Adaptive SPAD Imaging with Depth Priors

Ryan Po
Carnegie Mellon University
rlpo@andrew.cmu.edu

Ioannis Gkioulekas
Carnegie Mellon University
igkioule@cs.cmu.edu

Abstract

Single-photon Avalanche Diodes (SPADs) have become increasingly popular in depth sensing due its high temporal resolution and photon efficiency. It is used in many applications such as fluorescence lifetime imaging, NLOS imaging and autonomous navigation. Especially in modern autonomous vehicles, SPAD based LiDARs are commonly part of a multi-sensor depth sensing pipeline that include other forms of depth-sensing equipment, i.e. stereo cameras. Under high ambient light, SPADs perform poorly when operating over a large range due to the 'pile-up' effect. This project aims at exploring how measurements from other sources can help establish priors on depth that can drastically increase the performance of SPADs by reducing the effect of pile-up in depth-sensing scenarios. Another key insight addresses why SPADs alone can not achieve a similar improvement in performance in the absence of such depth priors.

1. Introduction

Before diving into how we can utilize depth priors to achieve better SPAD performance for depth-sensing, it is important to understand the fundamental problems that lie in transient imaging for SPADs. Due to the single-photon sensing nature of SPADs, under strong ambient light, we run into the problem of 'pile-up'. This section serves as a brief introduction to the pile-up effect, and existing ways of compensating for this effect.

1.1. Pile-up Effect

For depth-sensing application of SPADs, a light source, usually a laser, is fired and the transient is recorded. The peak of the recovered transient is then treated as the time at which the laser pulse returns to our sensor from the source and thus depth can be inferred from time-of-flight. Under low light settings, the raw measurements of SPADs closely resemble the transient that we would expect, with photons only arriving in the time bin where the laser reaches the sensor. Under high ambient light however, the effect of

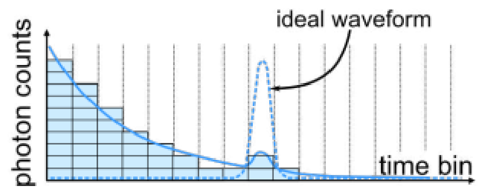

Figure 1. **Visualization of Pile-up Effect.** Early arriving photon causes the formation of skewed transient. The peak representing the laser pulse is buried in the tail of the transient. Figure courtesy of Gupta et al.

'pile-up' causes the measured transient to become skewed. This effect is caused by the fact that SPADs are saturated by the first incoming photon, therefore photons detection at later bins not only relies on the level of light at that bin but also on whether a photon was detected in an earlier time bin. In fact, the model of photon arrivals for SPADs is a well-studied topic, for example in [1], it has been proven empirically that photon arrivals can be roughly modelled as a poisson random variable. Then given a transient with B bins and ideal transient measurements of λ_i at each bin, the photon arrival rates for each bin with pile-up is:

$$\hat{\lambda_i} = e^{-\sum_{j=1}^{i-1} \lambda_i}(1 - e^{-\lambda_i}) \tag{1}$$

The above effect can also be visualised as shown in Figure 1, where we can see that the effect of pile-up causes the transient to become skewed, and the expected peak due to the laser pulse is buried within the tail of the measured transient.

1.2. Generalised Coates' Estimate

Fortunately, there are methods for recovering the true transient from the piled-up transient as explored in [1, 3, 2]. All of these techniques stem from one estimator as mentioned in [3], namely the generalised Coates' Estimator, which is in essence the maximum-likelihood estimate for the original transient. If we let D_i be the total number of photon detection opportunities and time bin i and N_i be the total number of photons detected at time bin i. Then an estimate, named q_i, for $\hat{\lambda}_i$ in equation (1) is as follows,

$$q_i = \frac{N_i}{D_i} \tag{2}$$

Then by taking the MLE, we have an estimator, named r_i, for the original transient measurement λ_i:

$$r_i = \log\left(\frac{1}{1-q_i}\right) \tag{3}$$

The above estimator is a great generalisation, since as long as the ideal transient does not change, we can combine multiple different SPAD measurements with different active times and shifted gates into one transient. A key takaway from the above estimator is that as D_i gets larger, we are essentially getting more information for that particular bin, therefore bins with more opportunities to detect photons, namely the bins closer to the beginning of the transient measured by the SPAD will get have more accurate recovered values. Our method will leverage this fact to get better measurements around the signal position to improve performance in depth recovery.

1.3. Existing Work

A brief mention on existing work that we will also incorporate in the methods described below.

1.3.1 Optimal Scalar Attenuation

Due to the pile-up effect, it is the standard to attenuate all light coming into the sensor so that a photon is detected during approximately only 5-10% of laser cycles. However in Gupta's work [2], it is proven that there is an optimal scaling factor for the transient that is based off the strength of ambient light and the number of bins. Throughout the rest of this project, this optimal attenuation factor will be applied whenever possible.

1.3.2 Asynchronous Acquisition

This method [3] is also used for compensating pile-up, where a shifted gate is utilized so that every bin gets the chance to be at the front of the transient, and merging multiple measurements of together to get an overall better measurement at each bin than synchronous acquisition. Later on we will discuss how we can utilise depth priors and apply it to asynchronous acquisition.

1.3.3 Comparison Baselines

An important note is that we will be using synchronous acquisition with optimal attenuation applied as the baseline for all comparisons. Therefore whenever the term 'conventional' is used, it is referring to synchronous acquisition with optimal attenuation. Although asynchronous acquisition performs better, the below comparisons are done where both conventional and proposed techniques use synchronous acquisition, since asynchronous acquisition can be applied whether or not a shifted gate is present. There is also an extra section addressing how asynchronous acquisition can be applied to the methods described below.

2. Depth Adaptive Gating

The first contribution of this project is known as depth adaptive gating. Essentially, by utilising depth priors of varying sources we can determine an optimal shift in the gate and shortening of the active time in order to achieve better results in depth recovery performance.

2.1. Method

The method itself is quite simple, given a depth measurement from a prior source, we establish a prior on the depth based on this depth estimate and find an optimal gate shift and active time. The general process is visualised in Figure 2. By shifting the gate and adjusting the active time of the SPAD sensor, we are essentially limiting the range of the SPAD. The positive effects of doing so are two-fold. Firstly, the shifting of the gate means that the signal is closer to the beginning of the transient measured by the SPAD, and as mentioned in section 1.1, this will lead to better measurements. The second effect is the decreased active time, the lower active time means that we are getting more laser cycles given the same integration time.

2.2. Sources of Depth Estimate

The initial depth estimate used to establish the depth priors can come from many different sources. Some methods include depth from stereo, NN-based monocular depth and depth from defocus. It is important to note that not all depth estimates can be used, the criteria for the depth estimate will be discussed in a later section. For this project, we have chosen to use depth from stereo and a probabilistic NN-based monocular depth method [5] to establish depth priors, the results of this will be discussed later.

2.3. Ideal Gating under Gaussian Prior

Before we experiment with how this method works with real measurements, it is interesting to look into what the optimal gating strategy is given certain assumption. To be precise, we decided to explore what the optimal gate shift would be if the errors from the depth estimates were gaussian distributed. We did a sweep of different active times and gating positions. For gaussian distributed depth errors of differing variances, we set set the gate so that the estimated depth is a scalar multiple C of the variance. The

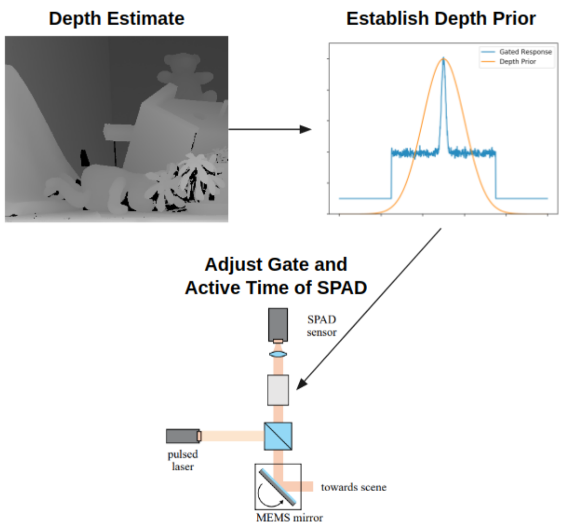

Figure 2. **Visualization of Method.** Depth estimate is used to establish depth prior from which the optimal shifting scheme is derived. This shifting scheme is then applied during the SPAD acquisition process.

active time of the SPAD is then set to be double of this distance from the beginning of the transient to the estimated depth. After doing a scan over different values of C, the best performing value was is used. A visualisation of the performance of varying values of C is shown in Figure 3.

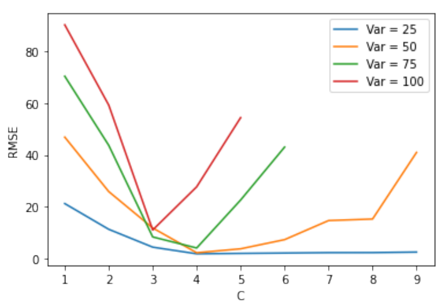

Figure 3. **RMSE of simulated transients for different values of C for estimators of differing variances**

2.3.1 Results

Under the gaussian error assumption, the results obtained work very well. In Figure 4, we show the results from sample renders given the gaussian depth error assumption, where the variance of the depth error is set at 75 bins and the integration time is set at 1 ms. With the depth prior, as shown in Figure 4, our proposed technique heavily outperforms the conventional technique without the adaptive gate. However, this behaviour is to be expected since a depth prior does provide more information.

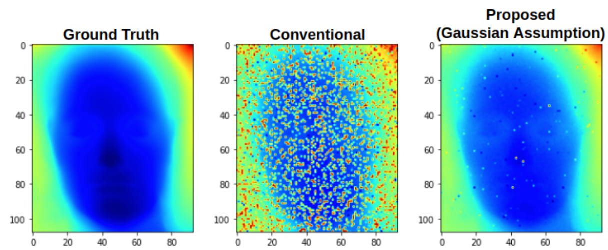

Figure 4. **Renders with Gaussian Error Assumption** Conventional SPAD acquisition techniques don't work as well over a large range, by reducing the range using a probabilistic approach we can get much better results.

There is one very interesting observation here, it is that the depth recovery of our proposed method with half the integration time still performs better than the conventional method, with an RMSE that is over 2 times lower. This sparks the question, what if we use the first half of the pulses to form a depth prior. Note the fact that the 75 bin variance is actually larger than the RMSE of the depth recovered through conventional methods, meaning the first half of the pulses would give us a better depth estimate than what we are currently using. Unfortunately, this depth estimation does not satisfy certain requirements which we will address in a later section.

2.4. Experiments with Different Depth Estimators

In reality, the errors for the depth estimators that can be used for are not exactly gaussian, but for some common and robust methods for depth recovery, we can observe that their errors do follow some gaussian-like distribution. In our experiments, we tried using two methods of depth recovery methods to establish our depth priors, depth from stereo and NN-based monocular depth.

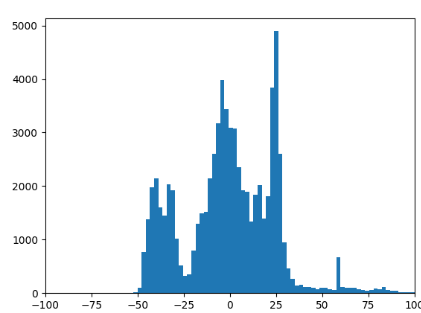

Figure 5. **Distribution of errors for depth from stereo**

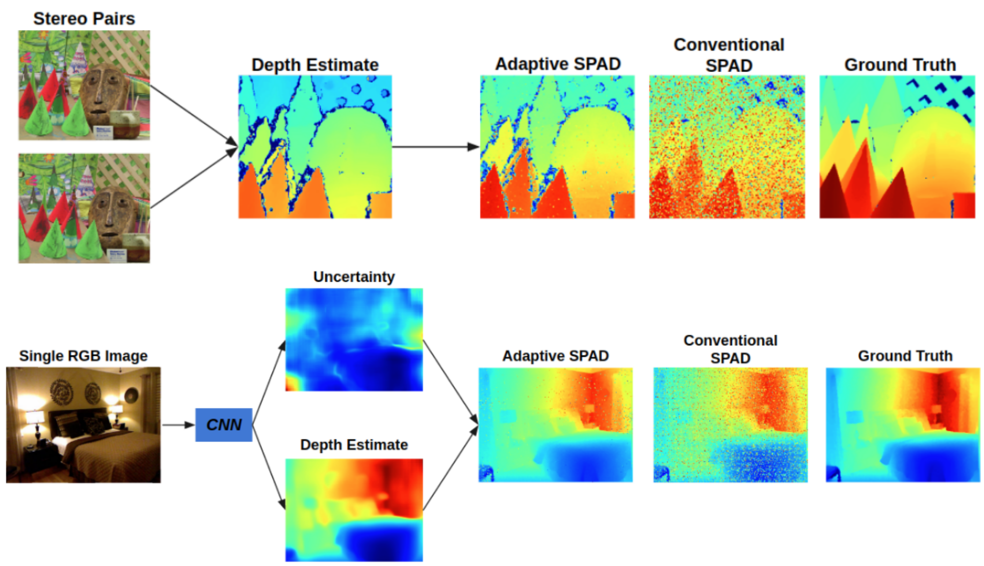

Figure 6. **Pipeline for stereo and monocular depth priors**

2.4.1 Depth from Stereo

Given two rectified images taken from the same plane, we can simply obtain depth from measuring the disparity of features. In terms of the distribution the errors for stereo depth, from Figure 5 we can observe that the errors roughly follows a zero-mean gaussian distribution. Therefore, we can apply the values for C that we derived in section 2.3. The resulting render can be seen on the upper right of Figure 6. Note that, wherever we get a reasonable depth estimate, the adaptive spad measurement manages to greatly reduce the noise that can be observed in the conventional SPAD renders.

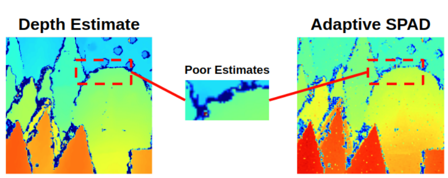

Figure 7. **Illustration of effects of bad priors**

However, in the above renders we some artifacts around the edge for our proposed method. In theory, an ideal shifted gate should never perform worse since at worst we can just leave the gate completely untouched. However issues occur when we get poor estimates of depth from our prior. In this case, this is caused by pixels that are occluded between the stereo image pairs. This problem could be circumvented if there was some measure of 'confidence' for our depth estimate, however in the case of depth from stereo, that is not available. This means that edge artifacts from stereo are then transferred to our SPAD measurements as shown in Figure 7.

2.4.2 NN-Based Single RGB Depth

With the progression of NN-based monocular depth-recovery, we don't even need stereo pairs to recover depth, instead we can just pass a single RGB image into a neural network to obtain reasonable depth estimates. There is also an added benefit in using a NN-based approach, the presence of a confidence metric. With the work of Xia et al. [5], which takes a probabilistic approach at NN-based monocular depth, not only can we recover an estimated depth map, but also an 'uncertainty' image as shown in Figure 6. Using this uncertainty matrix, we can adjust our active time accordingly.

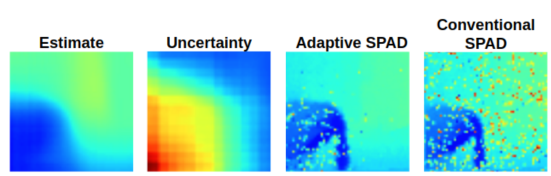

Figure 8. **Example of performance under high uncertainty**

As shown in Figure 8, despite having high uncertainty for the depth priors on the bottom left of the full image, the adaptive method will never perform worse than the conventional method.

3. Async. Acquisition with Depth Priors

As mentioned above, another method for compensating the effects of pile-up is by applying the method of asynchronous acquisition as proposed by Gupta et al. [3]. Without a prior on depth, asynchronous acquisition is done with a uniform shifting strategy as it is assumed that there is an equal likelihood that each bin is the signal bin. There are two ways we can apply asynchronous acquisition to the method above. The simplest way, and likely the most effective one, is to simply use the above method to reduce the range of the SPAD and then use asynchronous acquisition with uniform shifting to recover depth over this shortened range. Another method utilises the depth prior to assign different number of pulses to different shifting positions. Instead of a uniform shifting strategy that makes sure every time bin gets the be at the front of the transient the same amount of laser pulses, we instead distribute the laser pulses based on the depth prior, where we spend more laser cycles on bins we expect the signal to be at. In experimentation, both results show improvement, but the latter method performs quite similarly to asynchronous acquisition, with only a 10-20% gain in performance despite needing a depth prior.

4. Batch-based Adaptive SPAD

4.1. Motivation and Theory

Since the above methods utilise depth estimates derived from other sources, it is natural to consider whether we can utilise earlier pulses from our SPAD measurements to establish a prior on depth that we can use to adapt our gate during the acquisition process. We call this method batch-based or pulse-based adaptation, where we adapt the SPAD acquisition process every fixed number of pulses with the depth prior established using the previous pulses. Unfortunately this does not work, and the reason behind this paves way for defining a class of depth estimators that work well with our method.

4.2. Error Distribution of SPAD Estimate

In order to understand why batch-based adaptation does not work, we must understand the distribution of errors in SPAD-based depth recovery. We can estimate the probability of an error occurring at each time bin using the Chernoff Bound on gaussian random variables as mentioned in [2]. This Chernoff Bound is reliant on knowing the variance of the Coates' which we do not have a closed form value for, but we can utilise the CR-bound as derived by Adithya et al. [1], where the bound is empirically proven to be tight. The CR-bound for the Coates' Estimator is as follows,

$$Var(\hat{\lambda_i}) \geq \frac{1}{s}\frac{1-e^{-r_i}}{e^{\sum_{j=1}^{i} r_i}} \tag{4}$$

Where s is the total number of laser cycles that elapse during the integration time. There are 2 key takeaways from the above bound: 1. variance increases as we approach later time bins and 2. assuming constant ambient lighting, r_i will be roughly equal across all time bins other than at the signal bins (and a few bins surrounding the signal bin due to jitter). That means the variance of each bin will increase at a constant factor for most bins. This theory is supported by Figure 9. where the variance of the Coates' estimate is shown. You can see that other than the bins that are close to the signal bin, the rest of the bins have variances that are increasing in a linear fashion. Therefore we can conclude that the variances at most time bins is roughly independent of the signal position.

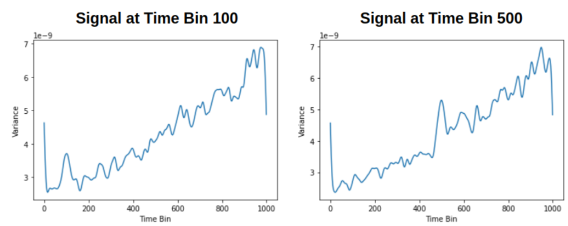

Figure 9. **Variance of Coates' Estimate at different time bins**

With a suitably tight bound for variance, we can apply the Chernoff bound on the probability of an error occurring at each time bin. As derived in [2],

$$\mathbb{P}(r_\tau < r_i) \leq \frac{1}{2}\left(-\frac{(\Phi_{sig}-\Phi_{bkg})^2}{2(\sigma_\tau^2+\sigma_i^2)}\right) \tag{5}$$

Where τ is the signal bin and Φ_{sig} and Φ_{bkg} are the laser intensity and background intensity respectively. Although the above is not precisely the probability that bin i is detected as the erroneous depth, it serves as a good enough estimate. Note that the above probability is inversely proportional to the variance of the Coates' estimate at time bin i, therefore we can say that the probability of a time bin causing an error in depth recovered increases as we get to later bins. Also note that in the above expression, the only value that changes across different time bins i is σ_i^2 the variance of the Coates' Estimate at time bin i. Since we established that variances at bins far from the signal bin is roughly independent of the signal position, this infers that the probability of any bin not close to the signal bin causing an error is also roughly independent of the signal position.

4.3. Why doesn't it work?

The key takeaway from the above analysis is that, other than bins that are very close to the signal bin, the probability of a bin causing an error is independent of the position of the signal bin. This means if we get an erroneous depth from our SPAD measurement that is far away from the true signal bin, we are given no information as to where the correct signal bin would be, meaning that we can not establish any priors based on this estimate. On the other hand, if we get an erroneous depth that is close to the signal bin, this would mostly be due to the effect of jitter, in which case setting up a prior based on this value won't give a lot of improvement anyway since our estimate is already close to the true signal bin.

4.4. Estimator Criteria for Depth Prior

Therefore we can conclude that a suitable estimator for establishing a prior must always provide information about the position of the true signal bin even when our estimate is far from the true signal bin. In the gaussian error assumption and probabilistic monocular depth cases, this is clearly true, since we are given information of 'how likely' our estimate is correct and we can shape our priors based on these likelihoods. For depth from stereo, we can see that for most points the errors follow some gaussian-like zero-mean distribution that we can model around. But in the case of SPADs, errors that are not close to the time bins provide zero information about the location of the true time bin and therefore does not meet the criteria.

5. Multi-sensor Fusion Pipelines

Given most depth sensing pipelines have more than one on-board sensor extensive amounts of work has gone into merging measurements from different sources to obtain better depth maps. An example of this would be the work done by Siddiqui et al. [4]. However, while they explore how to merge SPAD and stereo measurements, the final results of the merging depend greatly upon the raw measurements. The method described in our project aims at improving this raw measurement and will therefore also benefit the performance of such fusion pipelines.

6. Conclusions

To summarize, in this project we discussed the how we can utilise depth estimators from different sources to establish depth priors that motivate gate shifting schemes that can drastically increase SPAD performance in depth-sensing scenarios. We also discuss the necessary criteria that prior depth estimators must have in order to yield good results using our method.

References

[1] M. B. A. T. A. V. Adithya Kumar Pediredla, Aswin C. Sankaranarayanan. Signal processing based pile-up compensation for gated single-photon avalanche diodes, 2018.

[2] A. V. M. G. Anant Gupta, Atul Ingle. Photon-flooded single-photon 3d cameras, 2019. CVPR.

[3] M. G. Anant Gupta*, Atul Ingle*. Asynchronous single-photon 3d imaging, 2019. ICCV.

[4] M. O. Talha Ahmad Siddiqui, Rishi Madhok. An extensible multi-sensor fusion framework for 3d imaging, 2020. CVPR.

[5] A. C. Zhihao Xi, Patrick Sullivan. Generating and exploiting probabilistic monocular depth estimates, 2020. CVPR.