Adaptive Gating for Single-Photon 3D Imaging: Supplementary Material

Contents

1. Potential negative societal impact 1

2. Timestamp versus histogram measurements 1

3. Assumptions in probabilistic model 2

4. Alternative reward functions 2

5. Proof of Proposition 1 2

6. Code and data 3

7. Experimental details 3

8. Exposure time comparisons and adaptive exposure 5

9. Effect of estimators and exposure time 6

10 Effect of dead time 7

11 Per-pixel depth uncertainties 8

12 Model mismatch 9

13 Additional figures 10

1. Potential negative societal impact

Our proposed depth sensing system and algorithm are similar to existing LiDAR systems, and thus share similar potential negative societal impacts. In particular, when depth outputs from LiDAR are used to guide decision-making in artificial intelligence systems, such as autonomous vehicles, it is important to account for situations where LiDAR does not perform well. Examples of such situations include bad weather conditions (e.g., snow and rain), strong ambient light (which our work mitigates), and scene elements with very low reflectivity. In the latter situation, it is important to draw attention to the fact that LiDAR performance can vary for different skin types: In particular, darker skin colors reflect less light, and thus result in lower signal to background ratio.

Additionally, LiDAR systems employ high-power lasers, which have associated safety hazards (e.g., risk of blindness). These are typically mitigated using short-wavelength infrared laser wavelengths, where safe power limits are a lot higher. We note that, in our experiments, we used a visible laser due to equipment constraints (we did not have available a laser or SPAD operating at the short-wavelength infrared range); but this is not an inherent limitation of our algorithm.

2. Timestamp versus histogram measurements

Prior work on SPAD-based LiDAR [1–3] typically assumes that the LiDAR system returns not a detection sequence $\bar{s}$, but a *detected photon histogram* $N[\tau] \in \mathbb{N}$, $\tau \in \{0, \ldots, \mathrm{T}-1\}$: each histogram entry is equal to the number of photons

detected during the corresponding temporal bin after P cycles. Given the gate sequence $\bar{g}$, it is easy to convert the detection sequence $\bar{s}$ into the detected photon histogram $N[\tau]$, but the reverse conversion is not possible. Additionally, the probabilistic models for the detection sequence $\bar{s}$ in Section 3.1 of the main paper and detected photon histogram $N[\tau]$ in Gupta et al. [1] are consistent with each other, assuming that detection times are conditionally independent given gating times. Both our proposed adaptive gating scheme in Section 4 of the main paper and the uniform gating scheme of Gupta et al. [1] make use of this conditional independence condition.

We opt to use detection times in our presentation, as this simplifies the discussion of our adaptive gating scheme. We also note that modern TCSPC systems provide access to individual detection times [7]. However, our adaptive gating scheme would be applicable even if only measurements of the detected photon histogram were available. In that case, Algorithm 1 of the main paper would use a depth posterior conditioned on the detected photon histogram $N[\tau]$ instead of the detection sequence S. Such a posterior could be computed (and updated) using the Poisson-multinomial distribution likelihood model for the detected photon histogram $N[\tau]$ from Gupta et al. [1].

3. Assumptions in probabilistic model

The probabilistic model of Section 3.1 of the main paper makes three important assumptions, also implicit in the prior work we derive this model from [1,2]. First, it assumes that the laser emits pulses of infinitesimal duration. Second, it assumes that the TCSPC operation is free from electronic timing noise (*jitter*) and other non-idealities such as afterpulsing [4]. Third, it assumes that the scene transient $\lambda_d[\tau]$ is due to only direct illumination, and is free from indirect illumination effects such as interreflections and subsurface scattering.

We note that we use the infinitesimal-pulse and no-jitter assumptions to simplify exposition, and that they are not essential for our analysis: We can easily adapt the algorithms we introduce in Sections 4-5 of the main paper to use a probabilistic model that replaces the scene transient $\lambda_d[\tau]$ of Equation 2 of the main paper with a version that is convolved with arbitrary pulse and jitter shapes. This modification does not increase algorithm complexity, and is analogous to the approach of Heide et al. [3] for accounting for finite pulses. Accounting for other TCSPC non-idealities and for indirect illumination is more difficult; we leave these as future research directions.

4. Alternative reward functions

We note that an alternative to the reward function of Equation (15) of the main paper is

$$R^*\left(\tilde{d}, g\right) \equiv -\mathbb{E}_{s\sim p_{\tilde{d}}^{g}(s)}\left[\mathcal{L}_0\left(\hat{d}_{\text{MAP}}\left(\bar{s}_p, \bar{g}_p\right), \tilde{d}\right)\right]. \tag{1}$$

That is, we compute a reward using a depth estimate from not just the next detection time s_p, but also from all already recorded detection times $\bar{s}_{p-1}$. We do not use this reward function for two reasons: First, it increases the computational complexity of Algorithm 1 of the main paper, as we can no longer use Proposition 1 of the main paper to analytically select the gate g_p that maximizes R^*; instead, this maximization must be done numerically by exhaustively evaluating the reward function for all possible gate values. Second, we found empirically that using this reward function results in less effective exploration during the early stages of Thompson sampling.

5. Proof of Proposition 1

We prove Proposition 1 of the main paper, which we restate here for convenience.

Proposition 1. *The solution to the optimization problem*

$$\tilde{g} \equiv \underset{g\in\{0,\ldots,\mathrm{T}-1\}}{\operatorname{argmax}} R\left(\tilde{d}, g\right), \tag{2}$$

for the reward function

$$R\left(\tilde{d}, g\right) \equiv -\mathbb{E}_{s\sim p_{\tilde{d}}^{g}(s)}\left[\mathcal{L}_0\left(\hat{d}_{MAP}\left(s, g\right), \tilde{d}\right)\right], \tag{3}$$

equals $\tilde{g} = \tilde{d}$.

Proof. We note first that, for a single measurement s, $\hat{d}_{\text{MAP}}(s, g) = s \bmod \mathrm{T}$. Then,

$$R\left(\tilde{d}, g\right) \equiv -\mathbb{E}_{s \sim p_{\tilde{d}}^{g}(s)}\left[\mathcal{L}_0\left(\hat{d}_{\text{MAP}}(s, g), \tilde{d}\right)\right] \tag{4}$$

$$= \sum_{s=g}^{g+\mathrm{T}-1} p_{\tilde{d}}^{g}(s)\left(1 - \delta_{\tilde{d}, s \bmod \mathrm{T}}\right) \tag{5}$$

$$= \sum_{s=g}^{g+\mathrm{T}-1} p_{\tilde{d}}^{g}(s) - p_{\tilde{d}}^{g}\left(\tilde{d}\right) \tag{6}$$

$$= 1 - p_{\tilde{d}}^{g}\left(\tilde{d}\right), \tag{7}$$

where in the last step, we used our assumption that the SPAD always detects a photon during a period equal to the laser's pulse-to-pulse period after the gate (Section 3 of the main paper). This is not an essential assumption, and removing it will simply change the constant 1 to a number smaller than 1. Additionally, when we write $p_{\tilde{d}}^{g}\left(\tilde{d}\right)$, we slightly abuse our notation to simplify exposition: The first argument should be $\tilde{d}$ if $\tilde{d} > g$, and $\tilde{d} + \mathrm{T}$ otherwise, to account for the fact that $s \in \{g, g + \mathrm{T} - 1\}$.

Thus,

$$\tilde{g} \equiv \underset{g \in \{0, \ldots, \mathrm{T}-1\}}{\operatorname{argmax}} R\left(\tilde{d}, g\right) \tag{8}$$

$$= \underset{g \in \{0, \ldots, \mathrm{T}-1\}}{\operatorname{argmin}} p_{\tilde{d}}^{g}\left(\tilde{d}\right) \tag{9}$$

$$= \tilde{d}, \tag{10}$$

where the last step follows from inspection of Equation (5) of the main paper. □

6. Code and data

We provide anonymized code and data to reproduce select figures in the paper at the following link: https://anonymous.4open.science/r/adaptive_gating-D53A/README.md.

Link to data for the "Leaf" scene can be found here: https://www.dropbox.com/s/y7tx0wafwy6a751/leaf.npy?dl=0

We will release all of our data upon acceptance. We cannot provide the full scan data as part of the supplement, because of the size limitations for supplementary material.

7. Experimental details

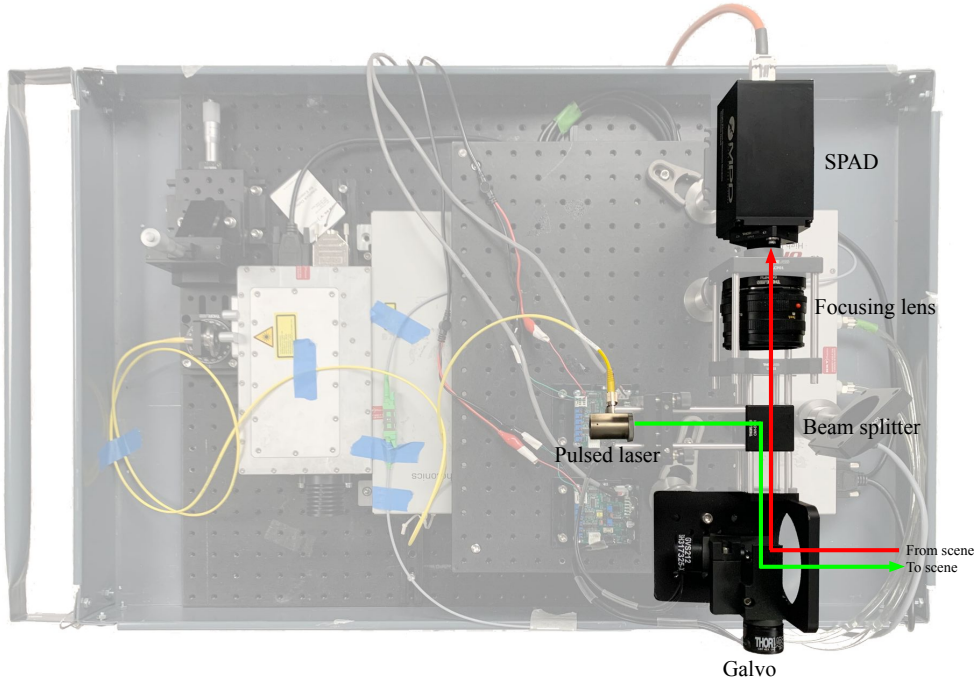
Prototype. Our SPAD-based LiDAR prototype comprises a fast-gated SPAD (Micro-photon Devices), a $532\,\mathrm{nm}$ picosecond pulsed laser (NKT Photonics NP100-201-010), a TCSPC module (PicoHarp 300), a programmable picosecond delayer (Micro-photon Devices), and a pair of galvo mirrors (Thorlabs GVS212) for beam steering. We operated the SPAD in triggered mode (adaptive gating) or free-running mode. Supplementary Figures 1-2 show photographs of our experimental prototype.

Dead time. In our experiments, we set the SPAD to operate at a dead time of $81\,\mathrm{ns}$. This is the default setting of the programmable dead time of the SPAD model we used for our experiments. We note that the minimum programmable dead time for our SPAD is $48\,\mathrm{ns}$. However, we found that operating the SPAD in free-running mode with at dead times smaller than $81\,\mathrm{ns}$ led to strong artifacts in the captured histograms and detection timestamps. Therefore, we performed all our experiments using the default setting of $81\,\mathrm{ns}$.

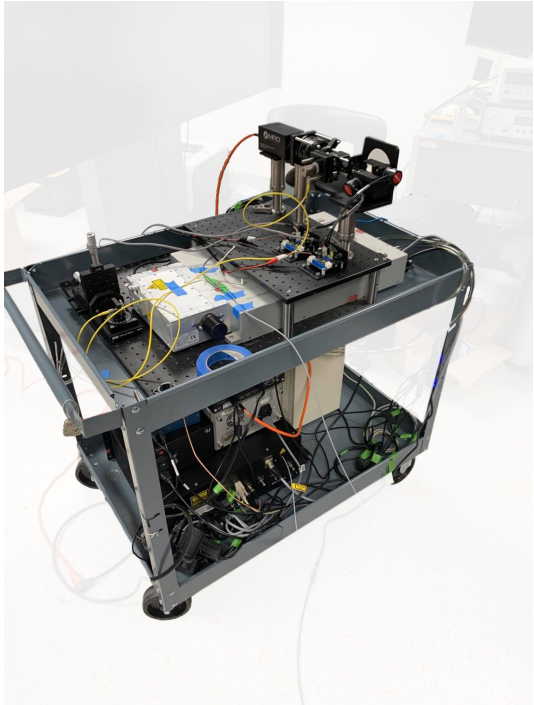
Time discretization. We follow Gupta et al. [1] and discretize our measurements to $\mathrm{T} = 500$ temporal bins, corresponding to a temporal resolution of $100\,\mathrm{ps}$. We set the laser pulse frequency to $20\,\mathrm{MHz}$.

Depth estimation. We follow Gupta et al. [1] (personal communication) and improve depth estimation using a post-processing temporal dithering step. To this end: First, we use either the Coates' or the MAP depth estimate, as we describe in Section 3.2 of the main paper, to form a depth estimate $\hat{d}$ at a resolution equal to the temporal bin resolution Δ. Then, we fit a Gaussian to the portion of the Coates' estimate of the scene transient (Equation (9) of the main paper) a few temporal bins

around the depth estimate $\hat{d}$. Lastly, we form a new depth estimate at sub-bin resolution, by selecting the max of the fitted Gaussian.

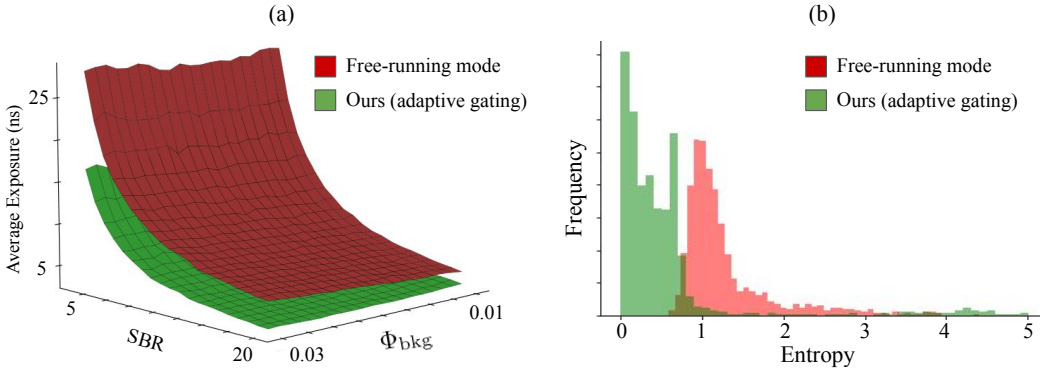


Supplementary Figure 1. **Top-down view of experimental setup.** Our SPAD-based LiDAR prototype comprises a fast-gated SPAD and a high-power $532\,\mathrm{nm}$ picosecond pulsed laser. The setup uses a TCSPC module (not shown) for timestampling photon detections, and a picosecond delayer (not shown) to implement gating. Lastly, the setup uses a pair of galvo mirrors for beam steering.



Supplementary Figure 2. **Side view of experimental setup.** We place our experimental setup on a cart, to make it easier to capture outdoor scenes.

8. Exposure time comparisons and adaptive exposure



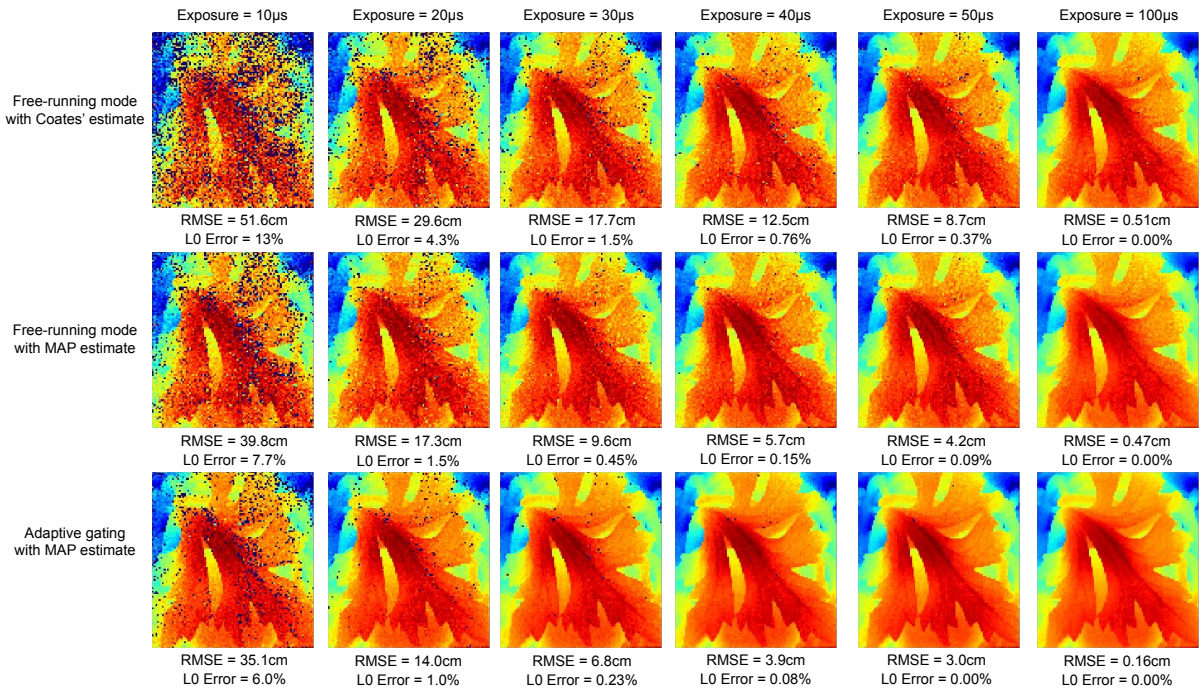
Supplementary Figure 3. (a) Average exposure for varying levels of ambient flux and SBR using adaptive exposure with an entropy threshold of 0.25. (b) Distribution of final depth posterior entropy of single-pixel experiment with $\Phi_{\text{bkg}} \approx 0.01$, $\Phi_{\text{bkg}} \approx 0.10$ and a fixed exposure of $100\mu s$.

In Supplementary Figure 3(a), we use simulations to show the exposure time needed by free-running mode and our adaptive gating scheme, to reach a fixed threshold in depth posterior uncertainty (as measured by the termination function), under different SBR and Φ_{bkg} conditions. We note that, in all cases, our adaptive gating scheme requires a shorter exposure time, thanks to the increased number of signal photons it detects. Combining adaptive gating with our adaptive exposure scheme allows us to take advantage of this improved performance automatically, without requiring us to determine a fixed exposure time to be used for each set of SBR and Φ_{bkg} conditions.

In Supplementary Figure 3(b), we show the depth posterior uncertainty (as measured by termination function) achieved by free-running mode and our adaptive gating scheme under a fixed exposure time, for multiple simulations of the *same*

underlying scene transient. We note that, for both algorithms, there is significant variation in the depth posterior uncertainty they achieve, despite the fact that the simulated scene never changes. This is due to the random nature of photon arrivals and, in the case of adaptive gating, the randomness in Thompson sampling. We observe that, in the majority of cases, adaptive gating achieves lower depth posterior uncertainty than free-running mode. We also observe that both algorithms have a "long tail" of instances where the achieved depth posterior uncertainty is higher. Our adaptive exposure procedure acts as a simple remedy for such instances: it allows the scanning algorithm to adaptively use longer exposure times in the rare cases when we are in the long tail of this distribution, and use lower exposure times in the majority of cases when we are in the main area of the distribution.

9. Effect of estimators and exposure time

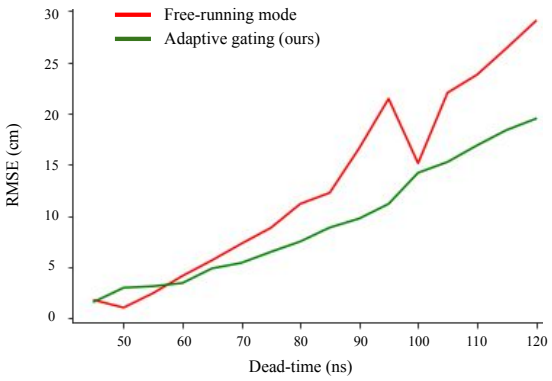


Supplementary Figure 4. **Depth reconstruction of "Leaf" scene under various exposure lengths.** We combine free-running mode with two different estimators, for different exposure times. The combination with the MAP estimator outperforms the combination with Coates' estimator for all exposure times. Additionally, adaptive gating outperforms both versions of free-running mode for all exposure times.

As we mention in Section 3.2 of the main paper, our proposed MAP depth estimator can potentially improve depth accuracy compared to the Coates' depth estimator of Gupta et al. [1], independently of how measurement acquisition is performed. To demonstrate this, in the top two rows of Supplementary Figure 4, we use the "Leaf" scene to compare the performance of free-running mode when combined with either the Coates' estimator, or our proposed MAP estimator. We note that the MAP estimator consistently outperforms the Coates' estimator. Therefore, in the main paper, for comparisons we always combine free-running mode with the MAP estimator.

We also compare both variants of free-running mode with our adaptive gating scheme (which uses the MAP depth estimator), for different exposure times. We observe that our adaptive gating scheme outperforms both variants of free-running mode under all exposure times.

10. Effect of dead time



Supplementary Figure 5. **Performance of SPAD acquisition modes under various dead time settings.** The performance of adaptive gating degrades as dead time decreases, but remains better than the performance of free-running mode for most dead time settings (including all settings higher than the practical lower bound of $81\,\mathrm{ns}$).

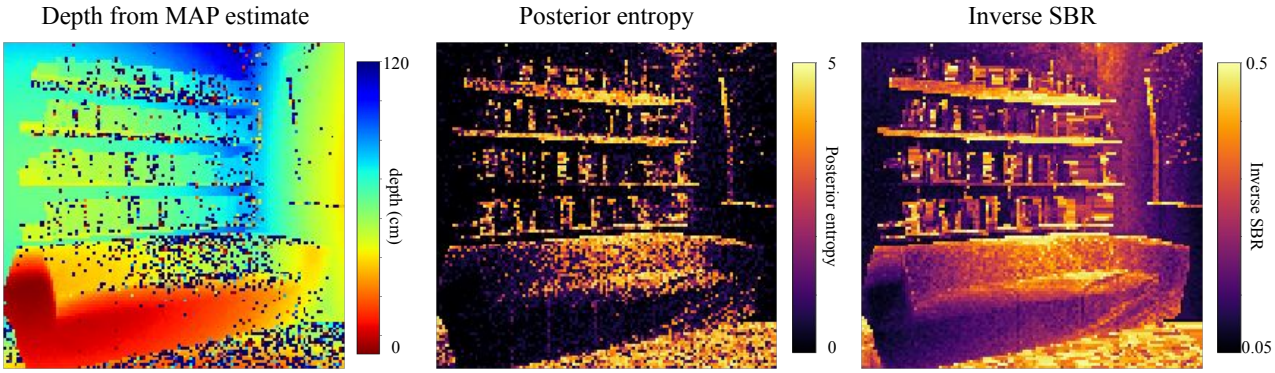
Adaptive gating has the undesirable effect of increasing the SPAD inactive time, because it keeps SPAD inactive during the period between the end of dead time in a SPAD cycle and the gate time at the subsequent cycle. By contrast, free-running mode minimizes SPAD inactive time, by activating the SPAD immediately after dead time ends. As a result, free-running mode results in a larger number of photon detections compared to adaptive gating, with this difference increasing as dead time decreases.

In Supplementary Figure 5, we use simulations to evaluate the effect of different dead time settings on depth accuracy, for free-running mode versus our adaptive gating scheme. We note that, even though free-running mode minimizes SPAD inactive time and results in more photon detections, it performs worse than our adaptive gating scheme for most practical dead time settings. As we mention in Supplementary Section 7, we could not reliably operate our SPAD at dead times lower than $81\,\mathrm{ns}$.

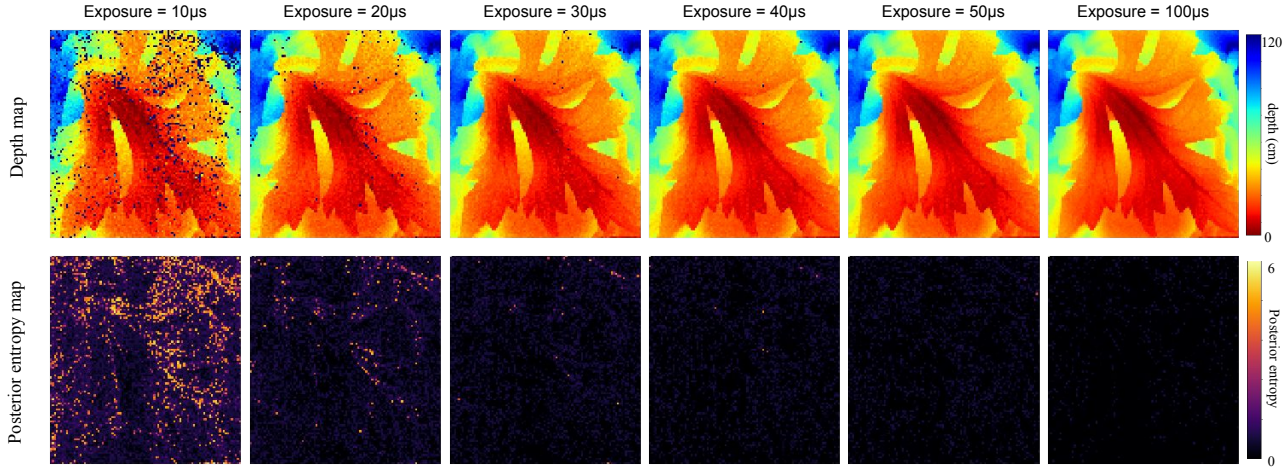
We note an interesting artifact when the dead time equals $100\,\mathrm{ns}$. At this point, the dead time is equal to the pulse-to-pulse repetition time T that we used for simulations. As a result, free-running mode effectively sets the "gate" for the next SPAD cycle at exactly the detection time of the last detected photon. This is exactly analogous to the operation of adaptive gating after the exploration stage is over (i.e., after the first few SPAD cycles), hence explaining the similar performance of the two acquisition schemes at this dead time setting.

It is important to point out here that, if it were possible to reduce dead time to zero, then the SPAD would be able to detect all incident photons, and thus pile-up distortion would not exist. Under such an ideal situation, free-running mode would be the optimal mode of operation for SPADs.

11. Per-pixel depth uncertainties



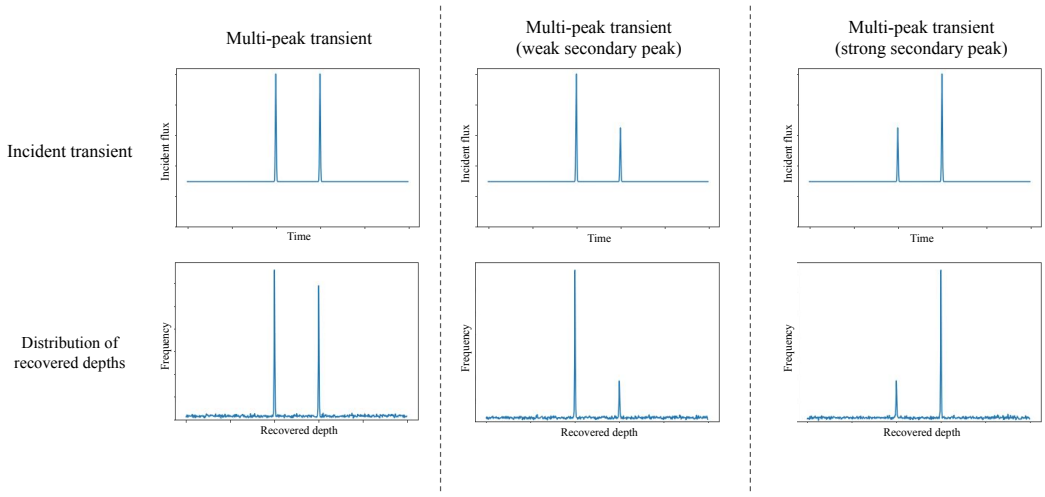
Supplementary Figure 6. **Depth reconstruction of "Office" scene with corresponding depth posterior entropy at each scene point.** Posterior entropy is high in regions where depth errors occur. As expected, regions with high posterior entropy and high error correspond to regions with high SBR.



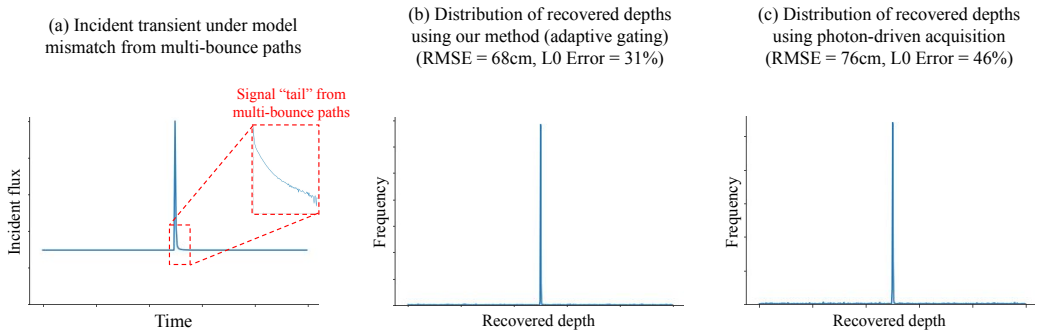
Supplementary Figure 7. **Depth reconstruction of "Leaf" scene at various exposures with corresponding depth posterior entropy at each scene point.** As exposure increases, the depth posterior entropy at each scene point decreases on average. Regions with high posterior entropy correspond to regions in the depth map with high depth errors.

Our adaptive gating scheme outputs not just a point estimate for depth, but also the entire depth posterior. This can be useful for downstream tasks that require both point estimates *and* uncertainties [5,6,8]. As a visualization, in Supplementary Figures 6 and 7, we show per-pixel depth uncertainties (computed as the entropy of the depth posterior) for the "Ofice" and "Leaf" scenes. In Supplementary Figure 6, we note that posterior entropy (and thus depth uncertainty) is highly correlated with SBR; in particular, as SBR increases, posterior entropy decreases. This is expected, given that higher SBR implies reduced pile-up artifacts. Additionally, in Supplementary Figure 7, we observe that as exposure time increases, depth uncertainty decreases.

12. Model mismatch



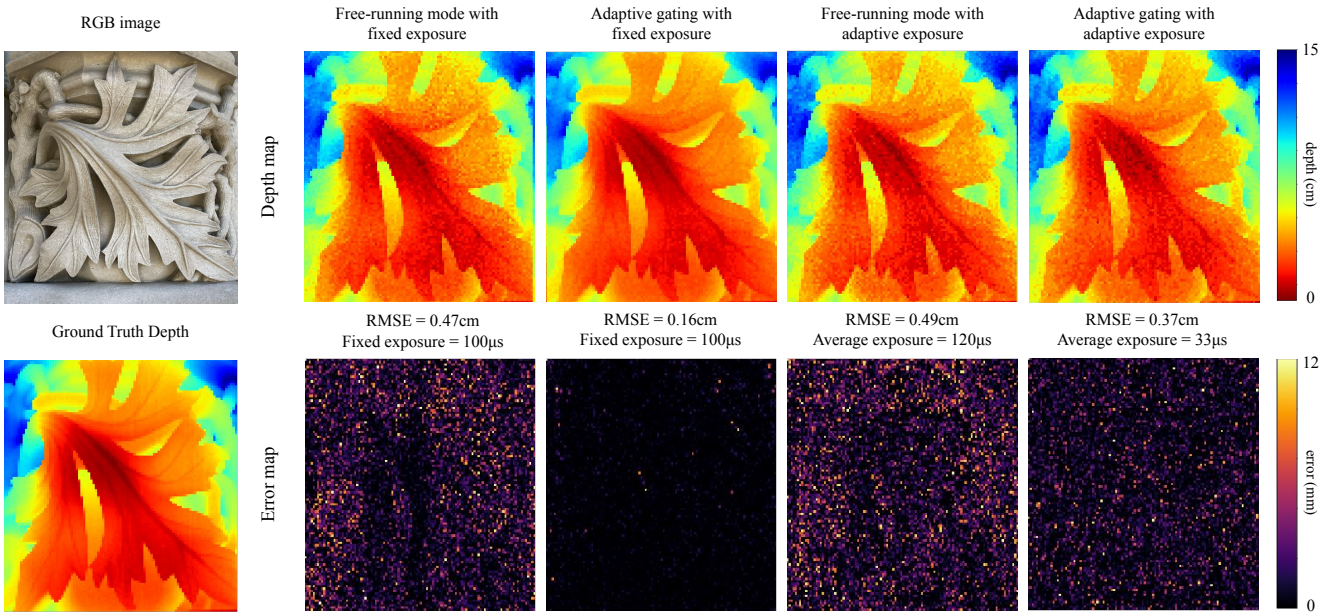
Supplementary Figure 8. **Behavior of adaptive gating under model mismatch: multiple peaks.** When multiple peaks are present in the incident transient, a model mismatch occurs to the single-peaked probabilistic model for the MAP estimator. The behavior of adaptive gating under such scenarios depends on the relative strengths of these peaks. Adaptive gating will probabilistically converge to one of these depths, with likelihood positively correlated to the signal strength at that depth.



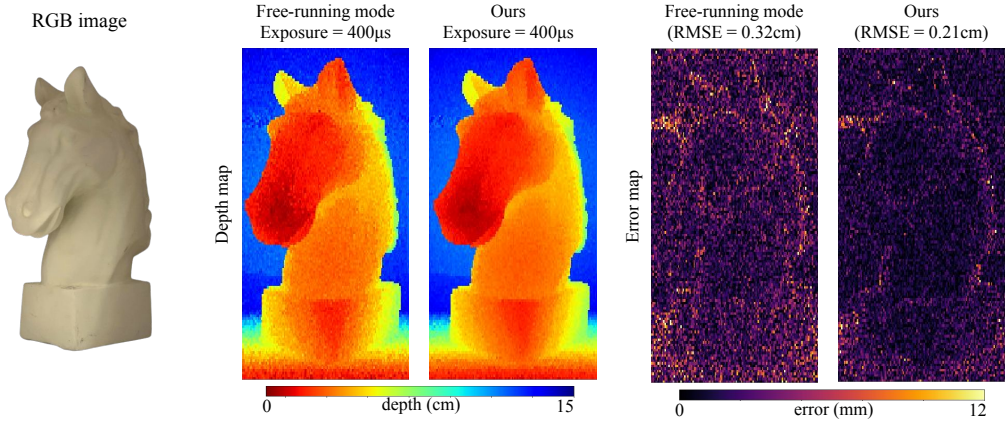
Supplementary Figure 9. **Behavior of adaptive gating under model mismatch: multi-bounce pahts.** (a) shows an example transient resulting from imaging a corner, where the presence of mult-bounce paths leads to a "tail" appearing after the true signal. (b) and (c) shows the distribution of recovered depths using our method (adaptive gating) and photon-driven acquisition respectively. Note that under model mismatch, our method still performs better than photon-driven acquisition.

In scenes where multi-bounce paths are common, e.g. a corner, the incident transient may not adhere well to the assumed model. In such cases, multi-bounce paths causes additional photons to return after the initial signal, resulting in a decaying "tail" that can be observed in Supplmentary Fig. 9(a). Since light returning to the SPAD after experiencing multiple bounces tend to be much weaker and also due to the confocal nature of our setup, multi-bounce artifacts are often very weak, meaning the transients still closely ressemble the assumed model. Supplmentary Fig. 9(b) and (c) illustrates how our method and photon-driven method are able to recover the correct depth under model mismatch, with our method still performing better despite deviation from our assumed model.

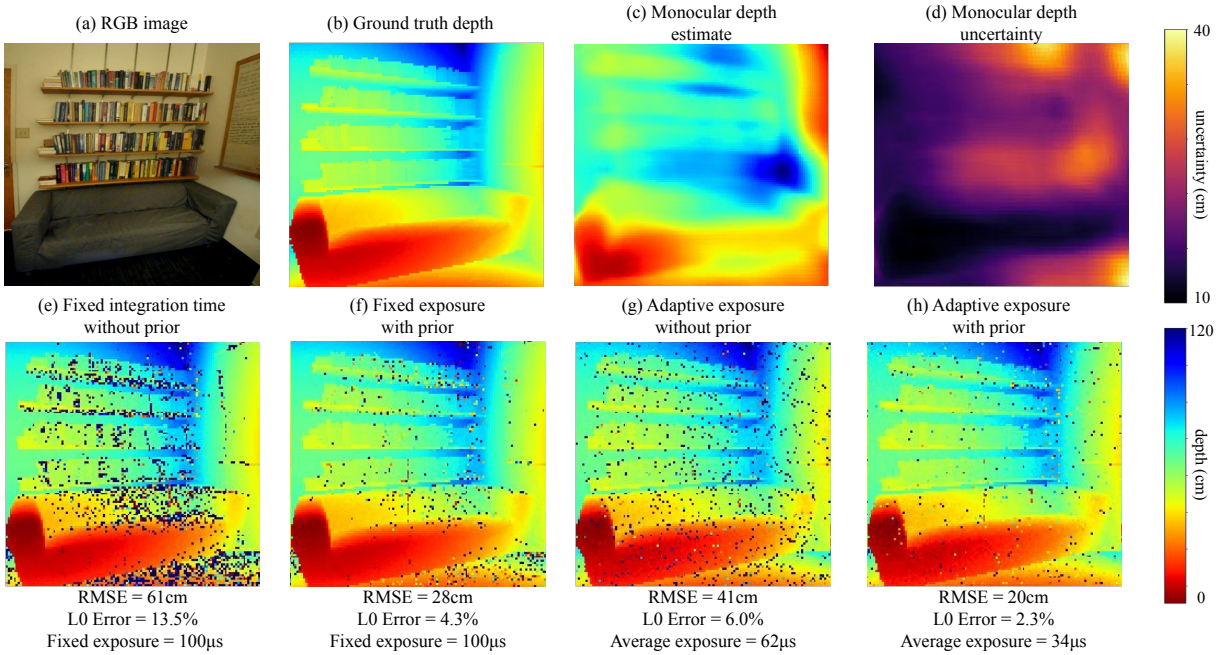
13. Additional figures



Supplementary Figure 10. **Depth reconstruction of "Leaf" scene with corresponding error maps.** Under equal fixed integration times, adaptive gating leads to lower errors across the entire scene compared to free-running mode. Using adaptive exposure, adaptive gating is able to achieve lower RMSE despite only using $\frac{1}{3}$ of the total exposure time.

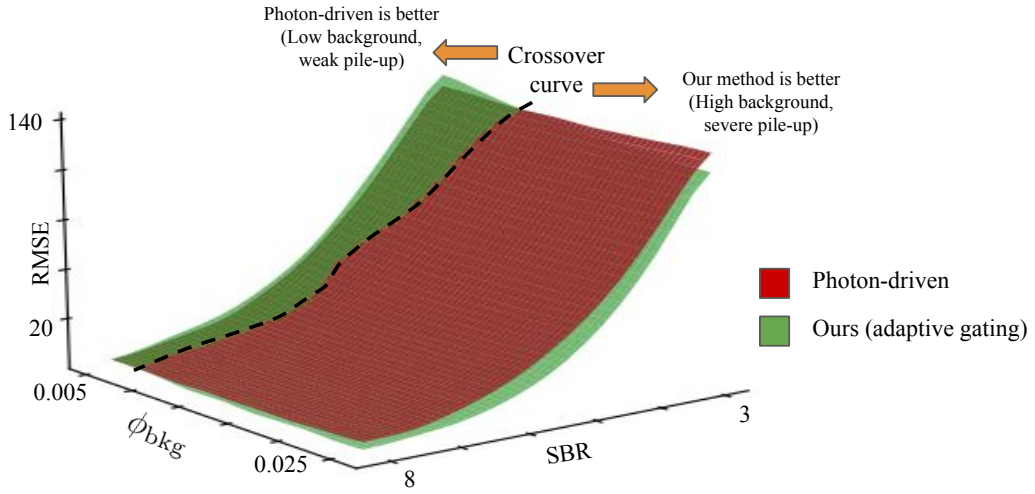


Supplementary Figure 11. **Depth reconstruction for "Horse" scene with corresponding error maps** Our method gives lower errors across all regions of the image.



Supplementary Figure 12. **Depth reconstruction of "Office" using monocular depth priors.** (c) and (d) show the depth estimates and corresponding uncertainties produced by the learning-based monocular depth technique of Xia et al. [8]. We use these results to construct a per-pixel Gaussian depth prior with mean equal to the depth estimates in (c), and variance equal to the uncertainties shown in (d). Using this prior with our adaptive gating scheme leads to significantly better depth reconstructions (e-f). When we additionally use adaptive exposure, using the depth prior also helps decrease total exposure time (g-h).

Supplementary Figures 10, 11, and 12 show additional visualizations for the experiments on the "Leaf", "Horse", and "Office" scenes we report in the main paper. We note that we use the indoor scene "Office", where pile-up artifacts are limited, to showcase the ability of our adaptive gating scheme to take advantage of priors from the monocular depth technique of Xia et al. [8]. Unfortunately, we were unable to obtain reliable depth and uncertainty estimates from the technique of Xia et al. [8] for outdoor scenes, likely because the neural network it uses is trained on a dataset of indoor scenes.



Supplementary Figure 13. **Performance comparison between adaptive gating and photon-driven acquisition** Adaptive gating consistently out-performs photon-driven acquisition except in scenarios of low background light. Adaptive gating facilitates better exploration of the depth range when pile-up effects are present, this explains the decrease in RMSE observed when background intensities are high. However, since optimal gate selection leads to additional SPAD dead-time, in low background light scenarios where pile-up effects are no longer observed, photon-driven acquisition performs better than adaptive gating.

References

[1] Anant Gupta, Atul Ingle, and Mohit Gupta. Asynchronous single-photon 3d imaging. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7908–7917, 2019. 1, 2, 3, 6

[2] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. Photon-flooded single-photon 3d cameras. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6763–6772, 2019. 1, 2

[3] Felix Heide, Steven Diamond, David B. Lindell, and Gordon Wetzstein. Sub-picosecond photon-efficient 3d imaging using single-photon sensors. *Scientific Reports*, 8, 2018. 1, 2

[4] Quercus Hernandez, Diego Gutierrez, and Adrián Jarabo. A computational model of a single-photon avalanche diode sensor for transient imaging. *ArXiv*, abs/1703.02635, 2017. 2

[5] Chao Liu, Jinwei Gu, Kihwan Kim, Srinivasa G Narasimhan, and Jan Kautz. Neural rgb (r) d sensing: Depth and uncertainty from a video camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10986–10995, 2019. 8

[6] Yaadhav Raaj, Siddharth Ancha, Robert Tamburo, David Held, and Srinivasa G Narasimhan. Exploiting & refining depth distributions with triangulation light curtains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7434–7442, 2021. 8

[7] Michael Wahl and Sandra Orthaus-Müller. Time tagged time-resolved fluorescence data collection in life sciences. *Technical Note. PicoQuant GmbH, Germany*, 2014. 2

[8] Zhihao Xia, Patrick Sullivan, and Ayan Chakrabarti. Generating and exploiting probabilistic monocular depth estimates. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 62–71, 2020. 8, 11