model [1].

As shown Garnett et al., the rank-ordered absolute difference (ROAD) statistics can accurately determine which samples came from the high-variance uniform distribution when samples are drawn from a mixture distribution described above. Let $\mathcal{A}_{ij}$ be the set of the absolute differences between the time-of-arrival value t_{ij} and the time-of-arrival values at its eight neighboring pixels. Then, the ROAD statistic for the $(i, j)^{\text{th}}$ pixel of the depth map is given as

$$\text{ROAD}(i, j) = \min_{x_1, x_2, x_3, x_4 \in \mathcal{A}_{i,j}} (x_1 + x_2 + x_3 + x_4)\,.$$

Computing this ROAD statistic simplifies to sorting the absolute arrival time differences in ascending order, and computing the sum of the first four values. Finally, as discussed in the paper, we obtain the set of corrupted (censored) image pixels using a binary hypothesis test which uses the reflectivity estimates computed in Step 1, and the ROAD statistic.

[1] Garnett, Roman, Timothy Huegerich, Charles Chui, and Wenjie He. *A universal noise removal algorithm with an impulse detector.* Image Processing, IEEE Transactions on 14, no. 11 (2005): 1747-1754.

Step 3: Depth estimation: The negative log-likelihood of signal photon arrival times is

$$\mathcal{L}_z(Z_{ij}; t_{ij}) \equiv -\log s(t_{ij} - 2Z_{ij}/c).$$

So, if the illuminated waveform $s(t)$ is log-concave. then the regularized maximum likelihood estimation of depth is computed by solving a convex optimization problem. For example, choosing the pulse to be in the family of generalized Gaussians such that $s(t) \propto e^{-(|t|/a)^p}$, where $p > 1$ and $a > 0$, leads to a convex optimization problem for regularized maximum likelihood estimation. Figure 2 shows the negative log-likelihood functions of generalized Gaussian distributions, which are log-concave, and the resulting negative log-likelihood functions are convex.

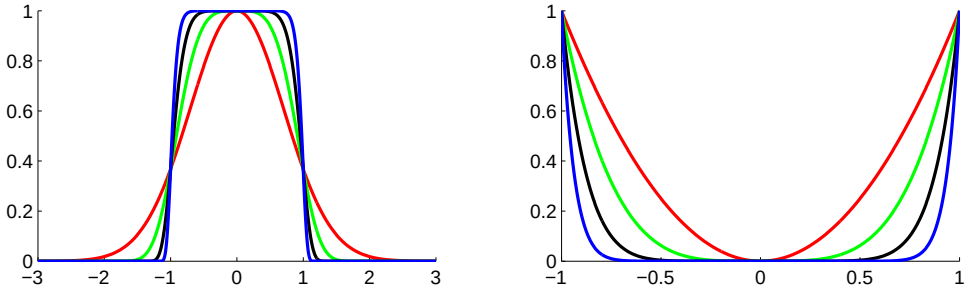


Figure 2: **Left:** Plot of generalized Gaussian functions with $p = 2$ (red), 3 (green), 4 (black), 5 (blue) with fixed $a = 1$ and amplitude 1. The generalized Gaussian function includes the Gaussian function ($p = 2$) and the square function (p large). **Right:** plot of negative log generalized Gaussian functions for the same p, a values.

Based on the prior that depth maps are smooth, the regularized MLE can be written as

$$\underset{\substack{\mathbf{D}=\{d_{ij}\} \\ 0 \le d_{ij} \le Z_{\max}}}{\arg\min} \sum_{\text{uncensored } (i,j)} -\log s(t_{ij} - 2d_{ij}/c) + \beta ||\Phi_Z \mathbf{D}||_1.$$

We note that even though we do not use the arrival times at the censored pixels, our regularized ML estimate the depth at the censored pixels by enforcing global sparsity.

Derivation of ML depth estimation error: At pixel (i, j), the pointwise maximum likelihood depth estimate based on the first photon arrival time is, $\hat{Z}_{ij} = c\,(t_{ij} - T_m)/2$, where $T_m = \arg\max_t s(t)$. In our experiments, we set the pulse to be mode-centered, hence $T_m = 0$. The root mean square error of the ML estimate as

$$\text{RMSE}(Z_{ij}, \hat{Z}_{ij}) = \sqrt{\mathbb{E}\left[(Z_{ij} - \hat{Z}_{ij})^2\right]}.$$

We earlier derived the statistics of t_{ij} in the cases when the detected photon originated due to background. Using these probability density functions, and the $\Pr$[detected photon is background noise], we derive

$$\text{RMSE}(Z_{ij}, \hat{Z}_{ij}) = \sqrt{\left(\frac{\alpha_{ij}S}{\alpha_{ij}S + B}\right)\frac{c^2}{4}T_p^2 + \left(\frac{B}{\alpha_{ij}S + B}\right)\frac{c^2}{4}\left(\left(\frac{2Z_{ij}}{c} - T_r\right)^2 + \frac{T_r^2}{12}\right)}.$$

As was in our experiments, the signal flux and the background noise flux were approximately equal, i.e., $S \approx B$ we assume that the probabilities of detecting a photon from either signal or background are equal. Therefore,

$$\text{RMSE}(Z_{ij}, \tilde{Z}_{ij}) \geq \frac{c}{2}\sqrt{\frac{1}{2}\left(T_p^2 + \frac{T_r^2}{12}\right)}.$$

Typically in range imaging applications, $T_r \gg T_p$. Hence, the root mean square error between true depth and the ML estimate based on the first photon observation is at least $cT_r/4\sqrt{6}$. Thus, even ML estimation requires a large number of detected photons at each pixel location to achieve sub-pulse width depth resolution under non-zero background illumination conditions,

Experimental Calibration

Measurement of pulse shape, $s(t)$: For depth estimation, our computational imager requires knowledge of the pulse shape, $s(t)$. This was obtained by directly illuminating the detector with highly attenuated laser pulses, and binning the photon arrival times to generate a photon counting histogram. Then, we obtain $s(t)$ by least square fitting the histogram with a Gaussian mixture function of order 3.

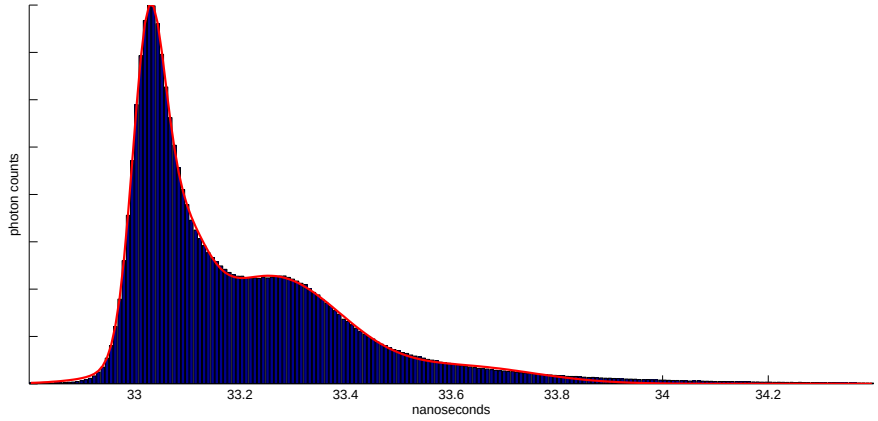


Figure 3: Histogram of photon arrival times (blue) and its Gaussian mixture function fit (red).

Radiometric calibration: The detection efficiency is the product of the interference filter's transmission and the detector's quantum efficiency, $\gamma = 0.49 \times 0.35 = 0.17$. A reference calibration for S, the average photon number in the backreflected signal received from a single laser pulse, was obtained as follows. All sources of background light were turned off, and the laser was used to illuminated a highly-reflective Lambertian surface at a distance of 2 m. The average number of transmitted pulses before a photon detection was found to be $\langle n_{calibration} \rangle = 65$. Using Equation 1 from the paper, with $\alpha_{calibration} = 1$ and $B = 0$, we find

$$\langle n_{calibration} \rangle = \frac{1}{1 - P_0(\text{calibration})} = \frac{1}{1 - \exp(-S)},$$

from which $\langle n_{calibration} \rangle = 65$ and $\eta = 0.17$ give $S = 0.09$.

For adjusting background illumination power, the laser was first turned off and all objects were removed from the scene. Then the incandescent lamp's optical power was adjusted such that the average number of background photons reaching the detector in a pulse repetition period was $B = 0.1$.

Additional figures detailing experimental setup and comparison with denoising methods

Ground truth datasets were computed by reducing background noise to a negligible level and using pointwise ML estimation with a large number of photons ($M \approx 1000$ photons-per-pixel).

For all processing methods used, the parameters were choses to minimize RMSE for depth maps and PSNR for reflectivity reconstruction.

Schematic of experimental setup

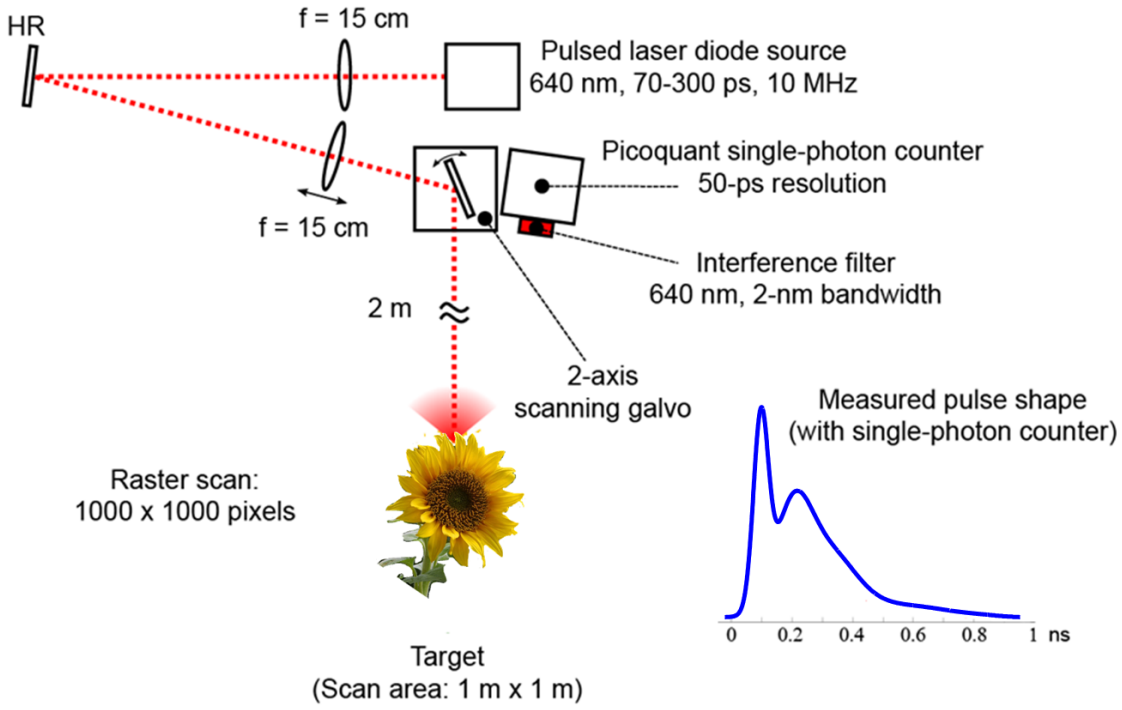


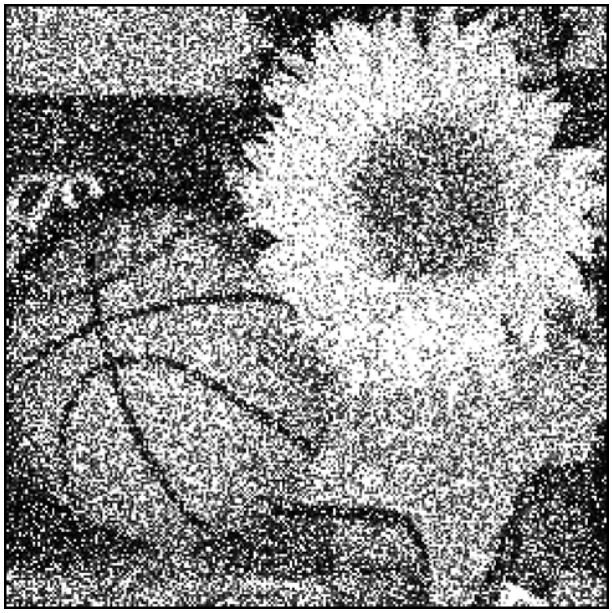
Figure 4: Schematic of the experimental setup showing the optical paths and pulse shape measured

Scene photograph of layered scene dataset

Figure 5: Photograph for layered scene dataset

Reflectivity images for layered scene dataset

(a) Ground truth.

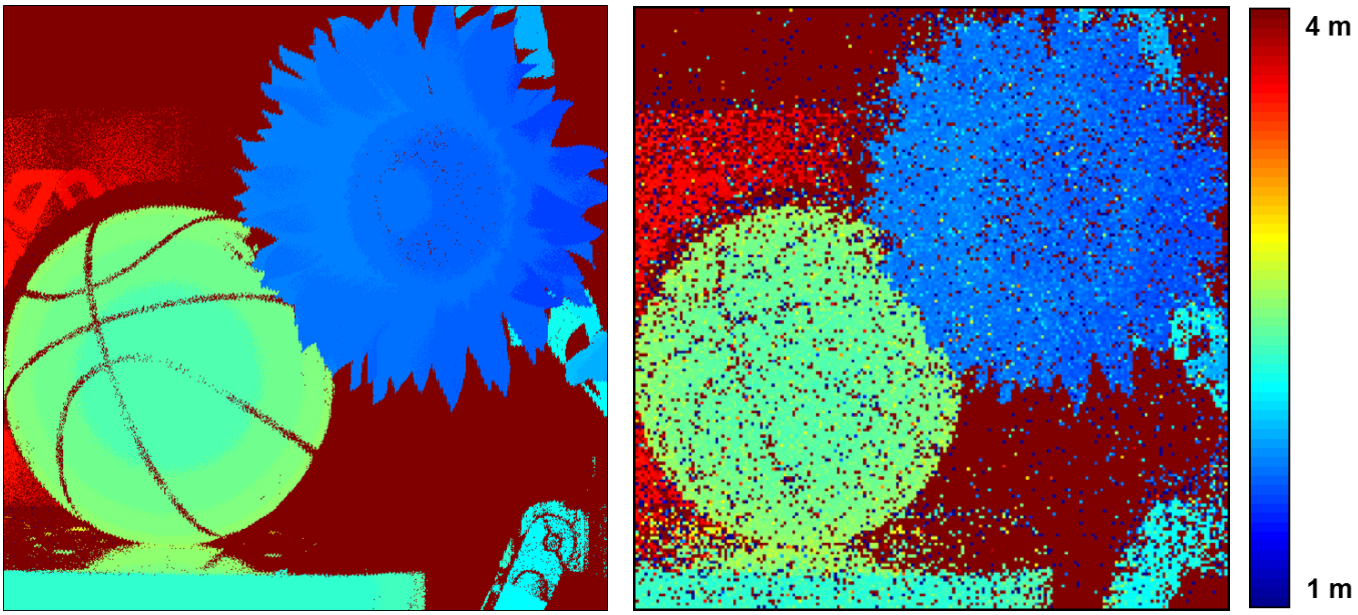


Pointwise ML estimate.

(c) Our processing: first-photon imaging.

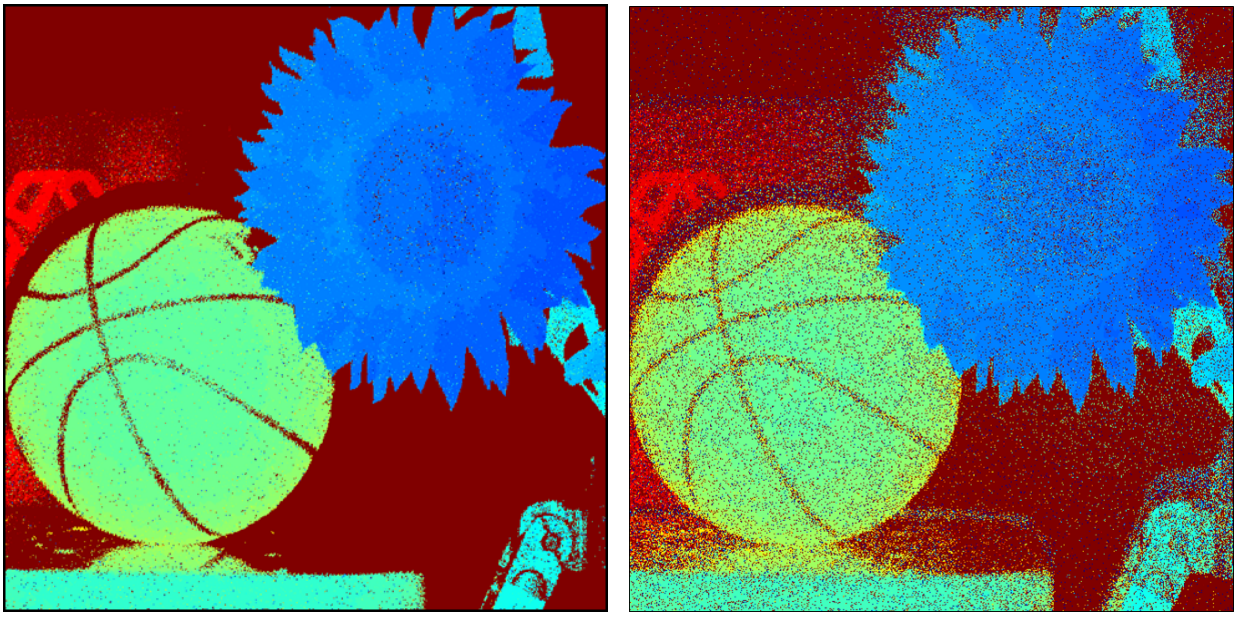
(d) BM3D with Anscombe transformation.

Depth maps for layered scene dataset



(a) Ground truth.

Pointwise ML estimate.

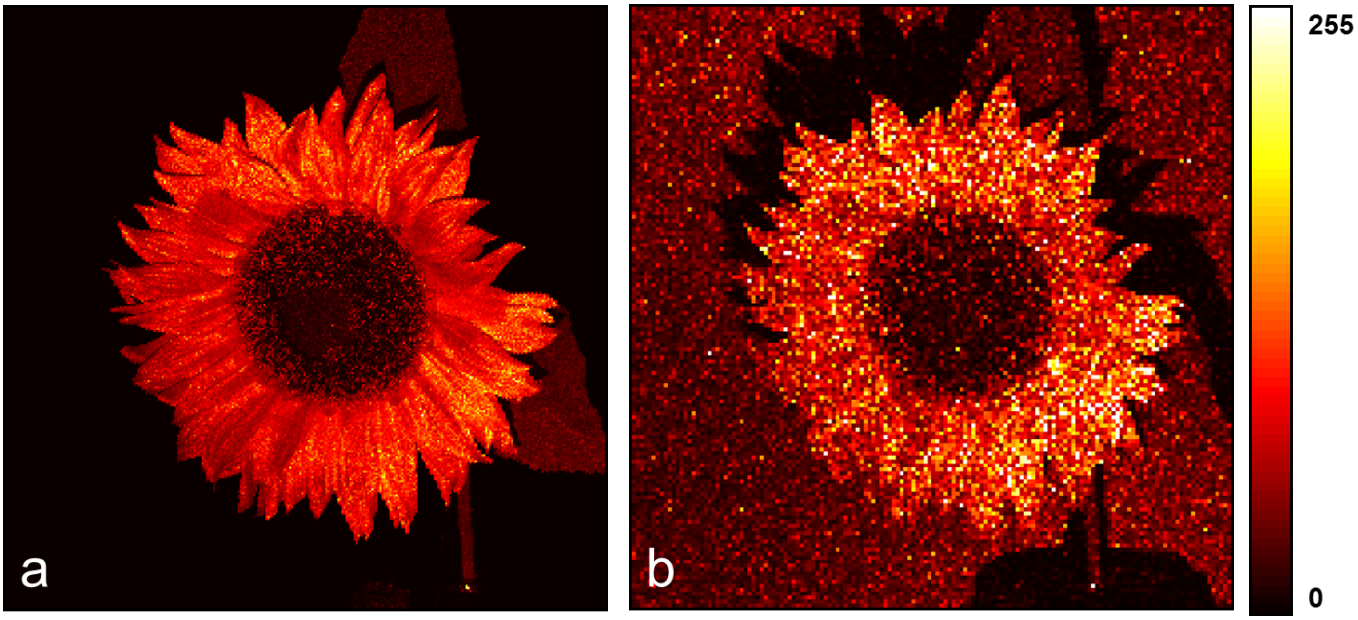


(c) Our processing.

(d) Median filtering.

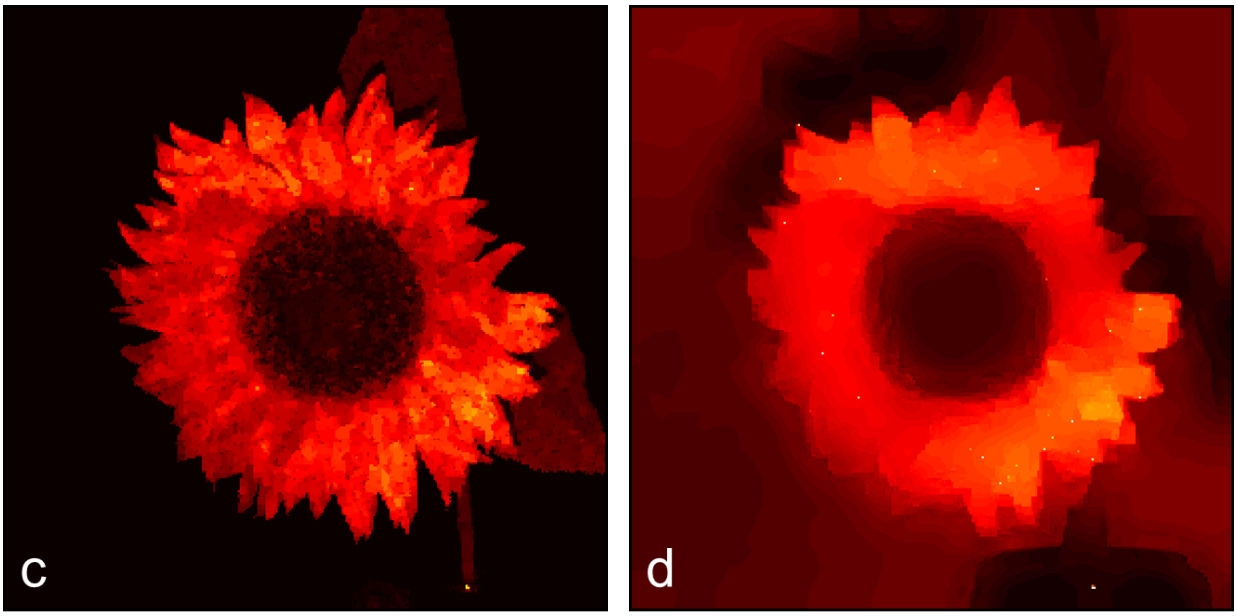
Scene photograph of sunflower dataset

Reflectivity images for sunflower dataset



(a) Ground truth.

Pointwise ML estimate.

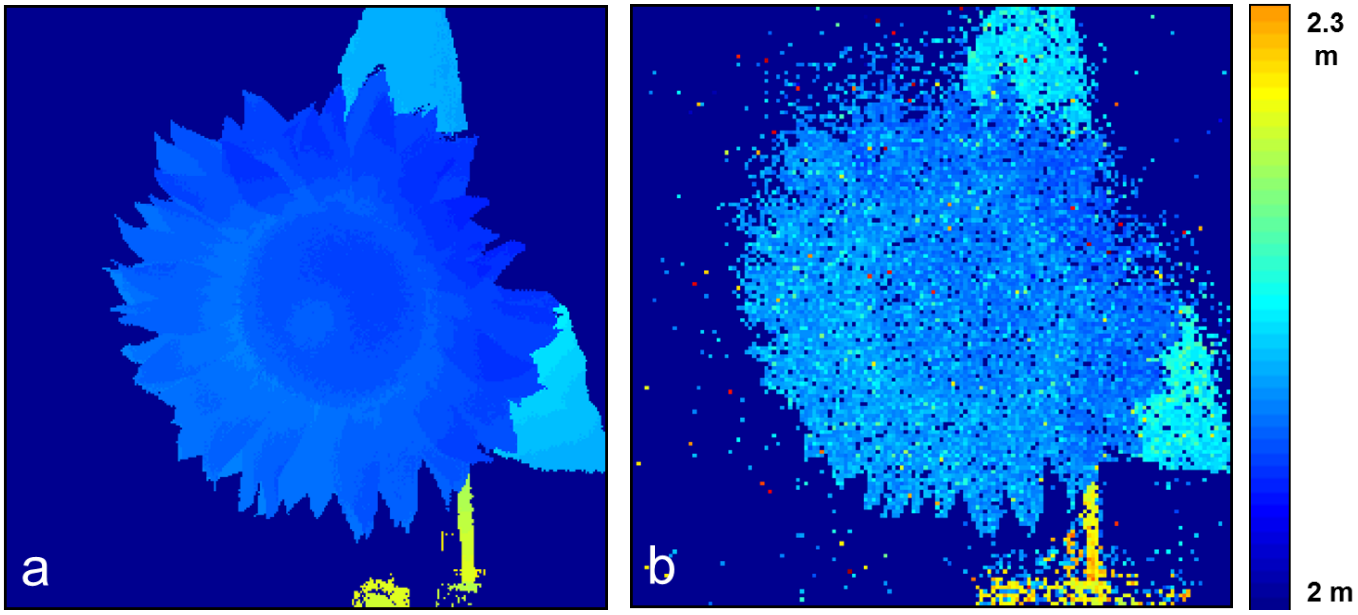


(c) Our processing.

(d) BM3D.

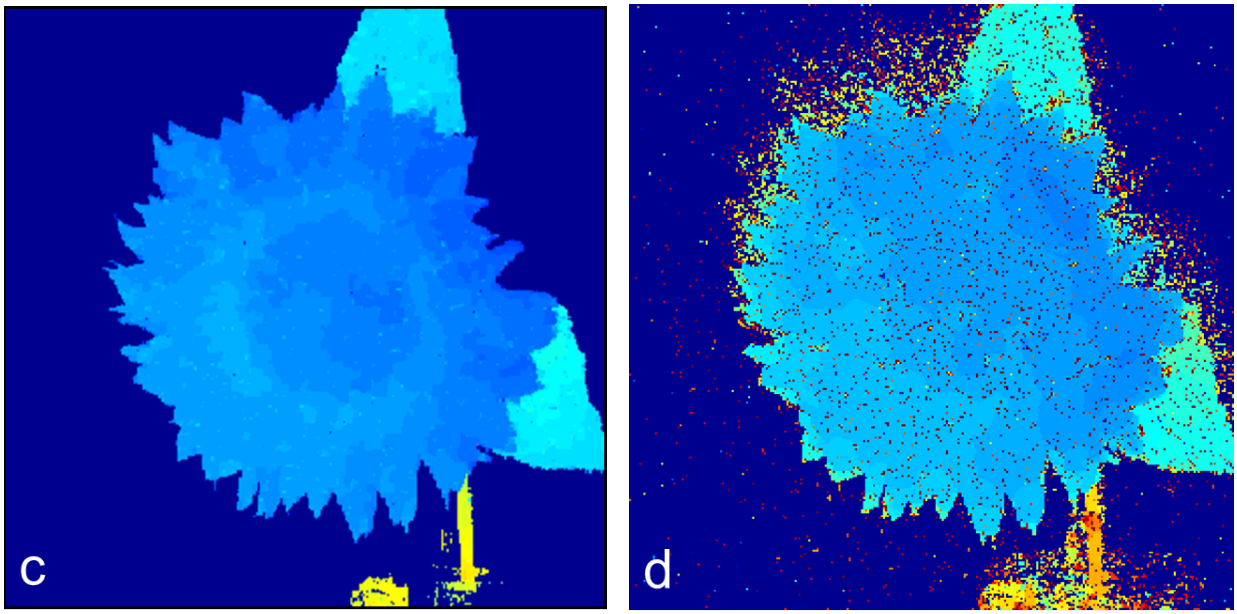
Figure 6: **Sunflower reflectivity image.** Our method rejects background and increases image contrast while retaining fine spatial features like flower petals. In comparison, BM3D reduces errors at the expense of oversmoothing and losing spatial features.

Depth maps for sunflower dataset



(a) Ground truth.

Pointwise ML estimate.



(c) Our processing.

(d) Median filtering.

Figure 7: **Sunflower depth map.** Our method rejects background and denoises while retaining fine spatial features like flower petals.

Scene Photograph of Mannequin Dataset

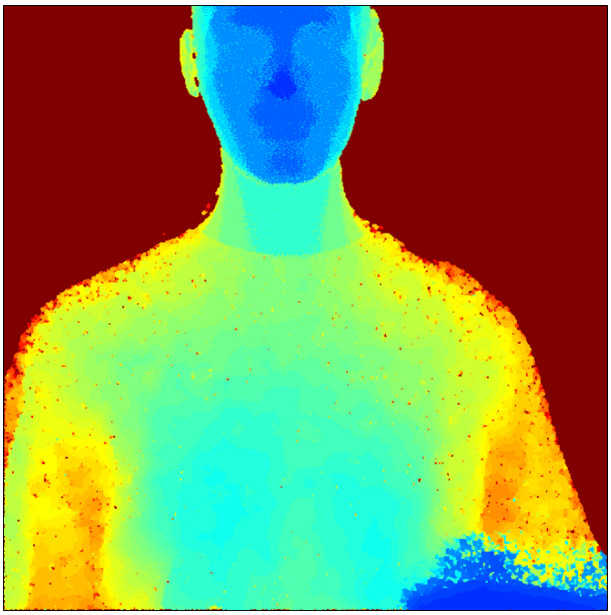
Reflectivity images for mannequin dataset

(a) Ground truth. Pointwise ML estimate.

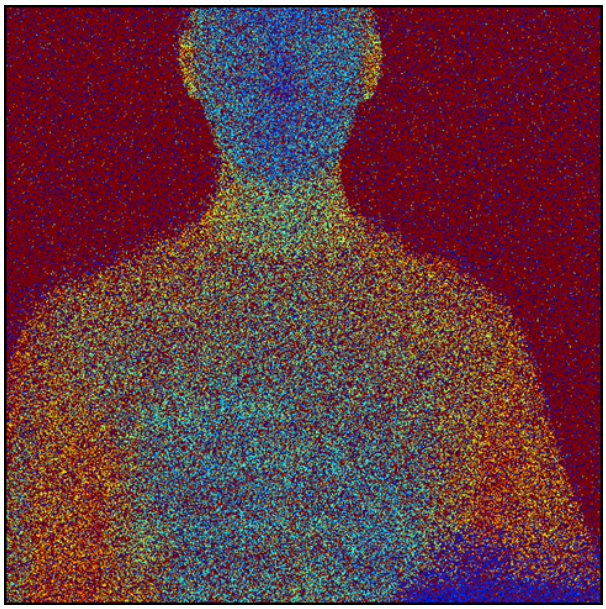
(c) Our processing. (d) BM3D with Anscombe transformation.

Figure 8: Note the recovery of heavily obscured text using our method.

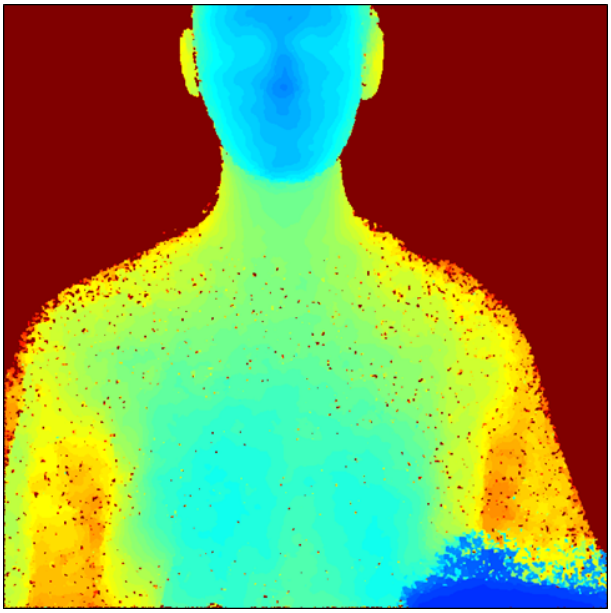
Depth maps for mannequin dataset



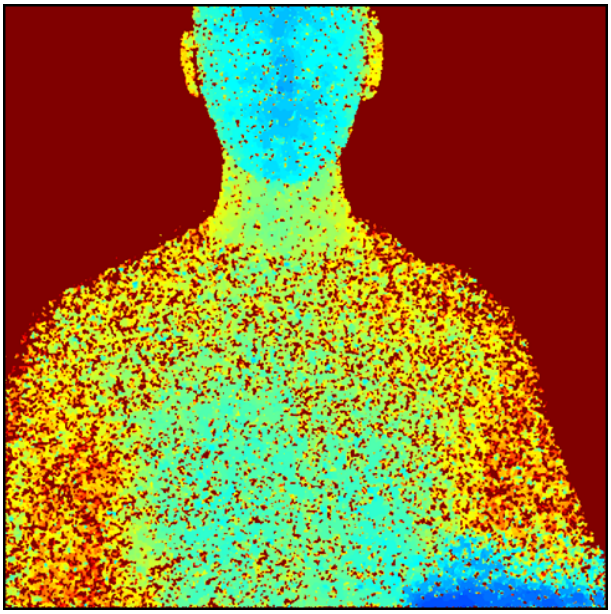
(a) Ground truth.



Pointwise ML estimate.



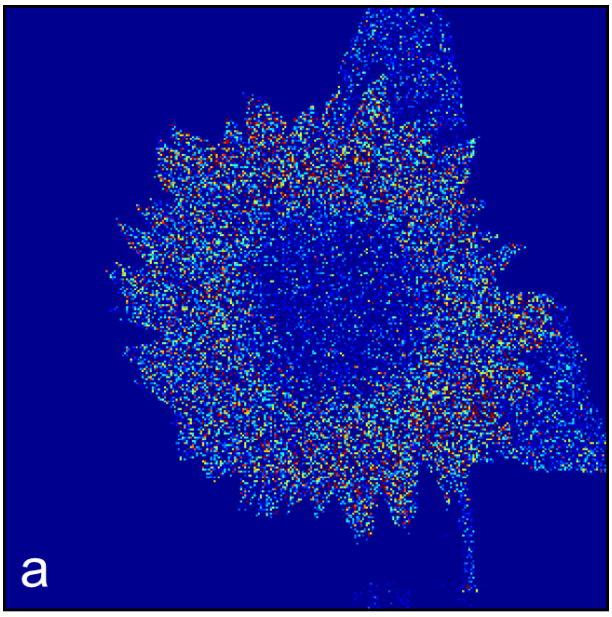
(c) Our processing.



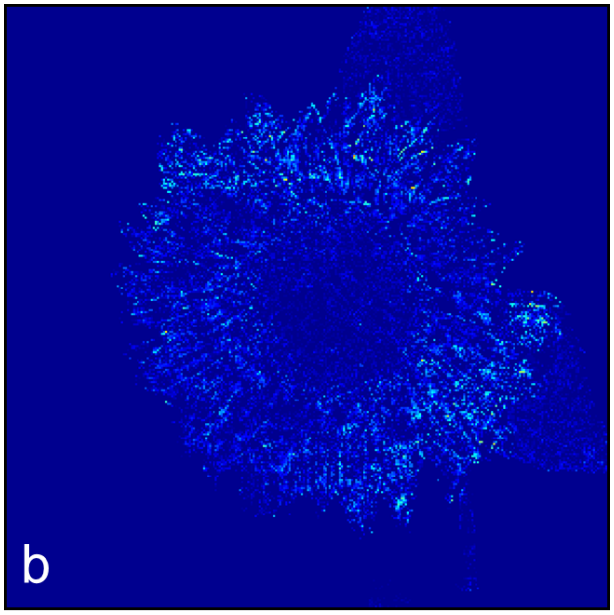
(d) Median filtering.

Additional figures and plots for error analysis

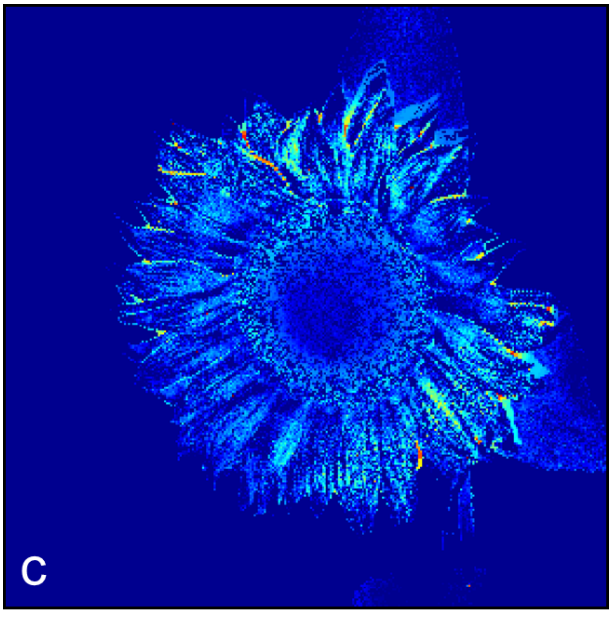
Reflectivity estimation error images for sunflower dataset



| Ground truth - pointwise ML |



| Ground truth - our method |

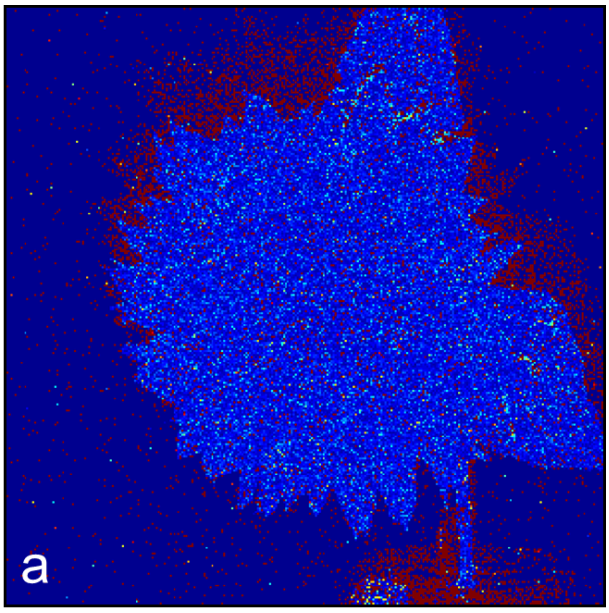


| Ground truth - regularized ML assuming AWGN |

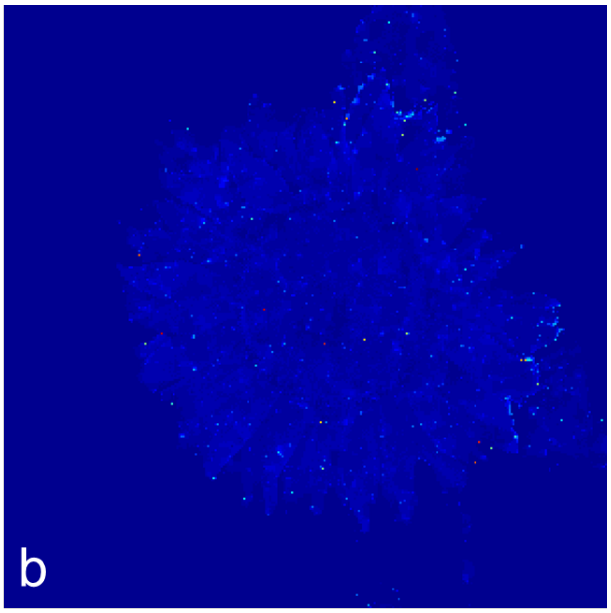
colorbar for (a)-(c)

Figure 9: Absolute difference images between reflectivity images. The high background error is masked out so avoid obscuring the details.

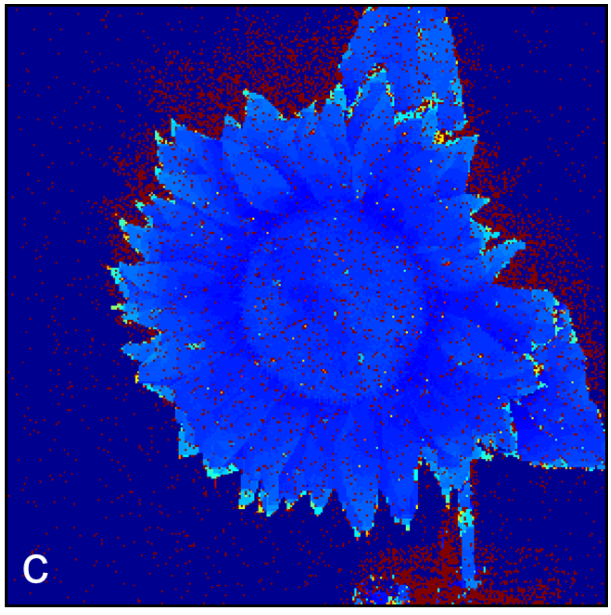
Depth estimation error images for sunflower dataset



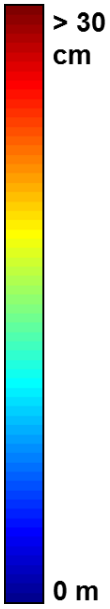
| Ground truth - pointwise ML |



| Ground truth - our method |



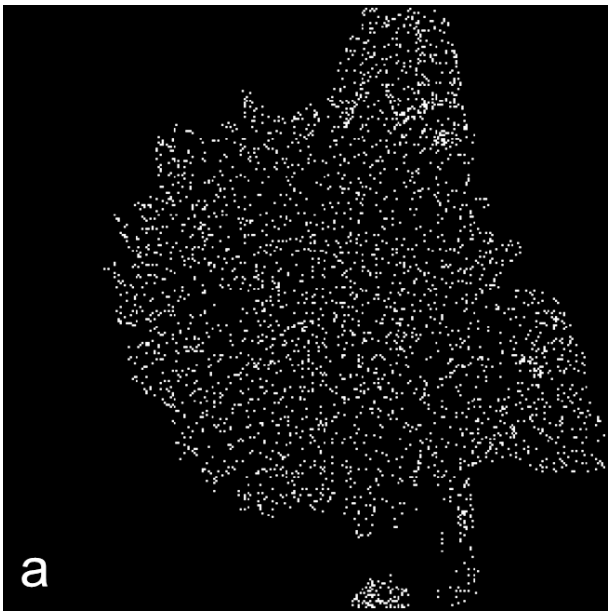
| Ground truth - regularized ML assuming AWGN |



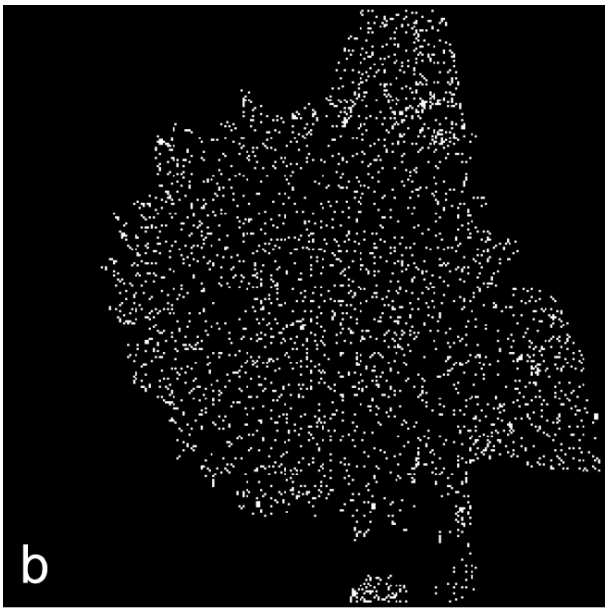
colorbar for (a)-(c)

Figure 10: Absolute difference between depth maps. G denotes ground truth.

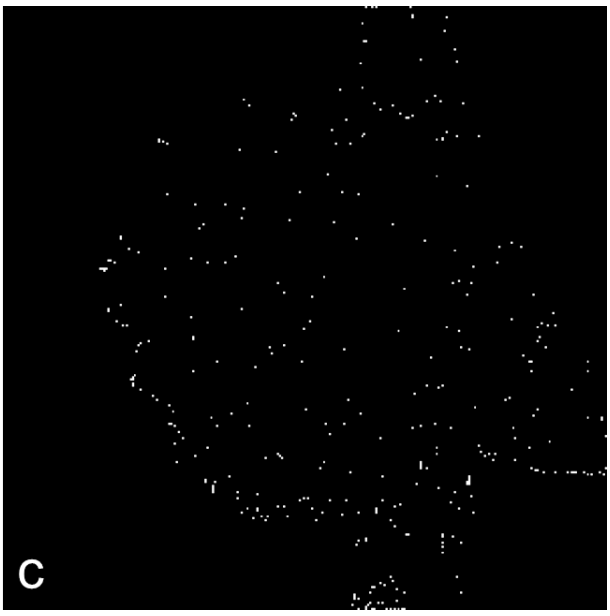
Noise rejection performance for sunflower dataset



Ground truth for noisy pixels



Our method based on ROADS



False positives and false negatives

Figure 11: (a) Noise photon labels obtained by thresholding high depth errors between ground truth and one photon per pixel data. (b) Noise photon labels identified by our framework (c) Pixelwise XOR of (b) and (c) indicating both false positives and false negatives. Note that our algorithm based on ROADS is highly successful at identifying noisy pixels.

Quantitative error analysis

For all processing methods used, the parameters were choses to minimize RMSE for depth maps and PSNR for reflectivity reconstruction.

	pointwise (or pixelwise) ML estimate	our method (first-photon imaging)	other denoising methods
sunflower depth	RMSE = 13.5cm	RMSE = 5.3mm	RMSE = 10.6cm (median filtering) 21.3cm (BM3D)
sunflower reflectivity	PSNR = 10.1dB	PSNR = 34.2dB	PSNR = 15.3dB (median filtering) 20.4dB (BM3D)
layered scene depth	RMSE = 15.7cm	RMSE = 6.8mm	RMSE = 11.8cm (median filtering) 19.4cm (BM3D)
layered scene reflectivity	PSNR = 7.6 dB	PSNR = 27.5dB	PSNR = 16.7dB (median filtering) 19.8dB (BM3D)
mannequin depth	RMSE = 21.2cm	RMSE = 2.4cm	RMSE = 14.7cm (median filtering) 27.3cm (BM3D)
mannequin reflectivity	PSNR = 11dB	PSNR = 35.9dB	PSNR = 11.5dB (median filtering) 18.3dB (BM3D)

Error distribution plots

Using the absolute error images for each dataset, the cumulative error distribution is computed. As shown in all the plots, our method dominates the curves for the other reconstruction methods.

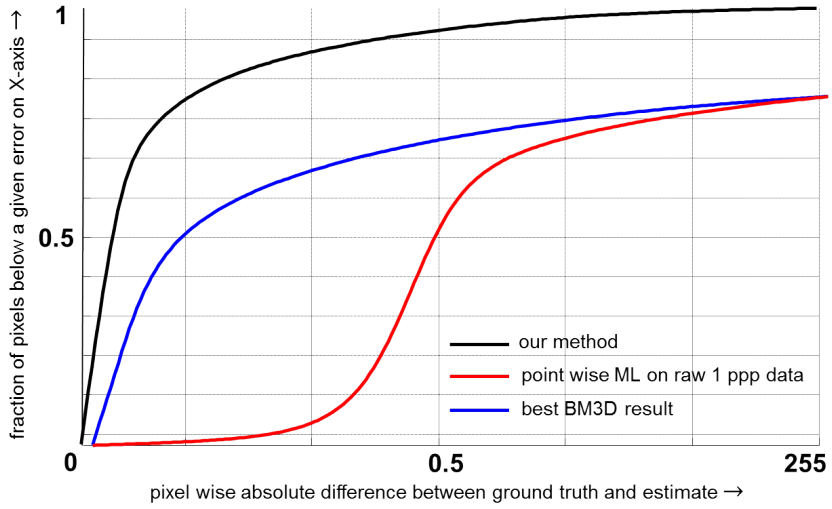


Figure 12: Mannequin dataset: reflectivity error cumulative distribution.

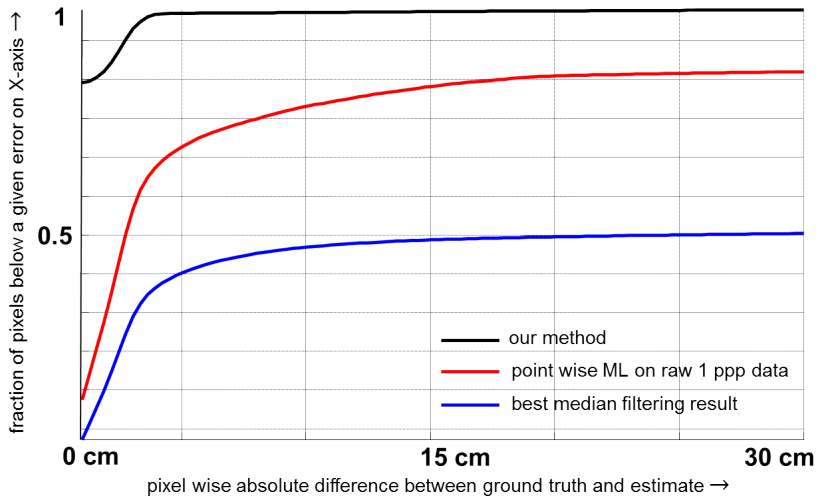


Figure 13: Mannequin dataset: depth error cumulative distribution.

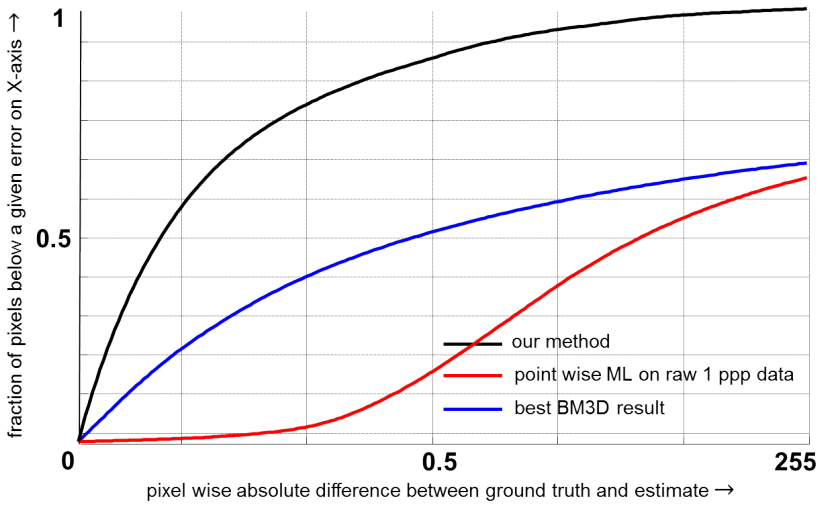


Figure 14: Layered scene dataset: reflectivity error cumulative distribution.

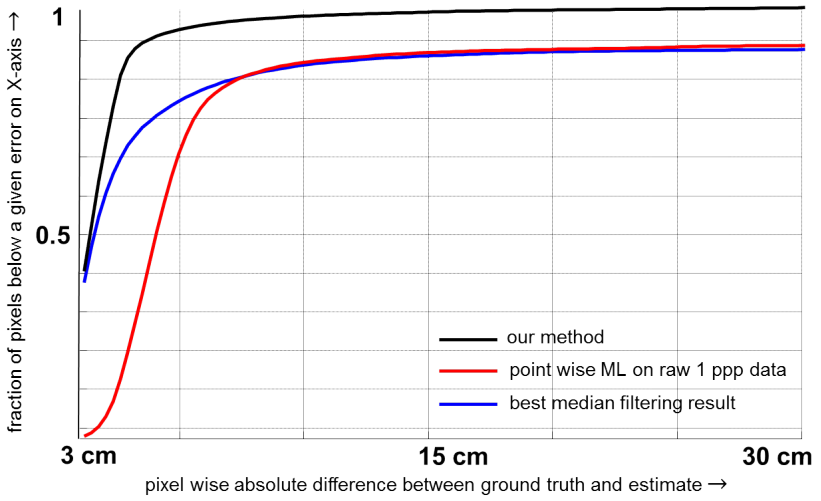


Figure 15: Layered scene dataset: depth error cumulative distribution.

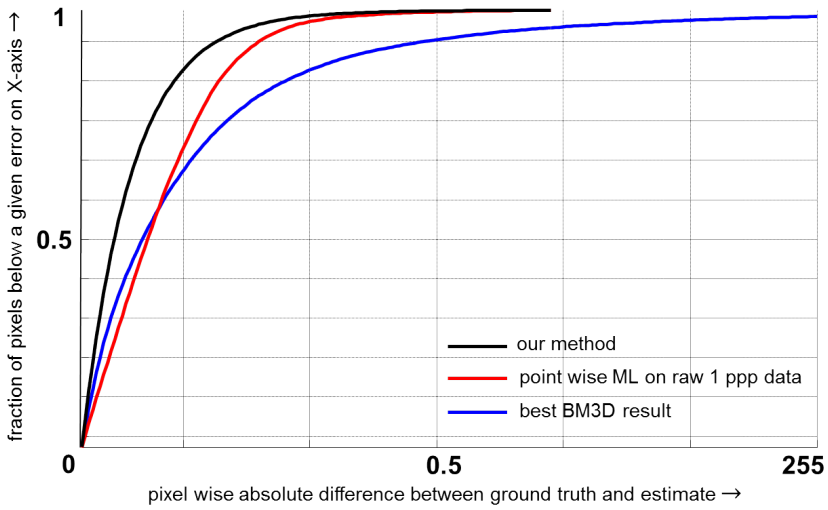


Figure 16: Sunflower dataset: reflectivity error cumulative distribution.

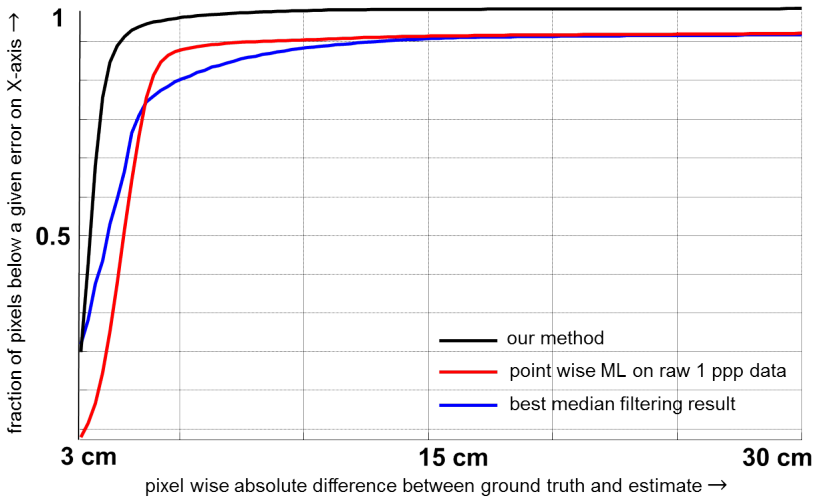


Figure 17: Sunflower dataset: depth error cumulative distribution.

Repeatability analysis

For testing repeatability, 500 independent first-photon datasets for the layered scene and mannequin were collected and processed using a fixed set of numerical and optical parameters across the datasets. The results for the layered scene dataset are discussed in the following figure (see movie `1696-kirmani-sup-movie3.mpg` for mannequin dataset repeatability test):

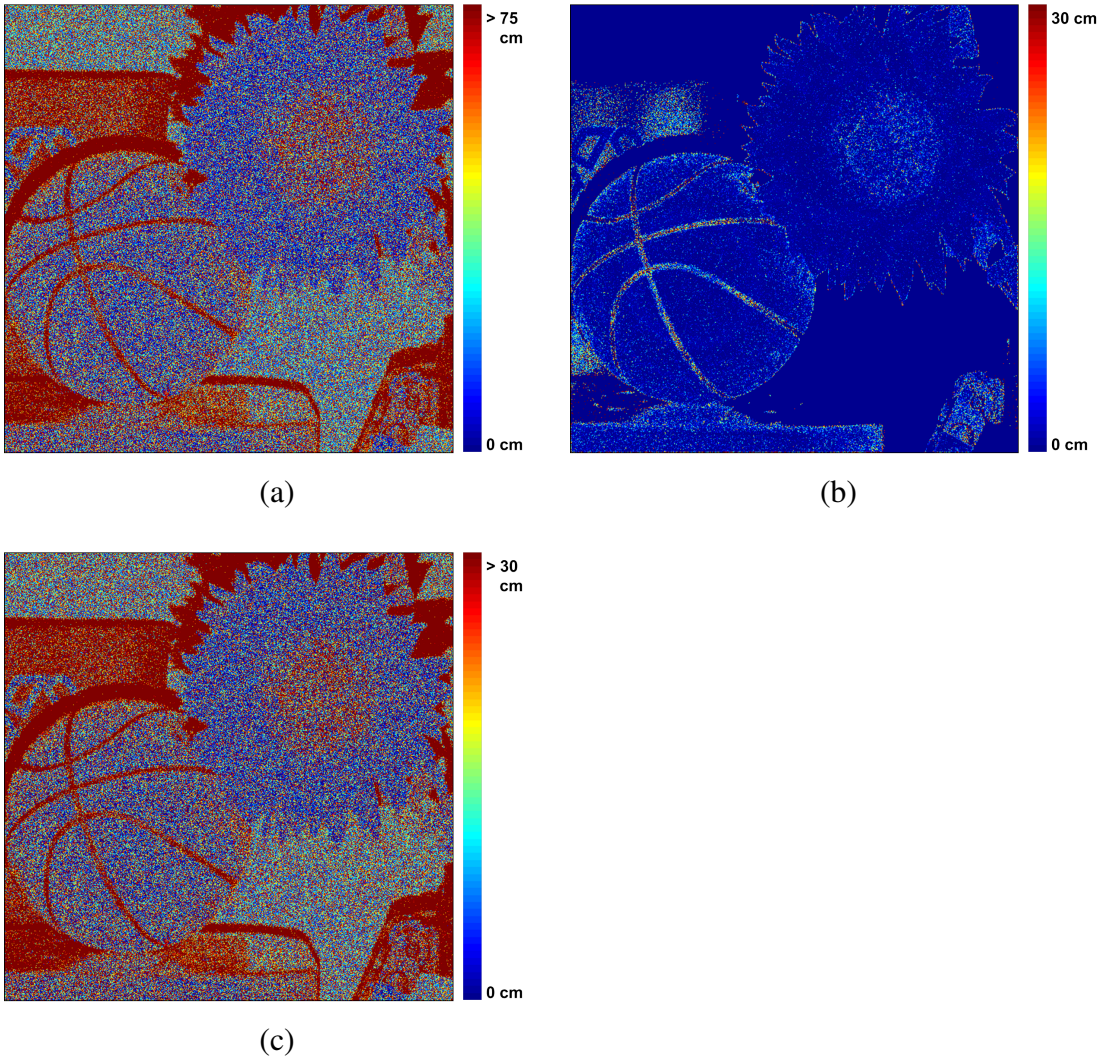


Figure 18: Pixelwise standard deviation of the posterior distribution computed by processing 500 first-photon data trials processed using: (a) Pointwise (or pixelwise ML) (b) Our proposed method (note the low standard deviation ($4-6$mm) observed throughout the image. This indicates that our method is robust and consistently improves estimation accuracy across independent trials. The darker pixels reduce the SNR and there estimation quality is poorer in these regions. The error is also high at the object edges and lateral faces. (c) Median filtering (which performed better than BM3D. Also note that the standard deviation is on the order 15 cm in most regions indicating that in the absence of a good noise model, other denoising methods fails to correct estimation errors.

Movie descriptions

Title for movie file named `1696-kirmani-sup-movie1.mpg`

Overview of the first photon imaging technique.

Title for movie file named `1696-kirmani-sup-movie3.mpg`

Video of the experimental setup and data collection.

Title for movie file named `1696-kirmani-sup-movie3.mpg`

Mannequin dataset repeatability test by processing 500 first-photon datasets using the proposed method and the pointwise ML technique.