Photon-Limited Time of Flight Depth Acquisition: New Parametric Model and Its Analysis

by

Vahid Montazerhodjat

B.S., Electrical Engineering, Sharif University of Technology, 2010

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

February 2013

Signature of Author: _____

Vahid Montazerhodjat
Department of Electrical Engineering and Computer Science
January 28, 2013

Certified by: _____

Vivek K. Goyal
Research Scientist, Research Laboratory of Electronics
Thesis Supervisor

Accepted by: _____

Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Committee for Graduate Students

Photon-Limited Time of Flight Depth Acquisition:
New Parametric Model and Its Analysis

by Vahid Montazerhodjat

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science
in Electrical Engineering and Computer Science

Abstract

As 3-D imaging systems become more popular, the depth estimation which is their core component should be made as accurate as possible at low power levels. In this thesis, we consider the time of flight depth acquisition problem at low photon counts. We first formulate the received light intensity at the photodetector as the convolution of the source intensity and another signal which we call the scene impulse response. We then present a new framework to calculate the scene impulse response given the scene depth. Using that framework, we propose a richer parametric model than the model being used in existing technologies. We use the maximum likelihood algorithm with our proposed model on simulated data to estimate the depth parameters of the scene. The simulation results show a uniformly better performance for our model relative to the conventional model.

Thesis Supervisor: Vivek K. Goyal
Title: Research Scientist, Research Laboratory of Electronics

Acknowledgments

I'd like to thank my adviser, Professor Vivek Goyal for all the great things I've learned from him directly and indirectly since the first day I joined his research group. One of the things which make Vivek stand out among a lot of advisers is the space and freedom he provides his students with to grow as individual researchers. Vivek is also one of the nicest people I have ever seen which makes working with him even a better experience. Also, it is great to know you can be friends with your adviser outside academia, play games together, have group dinners, super-bowl parties, election party, go hiking and curling. I'd also like to express my gratitude to Professor Jeffrey Shapiro and his student, Nick, who have always kindly and patiently helped me get better insight into the optics part of the joint project we were working on. Also, I always enjoyed talking to Dr. Franco Wong and his student, Dheera, who helped me understand the real optical experiments and their challenges better.

I'd like to take this opportunity to thank Professor Devavrat Shah for being a great academic counselor and always giving very useful advice, Professor Bob Gallager for his incredible patience during my discussions with him about his book and for being a great teacher and teaching sophisticated topics using the simplest ideas in his course. Special thanks to Professor Laurent Damanet in the math department, Dr. Baris Erkmen of JPL and Professor Sundeep Rangan of NYU Poly for brief but useful discussions during these five semesters. Also, many thanks to Professors Lee, Kassakian, and Berggren for the fun experience we had as the staff of 6.002 in fall 2012. I'm also very grateful of Professor Jonathan Kelner who I had a very great discussion about considering my research problem as a moment estimation problem and what the possible approaches to solve that problem would be.

Special thanks to our group administrative assistants over the past two and half years, Eric, Kathryn, Gabrielle. Also, very special thanks to Anne Hunter in the EECS undergraduate office, Janet Fischer and Professor Leslie Kolodziejski of the EECS grad-

uate office.

I'd like to thank all my friends who have so far made my journey at MIT and Cambridge wonderful; especially a lot of thanks to Dan (well, I should say Dr. Weller) who is a role model for me and was very determined to teach me how to play tennis and he finally did it! Also, my awesome office-mate John has changed my perspective of life by quite a lot and has helped me have a better understanding of what good research is about. John's always been challenging me in many different ways and that makes me realize I have so many flaws to work on to improve myself both in and out of academia. Ulugbek, I do not even know how to thank you because you are extraordinarily awesome. We had a great time when you were here and I always wanted to hear your take on different topics both in academia and outside school. Amirali, you are one of the best and nicest persons I got to meet here and I am very grateful of you for your good advice and all your help during my first two years here. Also, many thanks to Hossein, Moojan, Darya, and Soheil for helping me settle in when I came to Cambridge. Guys, you have no idea how helpful all of you were in the first few months of my stay here. Soheil, I want to thank you very specifically for being a great friend in many different ways and I hope I can beat you one day in foosball (still working on it). Also, chatting with Da about financial markets and recently, how to get out of one's comfort zone to grow the dimensions of their lives is always exciting. It's also been a pleasant opportunity for me to teach two courses during IAP's here, one with Elena and the other with Lawson; I've learned quite a lot from both of you. Also, Elena, as I have always told you, the balance you've managed between your athletic and academic life is simply amazing. Also, I'd like to thank my uncle and his family for the great time we had together when I visited them this past thanksgiving. And, many thanks to George, Randi, Arman and Omid (two of the funniest people I know in Cambridge), Joong-Bum, Surat, Grace, Kirmani, Andrea, Jonathan, Weifei, Jason, Guillame, Andy, Shervin, Vahideh, Marzieh and Hamed, Yuji, Xue, Sefa, Ali, Azarakhsh, Mina, Dr. Woohyuk Chang, Sepideh and Jouya, Peyman, Arash, Saman, Sepideh, Henna, Tong, Shane, Jason, Becky, Arya and finally Kyu for the great parties he organized last year.

Last but not least, I'd like to thank my parents, sister and brother who have supported me in every moment of my life. I owe all the accomplishments and success I have had during my life to my parents who are the most precious persons of my life.

Contents

List of Figures

Chapter 1

Introduction

RECENTLY, 3-D imaging has played a crucial role in a lot of different areas like the gaming industry, gesture-based technologies, etc. To get a 3-D image, one has to estimate the depth of different parts of the scene whose 3-D image is desired. Therefore, an important component of 3-D imaging systems is the depth estimation block which should work as precisely as possible depending on the application. In order to measure depth, existing raster scanning time of flight technologies scan the scene by moving a laser in a transverse plane and illuminate each part of the scene, get the reflections from the scene back and using those data, they estimate the depth of the illuminated part. In the end, they put these estimated depths together to form a 3-D image. Therefore, for every illuminated patch, a depth estimation problem should be solved. Let us consider the scene as a single patch in the rest of this work. Unfortunately, the current technologies assume the most basic model for the scene; they assume the scene is a fronto-parallel plane to the reference plane. In other words, they assume there is no depth variation in the scene. Now, if the scene is not a fronto-parallel plane, the model the existing technologies use becomes an approximation to the true scene which may be significantly different from this approximation and the estimation performance becomes very poor in that case.

In this thesis, we first formulate the received light intensity at the photodetector as the convolution of the source intensity and another signal called the scene impulse response which depends on the scene depth. We then employ useful tools from real analysis to create a framework where one can calculate the scene impulse response given the scene depth. The proposed framework to calculate the scene impulse response is very general and works for any scene depth (linear, piecewise linear, nonlinear) assuming the depth is precisely known as a function of transverse coordinates. Using that framework, we first derive the model which is utilized in existing technologies (referred to as the conventional model) and then derive a new model which precisely models a larger set of

scenes than the conventional model does. The proposed model assumes the scene to be a tilted plane and if the plane tilt is zero, that model is exactly the same as assuming the scene to be a fronto-parallel plane (i.e., the conventional model). Therefore, our proposed model reduces to the conventional model when the true underlying scene has no depth variation in it. Our model can also be considered as a linear approximation to non-planar (nonlinear) scenes and consequently is better than a constant approximation (i.e., the conventional model). Therefore, the proposed model in this work is closer than the conventional model to the true underlying scene whether or not the scene is planar (linear).

Here, one might ask why considering the depth variation in the scene is important and why we cannot simply ignore that and approximate the depth as constant over the area of interest. In other words, why can we not approximate the scene as a fronto-parallel plane to the reference plane? The answer to this question is simple: In the depth estimation problem, if we assume there were no depth variation while there is a significant variation, we would be using a totally mismatched model for the received signal and as is shown in Chapter 4, this can degrade the estimation results by a large factor. This is mainly the reason behind this work; to improve the model accuracy and to avoid getting wrong results due to significant model mismatch.

Also, one might argue that the available raster scanning depth acquisition systems do not probably experience any large model mismatch. This is because the area they illuminate on the scene at one time (referred to as pixel) is very small and there cannot be a huge depth variation in that small area. This argument might be correct when it comes to existing depth acquisition technologies. However, the model we are introducing in this thesis contains the conventional model (i.e., constant depth) and if the conventional model is a good approximation, as shown in Chapter 4, using our model, we would get the same results as the existing depth acquisition systems do. However, it should be noted that our model is more complex than the conventional model and hence, it has a higher computation cost in the estimation part. On the other hand, knowing that in our model, we allow depth variation in each pixel, one might increase the area of the laser beam which leads to less number of pixels required for a specified transverse area and consequently reduces the time needed to scan the whole transverse area (scene). Therefore, by employing our model, measurement time can be reduced and that would compensate for the extra time we need in the estimation part. Furthermore, the model we introduce here can be easily utilized for the experiments where we flood light the scene without any performance degradation (as long as the light inten-

sity is uniform over the scene area) while obviously, for a large scene, assuming there is no depth variation in the scene (i.e., using the conventional model) fails almost surely. Also, the model for the received light intensity that we have introduced in this work can be applied to full-waveform LiDAR problems similar to what the authors of [14] have done (they however have considered mixture of Gaussians which is different from our model). Therefore, the model introduced in this work can be put into practice in many different ways.

Outline. In Chapter 2, we review the literature on parameter estimation from the arrival times of a non-stationary Poisson process. We then formulate the depth acquisition process and propose a general framework to calculate the scene impulse response in Chapter 3. In Chapter 4, the maximum likelihood algorithm and two classes for the scene impulse response will be used to estimate the parameters of the scene depth and numerical simulations will be presented to verify the performance of the estimation algorithm. Chapter 5 concludes the thesis and briefly discusses possible future work.

Chapter 2

Background

Since in this thesis, we consider the direct detection framework in order to measure light intensity, we begin this chapter by describing the idealized model for direct detection photodetectors (optical receivers) while explaining the notation employed throughout this thesis. We also review the literature on relevant problems like signal detection, and signal parameter estimation where, for the receiver end, the idealized model is considered. We then present a relatively more complicated model for direct detection photodetectors and review a few research works in that area. In the final section of this chapter, we briefly talk about the third generation of LiDAR systems and how this thesis is related to the models the existing literature considers for that problem.

■ 2.1 Idealized Model for Direct Detection Photodetectors

The schematic in Fig. 2.1 represents the abstract model for idealized direct detection photodetectors (optical receivers), especially when the received light power at the detector (the number of received photons) is relatively low [24, 26, 27]. In this block diagram, $\lambda_0(t) \geq 0$ is the squared magnitude of the electromagnetic field at the receiver plane at time t, integrated over the aperture area and then normalized by the energy of a photon, i.e., the normalized power received through the aperture area at time t. It is also worth mentioning the model of Fig. 2.1 is valid only when the electromagnetic field is single-mode. If there are different modes present in the transmitted laser beam, the output should be a Poisson mixture instead [31, 32]. The overall effect of background (ambient) light and dark count in the detector is denoted by $\gamma \geq 0$ which is assumed to be constant over time and independent of $\lambda_0(\cdot)$. We shall refer to γ as *background noise* in the rest of this thesis. Also, $N(\cdot)$, which is a staircase function, is a single realization of the non-stationary (inhomogeneous) Poisson counting process with time-varying rate $\lambda(\cdot) = \lambda_0(\cdot) + \gamma$. In other words, $N(t)$ is the cumulative number of photons received in

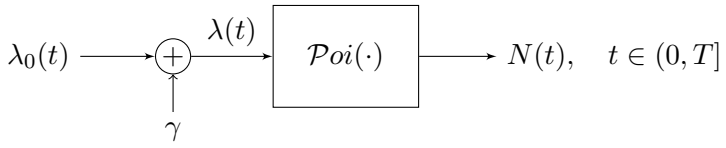


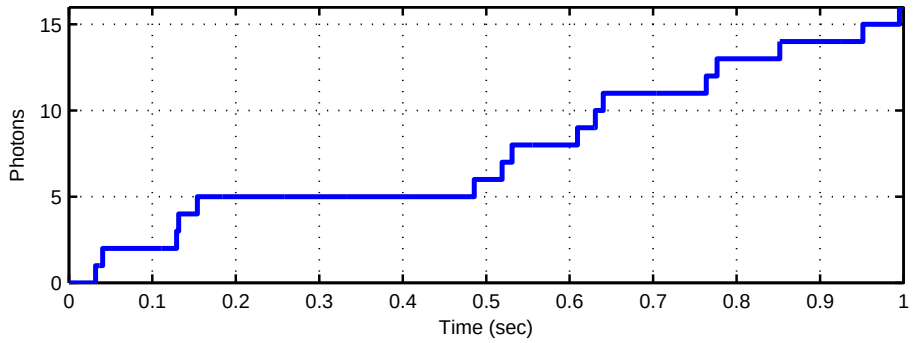
Figure 2.1. Block diagram for idealized direct detection photodetectors. The output is assumed to be an inhomogeneous Poisson counting process $N(\cdot)$ with rate $\lambda(\cdot)$ and to be observed only over the time interval $(0, T]$. The rate of the Poisson process, $\lambda(\cdot)$, is modeled as the sum of two independent terms: a constant term γ which represents the overall effect of ambient light and dark count of the photodetector and a second term $\lambda_0(\cdot)$ which models the normalized "signal" power received at the detector.

the time interval $(0, t]$. Throughout this work, we assume there is access to the output, $N(\cdot)$, only over the time interval $(0, T]$ unless otherwise specified. We also follow the convention and assume $N(0) = 0$ with probability one; i.e., there is no arrival before or at time $t = 0$ with probability one [10, 41]. As it is known in probability theory [10, 41], by utilizing the fact that it is continuous from right, the counting process can be completely specified by its associated point process (arrival process) which consists of the arrival times. An example for a sample path of a Poisson counting process along with its associated point process is presented in Fig. 2.2 where $\lambda(\cdot)$ over the time interval $[0, 1]$ is 15 photons per second. It is clearly observed that the discontinuity jumps in Fig. 2.2(a) (i.e., the arrivals in Fig. 2.2(b)) uniquely define the counting process. We are going to use this fact later in this chapter.

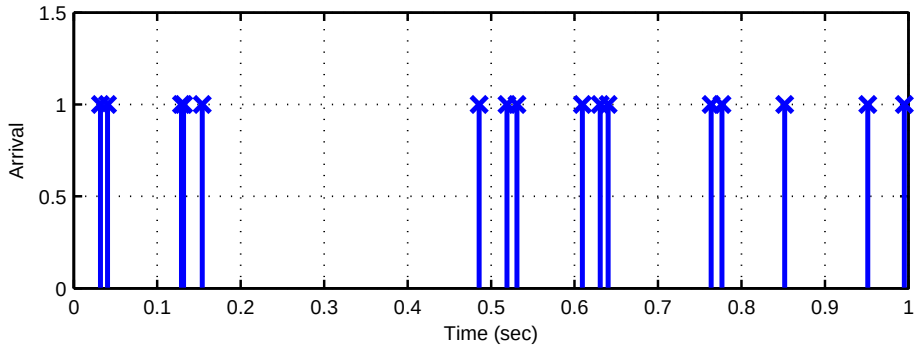
The block diagram of Fig. 2.1 was first introduced for the optical communications framework [24, 26, 27]. In that framework, $\lambda_0(\cdot)$ is the communication signal which carries information and hence, at the receiver, one should extract that information from the observations, i.e., the realization of the counting process. Based on how information has been embedded in the signal, the task of information extraction from observations is usually categorized into two distinct classes as follows:

- **Detection.** When the set of possible signals for $\lambda_0(\cdot)$ is countable, we need to "detect" the true underlying signal from the observations. For example, to send a bit, which can be 0 or 1, one can send nothing, i.e., $\lambda_0(\cdot) = 0$ or a constant positive signal, say, $\lambda_0(\cdot) = 1$, respectively. At the receiver end, we should detect whether $\lambda_0(\cdot) = 0$ or $\lambda_0(\cdot) = 1$ has been sent.

- **Estimation.** This scenario arises when all possible choices for $\lambda_0(\cdot)$ constitute an uncountably infinite set and the true underlying rate needs to be "estimated." As an example, suppose the receiver knows that the signal $\lambda_0(\cdot)$ is a Gaussian pulse shape with unit amplitude but the time shift of the Gaussian bump is not known



(a) An example of a Poisson counting process.



(b) The associated point process to the Poisson counting process shown above.

Figure 2.2. (a) A realization of cumulative number of photons (the Poisson counting process) and (b) its associated point process over time interval $(0, 1]$ where $\lambda(t) = 15$.

and has to be estimated from the observations.

In the previous example for the estimation problem, the set of all feasible solutions is a parametric set; i.e., the candidates for the signal have a known deterministic parametric form and the parameters themselves are unknown and to be estimated. This is a special case of what is called *parametric statistical inference* where the set of feasible solutions is parametrized and the estimation algorithm estimates the parameters in order to estimate the signal. On the other hand, for some problems, the possible solutions are not modeled as parametric signals and only some general properties about them are known, e.g., the signals are smooth, integrable, etc. This is an example of *non-parametric statistical inference* which is out of the scope of this thesis. We are going to model the problem in this thesis as a parametric problem and hence, parametric statistical inference framework shall be the core of this work.

When the parametric inference framework is employed, there are two possible cases for the parameters; they can either be constant or evolve over time. Because the considered problem in this thesis has parameters which are constant over time, in the rest of this section, we will talk about different research papers on the inference of parametric signals with time-invariant parameters from Poisson observations.

The research papers on signal detection [1, 16, 18, 38] and the estimation of signal parameters [5] from a single realization of the Poisson counting process in Fig. 2.1 date back to the 1960s. Utilizing the independence between the number of arrivals in non-overlapping time intervals in the Poisson counting process, $N(\cdot)$, the authors in [1, 16, 18, 38] devised different detection algorithms for unknown deterministic piecewise constant signals. They first divided the time interval $(0, T]$ into L bins, each of length $\frac{T}{L}$, assumed the signal $\lambda_0(\cdot)$ to be constant over each interval and then worked with the number of arrivals in each bin as the observables. From probability theory, those random variables are known to be independent and Poisson distributed [10, 41] and this fact probably was one of the main reasons the researchers wanted to work with the Poisson counts [5]. However, because they used the notion of binning the time axis and counting the number of arrivals in those bins, their algorithms were optimal only for piecewise constant signals and suboptimal for arbitrary positive signals in general[1] [5, 44]. On the other hand, Israel Bar-David worked with the Poisson arrival times themselves rather than the number of arrivals in the aforementioned bins. As was stated earlier in this chapter, the arrival times uniquely define the counting process $N(\cdot)$ [10, 41] and hence, working with the time of arrivals results in no information loss as opposed to working with Poisson counts in disjoint time bins. Intuitively speaking, there is always a finite number of arrivals in the interval $(0, T]$ and this makes working with the arrival times easier relative to working with the number of counts in disjoint intervals. This is due to the fact that a complete description of the sample space of the counting process requires infinitely many partitions for the interval $(0, T]$ [5] and that makes working with Poisson counts impractical if one wants to capture the whole sample space. On the other hand, if one wants to work with a specific partitioning of that interval, as stated above, the resulting algorithm will be suboptimal.

In 1969, in his seminal paper titled "Communication under the Poisson Regime,"

[1] In the estimation problem, one of the main issues regarding working with the Poisson counts in given time bins is that if the Poisson counts are not sufficient statistics for the to-be-estimated parameter, some information is lost due to binning [42]. For example, for the estimation of a parameter like the time shift of a known pulse shape, the algorithm accuracy would clearly be lower bounded by half of the length of each time bin [5].

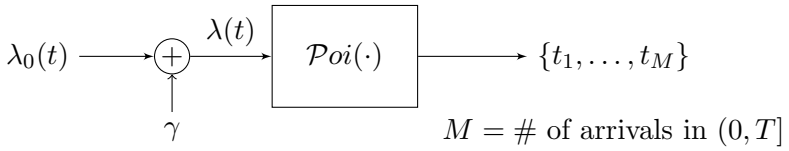


Figure 2.3. Because of the one-to-one correspondence between the counting process and its associated point process, considering arrival times $\{t_1, \ldots, t_M\}$ as the output of photodetector yields an equivalent model to the original model which has the counting process at its output.

Israel Bar-David considered an equivalent model to the original model of Fig. 2.1 for idealized direct detection photodetectors. The substitute model proposed by Bar-David is shown in Fig. 2.3 in which he considered the exact arrival times as the output of photodetector. We are going to refer to the model of Fig. 2.3 as the idealized model of photodetector in the rest of this thesis (it is called idealized because of the exact arrival times at the output of detector).

Bar-David derived the optimal detector for the detection problem where using the arrival times at the output of photodetector, one decides which signal $\lambda_0^{(1)}(\cdot)$ or $\lambda_0^{(2)}(\cdot)$ has been sent ($\lambda_0^{(1)}(\cdot)$ and $\lambda_0^{(2)}(\cdot)$ are two known arbitrary positive signals) [5]. He also addressed the estimation of the peak amplitude and time shift (delay) for a known pulse shape in a maximum likelihood framework and introduced a matched filter to estimate the time shift of the pulse from the Poisson arrival times at the output of direct detection optical receiver [5]. According to Verdú [44], Bar-David was the first person who had been able to find the optimal detector where the rate function was allowed to be an arbitrary positive signal and optimal delay estimator for deterministic signals with known shapes.

In [5], the author used the following three basic axioms for the Poisson process to derive the probability density function (pdf) of the Poisson arrival times. It should be noted that Pr(event name or description) stands for the probability of that event and $o(x)$ denotes any quantity that satisfies $\lim_{x\to 0} \frac{o(x)}{x} = 0$. Also, throughout this thesis, whenever there is a Poisson process, its rate is assumed to be $\lambda(\cdot)$ unless otherwise specified.

1. $\Pr\left(\text{one arrival in } (t, t+\Delta t]\right) = \lambda(t)\Delta t + o(\Delta t)$, as $\Delta t \to 0^+$.

2. $\Pr\left(\text{more than one arrival in } (t, t+\Delta t]\right) = o(\Delta t)$, as $\Delta t \to 0^+$; i.e., this probability goes to zero faster than Δt.

3. The numbers of arrivals in disjoint intervals are independent.

Using the above three axioms, it is possible to show that the number of arrivals in any interval is a Poisson random variable whose mean is equal to the integral of $\lambda(\cdot)$ over that interval and hence, numbers of arrivals in disjoint intervals are independent Poisson random variables [10]. Now, we are going to use these two facts to derive the probability density function of the arrival times $\{t_1, \ldots, t_M\}$[2], where M, number of arrivals in the interval $(0, T]$, itself is a random variable. Let us first define the event A as the following:

$$A = \text{exactly one arrival in each interval } (t_i - \Delta t_i, t_i], \quad i = 1, \ldots, M, \quad M > 0,$$
$$\text{and no arrival everywhere else in } (0, T],$$

where Δt_i's are positive and small enough that there is no overlap between any two consecutive intervals $(t_i - \Delta t_i, t_i]$ and $(t_{i+1} - \Delta t_{i+1}, t_{i+1}]$. Consequently, A can be interpreted as the intersection of $M+1$ independent events (these events are the number of arrivals in disjoint intervals and hence, using the third axiom, the events are independent). Therefore, probability of the occurrence of event A is simply the product of the probability of those $M+1$ independent events and we get

$$\begin{aligned}\Pr(A) &= \exp\left(-\sum_{i=0}^{M}\int_{t_i}^{t_{i+1}-\Delta t_{i+1}} \lambda(t)\mathrm{d}t\right)\prod_{i=1}^{M}\left(e^{-\int_{t_i-\Delta t_i}^{t_i}\lambda(t)\mathrm{d}t}\int_{t_i-\Delta t_i}^{t_i}\lambda(t)\mathrm{d}t\right)\\ &= \exp\left(-\int_0^T \lambda(t)\mathrm{d}t\right)\prod_{i=1}^{M}\left(\int_{t_i-\Delta t_i}^{t_i}\lambda(t)\mathrm{d}t\right), \quad M > 0, \end{aligned} \tag{2.1}$$

where $t_0 = 0$ and $t_{M+1} = T$. Therefore, we can use

$$p\left(\{t_i\}_{i=1}^{M}\right) = \lim_{\substack{\Delta t_i \to 0 \\ i=1,\ldots,M}} \left\{\Pr(A)\prod_{i=1}^{M}(\Delta t_i)^{-1}\right\}, \quad M > 0,$$

to derive the pdf of arrival times, $p\left(\{t_i\}_{i=1}^{M}\right)$, as the following (note that we use lower case letters for pdfs throughout this work):

$$p\left(\{t_i\}_{i=1}^{M}\right) = e^{-E}\prod_{i=1}^{M}\lambda(t_i), \quad M > 0, \tag{2.2}$$

[2]Without loss of generality, throughout this thesis, we assume arrival times are in ascending order, i.e., $0 < t_1 < t_2 < \cdots < t_M \leq T$. Also, we always define $t_0 = 0$.

where

$$E = \int_0^T \lambda(t)\mathrm{d}t \tag{2.3}$$

represents the expected number of photons over the whole observation interval, $(0, T]$. For the case where $M = 0$, which means no arrival, the probability is equal to e^{-E}. Now, let us assume the following parametric form for the rate of the Poisson process:

$$\lambda(t) = \gamma + \lambda_0(t; \alpha), \tag{2.4}$$

where $\lambda_0(t; \alpha)$ is a deterministic function of both t and α with a completely known form. t is time and $\alpha \in \mathbb{R}^n$ is a vector of unknown parameters and has to be estimated. Also, γ represents the background noise which is constant over time and known. Let us plug (2.4) into (2.2) and (2.3) to get the pdf of arrival times as the following:

$$p\left(\{t_i\}_{i=1}^M ; \alpha\right) = e^{-E(\alpha)} \prod_{i=1}^M (\gamma + \lambda_0(t_i; \alpha)), \quad M > 0, \tag{2.5}$$

where $E(\alpha) = \int_0^T \gamma + \lambda_0(t; \alpha)\mathrm{d}t$ represents the normalized energy (with unit of photons) received at the detector over the observation interval. Equivalently, $E(\alpha)$ is the expected total number of photons over the observation interval.

If one employs the maximum likelihood framework, estimating α is equivalent to solving the following mathematical problem:

$$\begin{aligned}
\widehat{\alpha} &= \underset{\alpha}{\operatorname{argmax}}\, p\left(\{t_i\}_{i=1}^M ; \alpha\right) = \underset{\alpha}{\operatorname{argmax}} \left\{\log p\left(\{t_i\}_{i=1}^M ; \alpha\right)\right\} \\
&= \underset{\alpha}{\operatorname{argmax}} \log \left(e^{-E(\alpha)} \prod_{i=1}^M (\gamma + \lambda_0(t_i; \alpha)) \right) \\
&= \underset{\alpha}{\operatorname{argmax}} \left\{ -E(\alpha) + \sum_{i=1}^M \log\left(\gamma + \lambda_0(t_i; \alpha)\right) \right\}, \quad M > 0.
\end{aligned} \tag{2.6}$$

Now, if we assume the received normalized energy $E(\alpha)$ does not vary with α, the first term in (2.6) is therefore constant with respect to α and can be removed from the optimization function. This is usually the case for ranging problems considered in the literature as we are going to see later in this chapter [20]. Also, as is stated in Chapter 4, in the depth acquisition problem we consider in this work, the received normalized

energy $E(\alpha)$ indeed does not vary with α. Now, by removing $E(\alpha)$ from (2.6), we get

$$\widehat{\alpha} = \underset{\alpha}{\operatorname{argmax}} \left\{ \sum_{i=1}^{M} \log\left(\gamma + \lambda_0(t_i;\alpha)\right) \right\}, \quad M > 0. \tag{2.7}$$

Now, one can be more specific and consider α as the time shift of $\lambda_0(t)$ as Bar-David did in [5]. To be more precise, let us assume $\lambda_0(t;\tau_0) = s(t-\tau_0)$, where $s(t)$ is a known positive signal and τ_0 is unknown. If we assume the feasible range for τ_0 is such that the integral $\int_0^T s(t-\tau_0)\mathrm{d}t$ does not change with varying τ_0 in that feasible set, we will get the same problem as in (2.7). This is the case where $s(t)$ is a short pulse (its support is much smaller than T) and the feasible range for τ_0 is small enough that $s(t-\tau_0)$ is zero outside the time interval $[0,T]$. Let us rewrite the optimization problem of (2.7) for this case as the following:

$$\widehat{\tau} = \underset{\tau}{\operatorname{argmax}} \left\{ \sum_{i=1}^{M} \log\left(\gamma + s(t_i-\tau)\right) \right\}, \quad M > 0. \tag{2.8}$$

The author of [5] considered this problem (which we refer to as the time shift estimation problem), assumed $s(t)$ to be a differentiable short pulse and derived the mean square error of the time delay estimate as the following

$$\mathrm{MSE}_\tau \triangleq E[(\widehat{\tau}-\tau_0)^2] = \left[\int_0^T \frac{[\dot{s}(t)]^2}{\gamma + s(t)} \mathrm{d}t \right]^{-1}, \tag{2.9}$$

where $\dot{s}(t)$ denotes the derivative of $s(t)$ with respect to t and $E[\cdot]$ is the expectation over observations t_i's. Now, because $s(t)$ is a positive signal, one can write $s(t) = E_s a^2(t)$, where $E_s = \int_0^T s(t)\mathrm{d}t$ is the expected number of "signal" photons due to $s(t)$, and $a(t)$ is a nonnegative signal satisfying the constraint $\int_0^T a^2(t)\mathrm{d}t = 1$. It should be noted that $a(t)$ has information about the shape of the signal and E_s carries information about the signal energy [5]. By plugging $s(t) = E_s a^2(t)$ into (2.9) and after some simplifications, for sufficiently large signal to noise ratio $\frac{E_s}{\gamma}$, we get

$$\mathrm{MSE}_\tau = \frac{1}{4} E_s^{-1} W_a^{-2}, \tag{2.10}$$

where W_a is the effective bandwidth of $a(t)$ in the frequency domain [5]; i.e., if we define $A(\omega)$ as the Fourier transform of $a(t)$, $W_a \triangleq \int_{-\infty}^{\infty} \omega^2 |A(\omega)|^2 \frac{\mathrm{d}\omega}{2\pi}$. As Bar-David pointed out in his paper [5], (2.10) yields nonzero MSE_τ even for infinite signal to noise

ratio ($\gamma = 0$) if E_s is finite (this is not the case for the time delay estimation problem when the noise is Gaussian). For the Poisson noise, we get nonzero MSE_τ for finite E_s because of the random nature of the Poisson process (even with zero background noise, the arrival times are random as opposed to noiseless observations in the Gaussian noise scenario) [5]. Also, another important result we can get from (2.10) is that if the pulse $s(t)$ is differentiable, the mean square error of the time shift estimate, MSE_τ, decreases with E_s^{-1} for large SNRs [5, 8].

In 1971, Bar-David used the relation for MSE_τ given in (2.10), in another paper [6], to calculate a lower bound on the capacity of a continuous time pulse position modulation (PPM) channel where the information is encoded in the continuous time shift of the transmitted pulse. We can assume the random variable τ to be the true underlying time shift of the pulse and $\widehat{\tau}$ to be its ML estimate. Now, we can write $\widehat{\tau} = \tau + \epsilon$, where ϵ can be considered as additive noise and is independent of the channel input τ. Then, we can use the rate distortion theory [9] and the fact that for all random variables with the same variance, Gaussian random variable has the maximum entropy to derive a lower bound for the channel capacity [6]. Comparing the lower bound derived in [6] to the bounds for discrete time PPM, we can conclude the continuous time PPM has an advantage over the discrete time PPM.

Israel Bar-David in 1975 found interesting results for the minimum mean square error in the time shift estimation problem with no background noise [8]. He considered three different types of signals for $s(t)$ (rectangular, exponential and Gaussian) and a uniform prior distribution for the time shift τ. In that setting, he then derived the minimum mean square error (mmse) time delay estimators for those three pulse shapes. According to his results [8], when $s(t)$ is a rectangular pulse, the mid range of arrival times is the mmse estimator, i.e., $\widehat{\tau}_{\text{mmse}} = \frac{1}{2}(t_1 + t_M)$ and for large enough signal energies (i.e., large E_s), minimum mean square error for the time shift (MMSE_τ) decreases with E_s^{-2}. If $s(t)$ is an exponential pulse (e.g., $s(t) = \frac{E_s}{D}\exp(-\frac{t}{D})u(t)$ where $u(t)$ is the unit step function), the mmse estimator for the time delay has the form of $\widehat{\tau}_{\text{mmse}} = t_1 - \frac{D}{M}$ and as in the rectangular pulse case, MMSE_τ is proportional to E_s^{-2} for large enough E_s. Finally, for the Gaussian pulse, the mmse estimator of the time shift is the center of gravity of the arrival times, i.e., $\widehat{\tau}_{\text{mmse}} = \frac{1}{M}\sum_{i=1}^{M} t_i$ and in this case, the minimum mean square error decreases only with E_s^{-1} as we saw earlier in (2.10) (note that the Gaussian pulse is differentiable). The author in [8] then concluded that in the time shift estimation problem without any background noise for pulse shapes with sharp discontinuities like rectangular or exponential signals, MMSE_τ has a sharper decay

($\propto E_s^{-2}$) than for differentiable pulses like Gaussian family where $\text{MMSE}_\tau \propto E_s^{-1}$ (for large enough signal energies E_s). It should also be noted that the exponential signal has frequently been used in the literature [8, 15, 41] because in the medical applications, the bio-signals can be well modeled as exponential signals; however, in this thesis, we do not consider those type of signals.

In 1975, parallel to Bar-David's research on the performance characterization of ML estimate for the time shift of a signal with known shape, Misra and Sorenson in [34] derived lower bounds for the mean square error of pulse position estimate where they considered both a box function and a step function for the signal $s(t)$. They used the general Ziv-Zakai lower bounds which had been derived in [46]. The lower bounds derived in [34] were tighter than the bounds Bar-David calculated in [7, 8] only for the low E_s region. Also, Hero in 1989 used the notion of entropy power and other similar information theoretic ideas as Bar-David used in [7] to derive possibly tighter lower bounds on the estimator performance where the energy of the signal was assumed to be independent of the parameters [20] (this is what we earlier assumed about $E(\alpha)$ being independent of α). Hero's lower bounds for the MSE of time shift estimate were tighter than the bounds in [34] and [8] for the low SNR region while Bar-David's lower bounds were optimal in the high SNR region among all derived bounds [7, 8, 20, 34]. Also, recently, the authors in [17] have proposed a model to predict the performance of the ML estimate for the time shift estimation problem where they have considered a generalized Gaussian model for the signal $s(t)$. Their model predicts the performance quite well even for low SNR's where the Cramér-Rao lower bound is not tight.

Up until now in this section, we have reviewed different research papers in which the set of parameters in the rate of Poisson process are assumed to be time-invariant. However, the parameters of the rate can be time-varying as Snyder in his paper [40] considered a Markov process over time for the rate of the Poisson process. These Poisson processes which are driven by another stochastic process are called doubly stochastic Poisson processes [33, 40, 41]. Snyder's paper started a large research arena on doubly stochastic Poisson processes where researchers considered different types of stochastic processes for the rate of the Poisson process.

There also is an extensive amount of research on pulse position modulation and pulse amplitude modulation for photon-limited optical communications specifically from the information theoretical point of view (cf. [44] and references therein). There are a couple of papers on non-parametric estimation of the rate of Poisson processes as well (cf. [2] and references therein). However, we are not going to consider these topics

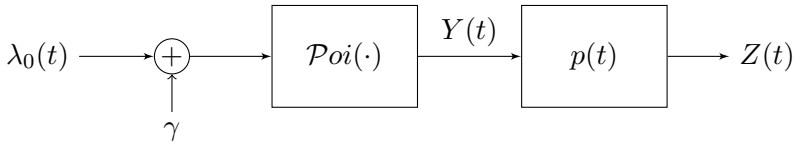


Figure 2.4. Shot noise model for photodetector.

because they are out of the scope of this thesis.

■ 2.2 Shot Noise Model for Direct Detection Photodetectors

In the previous section, we assumed the output of direct detection photodetectors to be the exact photon arrival times as shown in Fig. 2.3. However, because detectors have finite electrical bandwidth, we cannot have a train of impulses (each impulse located at one of the arrival times) at the output of detector. A more accurate and more practical model can be constructed by feeding that impulse train into a time invariant filter with a finite bandwidth as shown in Fig. 2.4 [4]. $p(t)$ represents that filter in the time domain and its bandwidth in the frequency domain is the same as the electrical bandwidth of the detector. $p(t)$ is also the envelope we would see at the detector output for each single photon arrival. $Y(t) = \sum_{i=1}^{M} \delta(t - t_i)$, where $\delta(\cdot)$ is the Dirac delta function. Also, we have $Z(t) = Y(t) * p(t)$, where $*$ denotes the convolution operator and because $Y(t)$ is an impulse train, we get $Z(t) = \sum_{i=1}^{M} p(t - t_i)$, where $Z(t)$ is called shot noise process [4]. Now, if the rate of the Poisson process is low enough, the spacing between t_i's will be large and therefore, one can get the exact arrival times by simply detecting the peaks of $Z(t)$. However, it is a different story when the Poisson rate is large, because in that case, t_i's tend to be close to each other and therefore, there will be overlap between adjacent pulses in $Z(t)$ and peak detection algorithms will fail most likely [21]. In the following part, we add another layer to our shot noise model for the detector and very briefly list a handful of research papers considering that model.

■ 2.2.1 Shot Noise Process Embedded in Gaussian Noise

The last modification to our model for a direct detection photodetector is adding white Gaussian noise independent of shot noise process to the output of detector as depicted in Fig. 2.5. $W(t)$ denotes white Gaussian noise and $X(t) = Z(t) + W(t)$ is the process we get to observe. This model complements the model of Fig. 2.4 by taking into account the existence of Gaussian noise at the output.

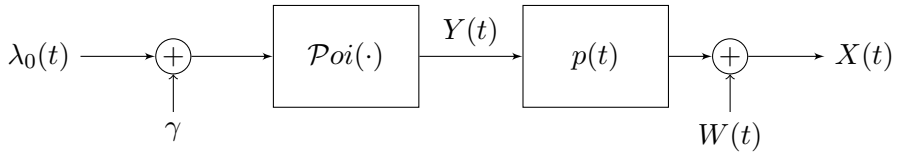


Figure 2.5. Shot noise process embedded in Gaussian noise.

Alfred Hero in his paper [21] titled "Timing Estimation for a Filtered Poisson Process in Gaussian Noise" investigated the problem of time shift estimation using the observations of Fig. 2.5. He again used information theoretic concepts to relate channel capacity and rate distortion theory to the mean square error in the estimation problem and derived lower bounds which matched Cramér-Rao lower bounds for high SNR's and were tighter than Cramér-Rao lower bounds for low SNR's.

There is also another paper written in 1990 where the authors considered the time shift estimation problem using the observations in Fig. 2.5 and compared the performance of a few linear and nonlinear estimators [22]. Then in 1994, instead of only theoretically analyzing the lower bounds on the estimator performance, Antoniadis and Hero proposed an efficient and approximate expectation-maximization (EM) algorithm to estimate the time-delay of the pulse in a filtered Poisson process in Gaussian noise [3]. Considering the exact arrival times as the hidden variables, the authors in [3] first formulated the exact EM algorithm. However, the expectation step turned out to be hard to evaluate and therefore, they approximated that expectation with a linear function which was easy to compute.

■ 2.3 Full-Waveform LiDAR

One of the most commonly used methods to acquire depth is the light detection and ranging (LiDAR) technique [11–14, 30, 36, 39, 45] where the scene is scanned by moving a laser source in a transverse area. Then, each illuminated point on the scene is considered as a pixel and using the returned (received) reflections at the photodetector, the distance of the pixel from the detector plane is estimated. In the end, the depths of different pixels are put together to create an image which is called the depth map of the scene (the depth map is precisely defined in the next chapter). Now, if the laser source is a very short pulse in the time domain, the received signal (light intensity) at the photodetector is a mixture signal which has a component (nearly) constant over time because of the presence of ambient light and also exhibits one or more peaks due

to the reflections of light from the object(s) present in the illuminated pixel. In the first two generations of LiDAR systems, the main goal was only to estimate a single range/ depth for each illuminated part (pixel). However, in the third generation LiDAR systems, in order to extract more information about the scene, there has been quite a lot of interest in using the whole received signal which is called the full-waveform in the literature [12–14, 30, 36, 45]. The LiDAR techniques which use the full-waveform at the detector are subsequently called full-waveform LiDAR methods [12–14, 30, 36, 45]. One major issue which comes with using the full-waveforms at the photodetector is the storage and transmission cost of such those signals. Therefore, extensive research has been carried out to efficiently compress the full-waveform [12–14, 30]. One method is to model the full-waveform as a parametric signal such that one can store and/or transmit the parameters of the received signal instead of the whole signal itself [12–14, 30].

The authors of [12, 13] have considered a sum of Dirac delta functions as the model for the received signal when the illuminating laser source has been assumed to be an impulse in time. Then, they have used finite rate of innovation (FRI) techniques along with Cadsow's denoising algorithm to estimate the locations of impulses in the full-waveform. As we are going to discuss in Chapter 3, the sum of Dirac delta functions corresponds to the scenario where the scene is assumed to be comprised of only fronto-parallel planes to the detector plane. Therefore, the estimation technique in [12] suffers from model mismatch if the true underlying scene is significantly different from the fronto-parallel assumption. However, as the authors of [12, 13] have noted, their model is one of the simplest models among the others in the literature because an impulse mixture requires fewer parameters than the other existing models in the literature.

Another commonly used model for the full-waveform is a mixture of Gaussian model which the authors of [11, 23, 30, 36] have considered. As it is stated in [30], sum of Gaussian signals is a pretty good model for most of mapping applications in urban areas. However, the Gaussian approximation always results in symmetric peaks which may not match the asymmetric peaks seen in real full-waveform LiDAR data. Therefore, there have been other papers where even more complicated models have been investigated for the same problem [14]. The authors of [14] considered a mixture of log-normal signals and also a mixture of generalized Gaussian signals for the full-waveform at the detector and showed improved performance in comparison to the Gaussian mixture model for the received signal. However, as was stated earlier in this section, these methods suffer from having too many parameters and hence, high computational complexity.

It should be noted that the proposed model of Chapter 3 can be considered as

another model for the received signal in full-waveform LiDAR systems. Also, we should mention that up to the best of our knowledge, the existing literature on full-waveform LiDAR has not considered the low-photon scenario while in this thesis, in addition to proposing a new model for the full-waveform, we have investigated the low-photon situation.

Now, after having reviewed the literature on similar problems as the problem considered in this thesis, we are going to introduce the model for our experiment and formulate the problem in the next chapter.

Chapter 3

Depth Acquisition Model

In this chapter, we first define depth function and depth map in Section 3.1 and then, in Section 3.2, we introduce a toy problem which sheds light on the time of flight depth acquisition problem. Then, in Section 3.3, we add different layers to that simple problem to make it more realistic and practical. The core of one of those layers is the impulse response of the scene which is defined as the normalized power received at the detector, due to an impulse illumination in time, and is modeled as a parametric signal in Section 3.4. Having known all those layers, it is straightforward to derive a relation between the scene depth (the scene impulse response) and the measurements at the output of photodetector. This relation is going to serve as the foundation of the inference algorithm proposed in Chapter 4 to estimate the scene depth. In the last section of this chapter, we briefly talk about how to modify the model for the scene impulse response to address depth acquisition with high range uncertainty.

■ 3.1 Depth Function and Depth Map

When it comes to depth acquisition in three-dimensional space, first, a reference plane is required from which the perpendicular distance of each point on the scene is measured. Let us define this reference plane as the x-y plane (equivalently, the $z = 0$ plane). Consequently, for each point with coordinates (x, y, z) on the scene, the associated depth is z assuming z is non-negative. Also, the scene is assumed to be nontransparent; therefore, if there are overlapping pieces in the scene (i.e., different points with the same transverse coordinates (x, y) but different longitudinal coordinates z's), the depth associated to that specific (x, y) pair is the distance of the closest point to the reference plane among those points. In other words, the associated depth is the smallest number among those z's. Therefore, for each transverse coordinate pair (x, y), there exists at most one depth z. Mathematically speaking, one may therefore conclude that depth is a

function of transverse coordinates, i.e., $z = f(x, y)$, where the field of view (FOV) in the experiment setting specifies the domain of function f. Also, the range of this function and how f maps its domain to its range are uniquely determined by the scene. In other words, in a depth acquisition experiment, for a specific illumination/detection setup, the field of view is fixed and one gets (possibly) different depth functions for different scenes. We are going to refer to the function f as the *depth function* throughout this work.

Throughout this chapter, we assume the light source in the experiment illuminates only a part of the scene which is D_x-unit long along the x axis and D_y-unit long along the y axis. Also, as we are going to see in Section 3.3, the field of view in our experiment is exactly the part of the scene illuminated by the source. Therefore, without loss of generality, the domain of the depth function is assumed to be $(x, y) \in [0, D_x] \times [0, D_y]$. Also, as we will see in Section 3.3.1, the depth function is required to be measurable which is the case for all physical scenes without any doubt.

Having defined the depth function and its domain, we can now define the *depth map*. The depth function is a function from a subset of $\mathbb{R}^2$ to its range which is a subset of positive real numbers. Throughout this work, we assume the range of depth function is always a subset of $[0, T]$. Therefore, we can make a gray scale image out of the depth function where the normalized depth associated to a point (x, y) determines the brightness of that point in the image. To be more precise, we use the affine transformation $b(x, y) = 1 - \frac{f(x,y)}{T}$ where $b(x, y)$ represents the brightness at point (x, y) in the image (0 corresponds to black and 1 represents white). We call the resulting image *depth map* of the scene where the closest point to the reference plane on the scene has the brightest intensity and the farthest point on the scene is the darkest point.

Having defined the depth function and depth map, we introduce the basic idea behind raster scanning time of flight cameras and how they measure depth of a scene in the following section.

■ 3.2 Basic Concept Behind Time of Flight Range Imaging

As shown in Fig. 3.1, let us assume there is a directional light source (drawn as a black cross in Fig. 3.1) and it starts sending a spatially-narrow light beam (i.e., with infinitesimal cross section area) at a reference time, say $t = 0$, with intensity $I_s(t)$ over time. Also, let us assume the laser source is pointing in some direction, say the positive direction of the z axis (the red arrow in Fig. 3.1 represents the laser beam). From

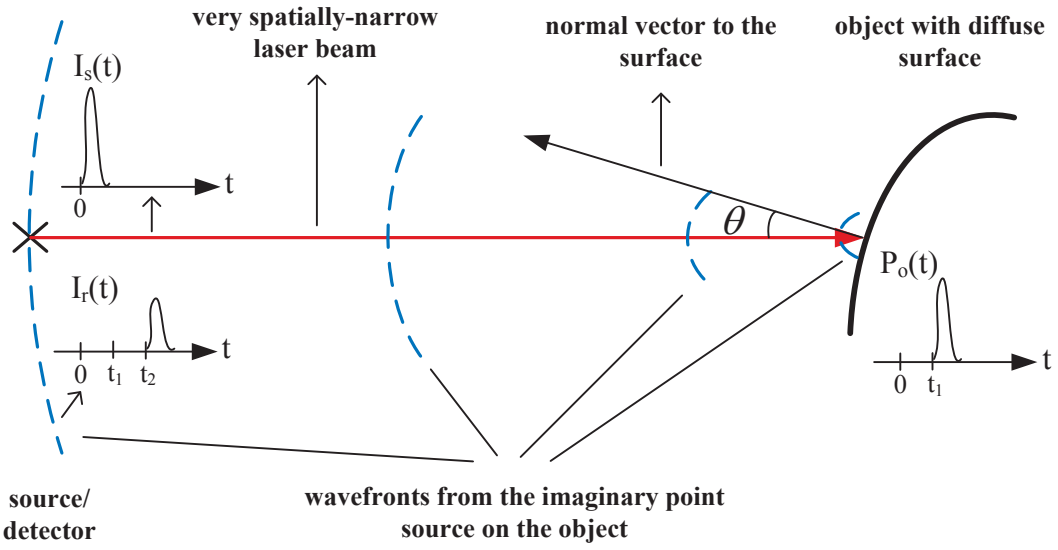


Figure 3.1. Basic idea behind time of flight range imaging.

the optics literature [19], it is well known that the speed of light is finite. Therefore, the time that the light beam arrives at a distance of d meters away from the source[1] is $t_1 = \frac{d}{c}$, where $c = 3 \times 10^8 \frac{m}{s}$ is the speed of light [19]. Therefore, mathematically speaking, we have

$$I_d(t) = \alpha I_s(t - t_1), \tag{3.1}$$

where $I_d(t)$ denotes the light intensity at the object surface which has a distance d from the source and α is the loss factor which is constant over time and decreases as the distance d increases. However, for a well designed laser source, α does not change drastically with d and is very close to one which means almost no loss. Throughout this work, we assume α to be constant with respect to d. As was indirectly assumed earlier, if there is an object located in the way of light beam in that distance, the light beam hits an infinitesimal area on the object and gets reflected off the object surface. Let us now assume the object surface is a diffuse surface, hence, the reflection of light off that small area of the surface can be modeled as a hypothetical point source located at that point with its power proportional to the received light intensity at that point [35]. To be more precise, the power of the imaginary point source on the object at time

[1] Distance is measured along the same axis as the light has been sent along, i.e., the z axis here.

t, denoted by $P_d(t)$, is

$$P_d(t) = I_d(t) A r \cos\theta, \tag{3.2}$$

where, $I_d(t)$ is the light intensity at the distance d from the source (received light intensity on the object at time t), A is the surface area on the object that the light beam illuminates and θ is the angle between the light beam and the normal vector at that infinitesimal area on the surface as shown in Fig. 3.1 and models foreshortening [35]. Also, r represents the reflectivity constant of the surface and is always less than or equal to one. Therefore, by plugging (3.1) into (3.2), we get

$$P_d(t) = I_s(t - t_1) \alpha A r \cos\theta, \tag{3.3}$$

where it is observed that as long as the object reflectivity and its position do not change over time, the proportionality constant (i.e., $\alpha A r \cos\theta$) stays the same. Hence, the power of the imaginary point source $P_d(t)$ is a time-delayed and attenuated copy of the source intensity $I_s(t)$. Furthermore, it is essential to note that the power of this imaginary source is proportional to $A\cos\theta$ which is the fronto-parallel area that the laser beam illuminates and therefore, is exactly equal to the cross section area of the laser beam itself. We are going to use this fact later in this chapter.

Now, the imaginary point source on the object emits light in all directions uniformly. Dashed blue lines in Fig. 3.1 represent wavefronts of the emitted light which are spherical. Also, let us assume there is a detector at the same place as the source location (the cross sign in Fig. 3.1); therefore, the light emitted from the imaginary point source on the object arrives at the detector at time $t_2 = \frac{d}{c} + \frac{d}{c} = \frac{2d}{c}$ and its intensity is equal to the power of the point source multiplied by $4\pi d^{-2}$ due to the uniform spread of power on a sphere of radius d. Therefore, we get the following equation for the received intensity at the detector

$$I_r(t) = \frac{1}{4\pi d^2} P_d(t - t_1) = \frac{\alpha A r \cos\theta}{4\pi d^2} I_s(t - t_2), \tag{3.4}$$

where $t_2 = \frac{2d}{c}$ and the second equality is a direct result of replacing $P_d(t)$ by (3.3). Because we assume the object position does not change over time, d is constant and therefore, the received light intensity at the detector $I_r(t)$ has the same shape (within an amplitude factor) in the time domain as the time delayed source intensity $I_s(t - t_2)$ does. Now, because the only difference (within an amplitude factor) between the sent and the received signals is the difference of their positions in time, if one were able to measure the exact light intensity at the detector, its time delay relative to the source

intensity could be precisely known. Hence, the distance of a very small part of the scene illuminated by the laser source from the source/detector plane could be measured accurately using this method.

It is worth noting the scheme explained above and shown in Fig. 3.1 is the basic idea behind raster scanning time of flight (ToF) ranging systems, specifically, raster scanning LiDAR systems [25, 39]. These systems usually have a laser source which can be moved on a grid in a transverse plane and by moving the laser source from point to point, they scan a transverse area on the scene and measure depths associated with the points in that area. For every position of the laser source, they face the same problem as the problem depicted in Fig. 3.1. In the end, after solving all those problems, they interpret the results obtained from each point as the depth of its corresponding pixel on the scene and form a depth map by putting those pixels together.

In the following section, we consider the experiment outlined in this section (i.e., measuring depth only for a small area or pixel) while we try to be more precise in modeling the physical experiment.

■ 3.3 Experiment Setup for Depth Acquisition and Its Model

Before delving into explaining the experiment setup and its different parts, let us first note that as we have seen in the previous section, when the illumination source and photodetector are in the same place, the time delay of the received signal at the detector relative to the source signal has the form of $\frac{2d}{c}$ where d is the distance of the point from the source/detector plane. For the sake of simplicity and to avoid carrying c in our equations, let us consider normalized speed of light such that the time delays are exactly equal to their corresponding distances. In other words, let us assume $c = 2$. Therefore, when it is said the delay is τ, we assume the distance of the point from the reference plane is τ and vice versa. Clearly, this does not limit the generality of our work and is only to simplify the equations.

Presented in Fig. 3.2 is a cartoon for the experiment setup we are going to build this whole work upon. The coordinate axes are drawn such that the y axis goes into the page and therefore, the figure can be interpreted as the top view of the experiment. First of all, as we stated before, because we want to measure the distance (depth) along the z axis, we illuminate the scene with light rays parallel to the z axis (red arrows in the figure). This can be done using different schemes; as in LiDAR systems, one may want to use a laser source which produces a spatially coherent light beam. The other

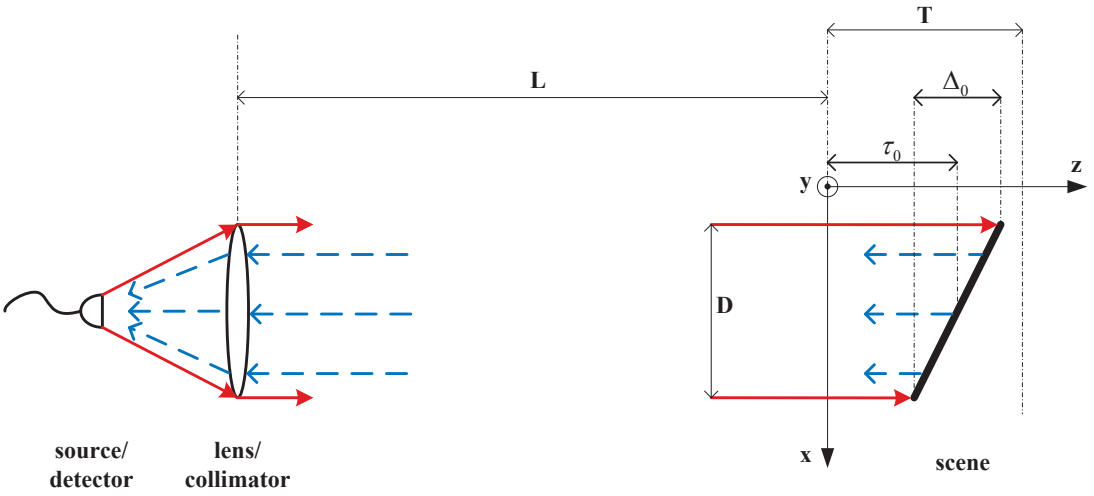


Figure 3.2. Schematic for the physical setup we are going to consider throughout this thesis to measure depth.

method is utilizing a point source and an optical collimator in front of the source as shown in Fig. 3.2. The output of the collimator is a light beam parallel to the central axis of collimator. Either way, the scene is illuminated with parallel light rays to the z axis and the light after hitting the scene is reflected back toward the illumination source. In Fig. 3.2, the scene is represented as a solid black line on the right and the dashed blue arrows represent the reflections.

Here, it should be noted that as was stated in Section 3.1, the illumination source is assumed to only illuminate the transverse area $(x, y) \in [0, D_x] \times [0, D_y]$. Furthermore, let us assume its intensity over that area is spatially-uniform and therefore, every point on the scene gets exactly the same intensity as other points do. We also utilize a lens to focus the reflected light into a detector with no spatial resolution and co-located with the source. It is also worth mentioning that when it is said source and detector or lens and collimator are co-located, in the physical experiment, they are not perfectly co-located; however, their separation is very small relative to other distances involved in the experiment. Therefore, for the sake of simplicity, one can as well assume they are in the same location and this is why every pair is drawn as one element in Fig. 3.2.

In addition, the scene is known to be somewhere between L and $L + T$ units away from the source/detector plane where we call L the offset distance. Without loss of generality, let us consider the source/detector plane to be at $z = -L$ and therefore, the uncertainty interval for the scene depth is $[0, T]$ as shown in Fig. 3.2. Furthermore,

let us assume the scene is a tilted plane relative to the reference plane, its center is at $z = \tau_0$ and its depth varies linearly from $\tau_0 + \frac{\Delta_0}{2}$ to $\tau_0 - \frac{\Delta_0}{2}$ when x goes from 0 to $D_x = D$ as shown in Fig. 3.2. It is clear that if the scene is a fronto-parallel plane to the reference plane, Δ_0 is zero and there is no depth variation along the x axis in that case.

Also, one might ask why, in Fig. 3.2, the reflections are drawn as light rays parallel to the z axis because as was seen earlier in Section 3.2 and Fig. 3.1, they must be radial with their origin at the reflection points. However, here, we assume the offset distance L to be very large and the detector area to be small enough that the curvature at the detector plane can be neglected. In other words, the received wavefront can be approximated as a flat plane with no curvature and parallel to the detector plane, hence, dashed blue lines are drawn as perpendicular to the lens plane. Therefore, we can assume all the reflected light from the top point ($x = 0$) in the scene comes into the detector with a time delay equal to $L + \tau_0 + \frac{\Delta_0}{2}$ and the time delay for the reflections from the bottom point of the scene ($x = D$) is $L + \tau_0 - \frac{\Delta_0}{2}$.

Now, having defined all different parts of the depth acquisition experiment, we are ready to improve the assumptions we made in Section 3.2 in order to model the low-photon depth acquisition problem more accurately. In the following section, we use linear systems theory and specifically, convolution to model the reflection of the light off the scene more precisely. We, then in Section 3.3.2, model the effect of low signal power received at the detector, dark count of the detector, existence of ambient light and undesired reflections from the parts of the scene outside of the field of view. Finally, modeling the effect of timing jitter of the detector is addressed in Section 3.3.3.

■ 3.3.1 A Model for the Received Light Intensity

Contrary to what we assumed earlier in Section 3.2, in general, the received signal (normalized power) at the detector, $r(\cdot)$, is not simply a time delayed copy of the source signal (within an amplitude factor). However, the received signal, in general, is a distorted copy of the source signal and the distortion is due to the fact that a laser source (or any other light source) cannot be very spatially-narrow, contrary to what was assumed in Section 3.2. Hence, there could be depth variation within the part of the scene illuminated by that light beam as shown in Fig. 3.2 where there is linear change in the depth along the x axis. Even if we could make a very spatially narrow light beam, it would not be desired to do so. As was stated in Section 3.2, the reflected power off the illuminated pixel is directly proportional to $A\cos\theta$ where A is

the surface area of the pixel and θ is the angle between the illumination beam and the normal vector to the surface at that pixel. We also recall from Section 3.2 that $A\cos\theta$ is basically the orthogonal area that the beam illuminates and consequently is equal to the cross section area of the laser beam itself. Therefore, for a fixed source intensity, it is desired to have a laser beam with an area as large as possible to boost the received power at the detector. Therefore, for a specified source intensity, there is a trade off between the received signal power at the detector and how constant the depth is or in other words, how undistorted the received signal is in comparison to the source signal. In this section, we are going to quantify the distortion caused by reflecting off an area of the scene which possibly has depth variation.

Let us consider the experiment shown in Fig. 3.2 and recall the transverse area which is being illuminated by the source is $(x, y) \in [0, D_x] \times [0, D_y]$, the offset distance is L (the reference plane is at $z = -L$) and the depth function maps its domain to a subset of $[0, T]$. Therefore, the distance of each point with coordinates $(x, y, f(x, y))$ on the scene from the source/detector plane is $d(x, y) = L + f(x, y)$. Also, let us assume the illumination source in Fig. 3.2 has uniform intensity over the area $[0, D_x] \times [0, D_y]$ and its normalized power (referred to as the source signal) is $s(t)$. Consequently, the source intensity is $I_s(t) = \frac{s(t)}{D_x D_y}$. Also, let us define the received signal $r(t)$ as the normalized power impinging on the photodetector. As discussed earlier, the received light intensity on the detector area is assumed to be uniform, i.e., $r(t) = A_D I_r(t)$ where A_D is the area of the detector[2]. Now, we are going to revisit (3.4) to find the relation between the source signal $s(t)$ and the received signal $r(t)$. In order to be able to use (3.4), a very necessary assumption to make is that the depth of the patch on the scene illuminated by the light beam is constant over that area. Therefore, let us discretize the scene depth into z_k's where $z_k = (k-1)\frac{T}{K}, k = 1, \ldots, K$ for a large enough K and use (3.4) for each z_k and its corresponding patch on the scene. For that purpose, for each $k = 1, \ldots, K$, we should replace d by $L + z_k$, $A\cos\theta$ by $\mu\left(\{(x, y)|f(x, y) \in (z_{k-1}, z_k]\}\right)$ where z_0 is defined as a negative number, say $z_0 = -1$ and $\mu(\mathcal{A})$ is the Lebesgue measure of set $\mathcal{A}$ and is equivalent to the area occupied by the set if the set is in $\mathbb{R}^2$. Also, recalling $t_2 = d$ (due to the normalized speed of light), we replace t_2 by $L + z_k$. Therefore, we

[2]This is because the offset distance L and the detector area A_D are assumed to be large and small enough respectively that the curvature of the wavefront on the detector plane can be ignored.

get

$$I_r(t;z_k) = \frac{\alpha r}{4\pi\left(L+z_k\right)^2} I_s(t-(L+z_k))\mu(\{(x,y)|f(x,y)\in(z_{k-1},z_k]\}), \quad k=1,\ldots,K, \tag{3.5}$$

where $I_r(t;z_k)$ represents the intensity at the detector at time t due to the reflections from the patch with depth z_k. The reflectivity constant is denoted by r and let us assume it is equal to one over the whole scene; i.e., scene is assumed to be white. Furthermore, as we stated earlier in Section 3.2, α, which is the forward path loss experienced by the laser intensity, can be very well approximated as constant with respect to the distance. Therefore, for the sake of simplicity, let us drop the constant factors in (3.5) to get the following simplified equation for the contribution of patch with depth z_k to the detector intensity.

$$I_r(t;z_k) = \frac{1}{\left(L+z_k\right)^2} I_s\left(t-(L+z_k)\right)\mu(\{(x,y)|f(x,y)\in(z_{k-1},z_k]\}), \quad k=1,\ldots,K, \tag{3.6}$$

Now, we should note that because the detector has no spatial resolution, all the contributions from different patches on the scene are added together to form the intensity at the detector, i.e., $I_r(t;K) = \sum_{k=1}^{K} I_r(t;z_k)$, where $I_r(t;K)$ denotes the received intensity at the detector when the scene depth is descretized into K equidistant points in the interval $[0,T]$. Now if we let K go to infinity, the approximation becomes precise and we get

$$I_r(t) = \lim_{K\to\infty} \sum_{k=1}^{K} \frac{1}{\left(L+z_k\right)^2} I_s\left(t-(L+z_k)\right)\mu(\{(x,y)|f(x,y)\in(z_{k-1},z_k]\}), \tag{3.7}$$

where $I_r(t)$ is the received light intensity at the detector. Now to obtain the relation between $s(t)$ and $r(t)$, let us assume the detector area to be one ($A_D = 1$) which results in $r(t) = I_r(t)$. It is also easily observed that (3.7) is the definition of the Lebesgue integral when the depth function f is measurable. Therefore, in (3.7), by substituting the sum with integral and replacing $I_s(\cdot)$ by $\frac{1}{D_x D_y} s(\cdot)$, we get

$$r(t) = \frac{1}{D_x D_y} \int_0^{D_y} \int_0^{D_x} \frac{1}{\left(L+f(x,y)\right)^2} s\left(t-(L+f(x,y))\right) \mathrm{d}x\mathrm{d}y, \tag{3.8}$$

where we can use linear systems theory and write $s(t-(L+f(x,y)))$ as the convolution of $s(t)$ with a time delayed impulse, $s(t-(L+f(x,y))) = s(t) * \delta(t-(L+f(x,y)))$, to get

$$r(t) = \frac{1}{D_x D_y} \int_0^{D_y} \int_0^{D_x} \frac{1}{(L+f(x,y))^2} \int_{-\infty}^{\infty} s(t-t')\delta\left(t' - (L+f(x,y))\right) \mathrm{d}t' \mathrm{d}x \mathrm{d}y, \tag{3.9}$$

where the innermost integral denotes the convolution integral. Now, we can change the order of integrals[3] in (3.9) to get the convolution of the source signal and the scene impulse response as the following

$$r(t) = s(t) * \widetilde{h}(t) = \int_{-\infty}^{\infty} s(t-t')\widetilde{h}(t')\mathrm{d}t', \tag{3.10}$$

where

$$\widetilde{h}(t) \triangleq \frac{1}{D_x D_y} \int_0^{D_y} \int_0^{D_x} \frac{1}{(L+f(x,y))^2} \delta\left(t - (L+f(x,y))\right) \mathrm{d}x \mathrm{d}y \tag{3.11}$$

is defined as the impulse response of the scene. The reason that we call $\widetilde{h}(t)$ the scene impulse response is that it is the received signal at the detector when $s(t) = \delta(t)$; i.e., when the illumination is an impulse in time domain.

It is worth noting that first we could have used linearity and time invariance properties of light propagation and reflection to conclude that the system (with its input as $s(t)$ and its output as $r(t)$) is linear and time invariant (LTI) and therefore, to characterize the relation between its input and its output, it suffices to know the impulse response of the system $\widetilde{h}(t)$. Then, we could have derived $\widetilde{h}(t)$ by directly evaluating the received signal at the detector due to an impulse illumination in time domain which would have yielded the same result as (3.11). From LTI systems theory, it is well known that the response to an arbitrary input signal $s(t)$ for the LTI system with the impulse response $\widetilde{h}(t)$ is $r(t) = \widetilde{h}(t) * s(t)$. This is how we formulated another time-resolved imaging problem in our earlier work [29]. We should also note that $\widetilde{h}(t)$ in (3.11) is an impulse $\widetilde{h}(t) = \frac{1}{(L+\tau_0)^2}\delta\left(t - (L+\tau_0)\right)$ if $f(\cdot,\cdot)$ is constant and equal to τ_0 over its whole domain. Otherwise, it would not be an impulse and therefore, $r(t) = s(t) * \widetilde{h}(t)$ would be a distorted copy of $s(t)$. Taking the distortion into account is very crucial for the

[3]The order of integrals in (3.9) can be changed because the integrand is absolutely integrable; i.e., it is a function in $\mathcal{L}_1\left([0,D_x]\times[0,D_y]\times\mathbb{R}\right)$.

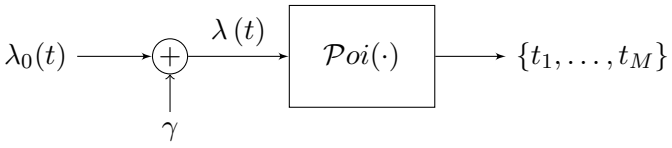


Figure 3.3. Simplified block diagram for a direct detection photodetector when the received light intensity at the detector is low and there exists background noise. $\lambda_0(\cdot)$ represents the normalized power of the impinging light on the detector area due to the reflections off the scene and γ models background noise. Also, the detector is assumed to output the exact photon arrival times $\{t_1, \ldots, t_M\}$ in the time interval $(L, L+T]$ where M itself is a random variable.

estimation problem in Chapter 4 because if $f(\cdot, \cdot)$ is assumed to be constant while it is not (meaning that there is distortion in the received signal), the estimation algorithm uses a mismatched model to estimate the depth and as we are going to see in Chapter 4, depending on how mismatched the model is, the results could be totally wrong. As was stated in Chapter 1, one of the main contributions of this work is modeling the distortion (the scene impulse response) as a parametric signal when there is linear depth variation in the scene. We are going to talk more about this in Section 3.4.

■ 3.3.2 Low Power, Ambient Light and Detector Dark Count

The framework outlined in Section 3.2 gives us a very good insight into the existing raster scanning LiDAR technologies and what those systems do to measure the depth of a pixel on the scene. However, the assumption made in that section about measuring the exact received light intensity fails to hold especially when the received light intensity at the detector is very low. The low power/intensity at the detector could be either due to the large offset distance L between the scene and the illuminating laser source which leads to a large intensity attenuation or because of power constraints at the source and hence, using low power to illuminate the scene to begin with.

Furthermore, in a real experiment, there always exists ambient light. Also, there could be reflections from other parts of the scene than the area (pixel) whose depth is to be measured. As was stated in Chapter 2, for this low power situation with the existence of background light, a more accurate model for the photodetector is the Poisson model we introduced in that chapter and which we have redrawn in Fig. 3.3 for future reference. The overall effect of ambient light, undesired reflections from other parts of the scene (outside the field of view) and the dark count of photodetector can be modeled as an additive independent rate γ which we introduced in detail in Chapter 2 and referred to as background noise. $\lambda_0(\cdot)$ is the normalized power (with unit of photons per second)

received from the reflections off the area of the scene whose depth is to be measured. As is observed in Fig. 3.3, the measurements at the output of photodetector are the photon arrival times. These measurements are random and their joint probability distribution is related to the depth function f which we are going to explain in more detail in Section 3.4. In the following section, we are going to explain how to modify $\lambda_0(\cdot)$ in order to take into account the effect of timing jitter in the detector.

■ 3.3.3 Detector Timing Jitter

Another imperfection of the detector is due to its timing jitter. Timing jitter in the detector causes uncertainty in the arrival times and its effect can be modeled as the convolution of the probability density function (pdf) of the timing jitter with the normalized power of received light at the detector [15]. To be more precise, if we assume the normalized power of received light at the detector to be $r(t)$ and the pdf of detector timing jitter to be $p_J(t)$, we get

$$\lambda_0(t) = r(t) * p_J(t), \tag{3.12}$$

where $*$ represents the convolution operator and $r(\cdot)$ is the received signal and $\lambda_0(\cdot)$ is the signal part of the rate of the Poisson process as shown in Fig. 3.3. This by itself makes the rate of the Poisson process different than a time delayed and attenuated copy of the source signal as was assumed in Section 3.2. In other words, in addition to the distortion we discussed in Section 3.3.1 due to the scene impulse response, the timing jitter in the detector also introduces distortion in the received signal.

Now, using (3.12), it is known that due to timing jitter in the detector, the signal part of the Poisson process rate is $\lambda_0(t) = r(t) * p_J(t)$, where $r(t)$ is given in (3.10). Therefore, by plugging (3.10) into (3.12), we have

$$\lambda_0(t) = s(t) * \widetilde{h}(t) * p_J(t) = (s(t) * p_J(t)) * \widetilde{h}(t), \tag{3.13}$$

where for the second equality, we have used the commutativity of the convolution operator. It is clearly seen that our model takes the jitter effect into account by just defining a new source signal $\tilde{s}(t)$ as the convolution of the original source signal $s(t)$ and the pdf of timing jitter $p_J(t)$, i.e., $\tilde{s}(t) = s(t) * p_J(t)$. Therefore, we can have the block diagram shown in Fig. 3.4 to model the relation between the scene impulse response $\widetilde{h}(t)$ and the measurements at the output of photodetector $\{t_1, \ldots, t_M\}$ obtained over the time interval $(L, L+T]$. Using this block diagram, one can write the distribution

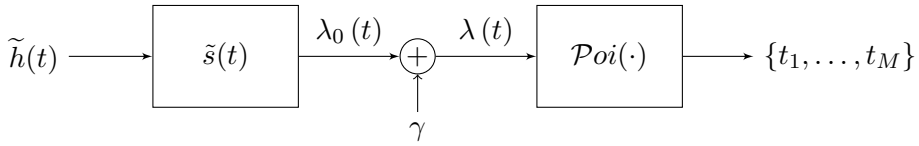


Figure 3.4. Block diagram for the depth acquisition procedure when the number of photons is low, there exists background noise, the detector has timing jitter and the scene has an impulse response $h(t)$.

of $\{t_i\}_{i=1}^{M}$ in terms of $\widetilde{h}(t)$ which is unknown and has to be estimated.

In the next section, we are going to consider a few parametric families for the depth function f and derive their corresponding impulse responses $\widetilde{h}(t)$ which will be parametric as well and then in Chapter 4, we are going to restrict $\widetilde{h}(t)$ to belong to two of those parametric families to form an estimation problem.

■ 3.4 A Parametric Model for the Scene Impulse Response

In the rest of this thesis, for the sake of simplicity, we assume there is no depth variation along the y axis in the scene and hence, $z = f(x, y) = f(x)$ with slight abuse of notation. It is worth noting that this assumption is only to simplify the mathematics to derive the proposed parametric model for the scene impulse response $\widetilde{h}(t)$ in this section and does not limit the generality of this work. Also, after introducing our model here, it is straightforward to verify the model is valid even for a broader set of scenes. In addition, assuming there is no depth variation along the y axis is a reasonable assumption in both following situations.

- If it is known that the scene consists of only planes perpendicular to a single plane, we can define the coordinate system such that the x-z plane is the plane which all the other planes are perpendicular to. Therefore, the distance of points with same x on the scene from the reference plane (the $z = 0$ plane) is the same regardless of y (e.g., the scene depicted in Fig. 3.2 falls in this category). Clearly, the above assumption is completely precise in this case.

- If the field of view along the y axis is so small that the scene depth variations along that axis can be neglected, one can say the depth function approximately is a function of x alone.

Now, let us assume we have the first case here with $D_x = D = 1$ and $D_y = 1$. Therefore, (3.11) can be rewritten as

$$\widetilde{h}(t) = \int_0^1 \frac{1}{(L + f(x))^2} \delta\left(t - (L + f(x))\right) \mathrm{d}x, \tag{3.14}$$

where L is the offset distance and $\widetilde{h}(t)$, the impulse response of the scene, at each time instant t is expressed as a functional of the depth function f. Also, because $f(x) > 0$ for all x in the domain of f, $\widetilde{h}(t)$ is zero before $t = L$. Now, because in the block diagram of Fig. 3.4, the mapping from $\widetilde{h}(\cdot)$ to $\lambda(\cdot)$ is time invariant, we can shift the time origin to L (i.e., shift $\widetilde{h}(\cdot)$ in the time domain) and therefore, $\lambda(\cdot)$ experiences the same time shift and instead of obtaining measurements over the time interval $(L, L + T]$, we can observe the output in the time interval $(0, T]$. Therefore, let us define $h_L(t) \triangleq \widetilde{h}(t + L)$ to get

$$h_L(t) = \int_0^1 \frac{1}{(L + f(x))^2} \delta\left(t - f(x)\right) \mathrm{d}x. \tag{3.15}$$

where $h_L(\cdot)$ is a time-advanced copy of $\widetilde{h}(\cdot)$. Because in the rest of this thesis, L is always constant and never changes, let us drop the subscript L and note that when we say $h(t)$, we really mean $h_L(t) = \widetilde{h}(t + L)$ as in (3.15).

Because we are using a detector with no spatial resolution, and as it is seen in (3.15), the mapping from $f(\cdot)$ to $h(\cdot)$ is not a one-to-one mapping. For example, it is easy to verify that $f(x)$ and $f(1 - x)$ result in the same $h(\cdot)$. This should not be surprising because in this experiment, if we get an impulse response $h(t)$ which is nonzero over an interval $[\tau_0 - \frac{\Delta_0}{2}, \tau_0 + \frac{\Delta_0}{2}]$, we do not know which part of the scene has contributed to a specific depth in that interval. In other words, we only know that specific depth exists in the scene and do not know where in the scene it is located. Therefore, if the scene is assumed to be a plane like in Fig. 3.2, some information about the orientation of the plane is lost in this experiment. For example, we do not know whether the top corner of that plane is closer to the scene or the bottom corner is closer. However, this is not a big issue if the proposed framework in this thesis is employed in a bigger problem. In that case, the considered problem in this work is equivalent to estimating the depth of a single pixel. Therefore, we can overcome the spatial ambiguity by using the spatial correlation among adjacent pixels and the fact that the depth map should be smooth to find the right orientation for each pixel. Another way to mitigate the

ambiguity is what the authors of CoDAC do [28]. They use another experiment on top of the experiment we have described here to determine which part of the scene is responsible for a specific depth in the scene. This is out of the scope of this thesis and is not going to be discussed in more detail here; however, curious reader is encouraged to refer to the CoDAC paper [28]. In this work, because we do not want to deal with this spatial ambiguity, we assume the function f is piecewise continuous (with finite number of discontinuities) and non-increasing in x (i.e., $\frac{\mathrm{d}f}{\mathrm{d}x}(x) < 0$ over each continuous piece and undefined at each discontinuity). This assumption guarantees a one-to-one relation between $f(\cdot)$ and $h(\cdot)$.

Finally, as we stated earlier in this chapter, we always consider the range of the depth function to be a subset of $[0, T]$. Therefore, T can be considered as the length of depth uncertainty interval. Now, let us assume that the offset distance L relative to the length of depth uncertainty interval T, is so large that $\frac{1}{L^2}$ and $\frac{1}{(L+T)^2}$ are almost the same. In other words, let us assume $\left(\frac{T}{L}\right)^2 \ll 1$. The physical interpretation of this assumption is that we are trying to localize an object in a very short longitudinal interval which is very far from the illumination/detection plane (equivalently, we are considering a depth acquisition problem with low range uncertainty). Therefore, we can approximate $\frac{1}{(L+f(x))^2}$ as $\frac{1}{L^2}$ for $x \in [0, 1]$. Also let us assume L to be one, therefore, we can simplify the equation for the impulse response as the following

$$h(t) = \int_0^1 \delta\left(t - f(x)\right) \mathrm{d}x. \tag{3.16}$$

In the rest of this thesis, except in Section 3.5, we are going to consider (3.16) for the scene impulse response which implies we are considering the depth acquisition problem with low range uncertainty. In Section 3.5, we investigate how to modify our model to be suitable for problems with high range uncertainties. Furthermore, it should be noted that $h(t)$ in (3.16) always integrates to one regardless of what the depth function f is (as long as the depth function is a measurable function) because

$$\int_{-\infty}^{\infty} h(t)\mathrm{d}t = \int_{-\infty}^{\infty}\int_0^1 \delta(t - f(x))\mathrm{d}x\mathrm{d}t = \int_0^1\int_{-\infty}^{\infty} \delta(t - f(x))\mathrm{d}t\mathrm{d}x = \int_0^1 1\ \mathrm{d}x = 1. \tag{3.17}$$

In the following two sections, we consider two different simplified models for the scene (its associated depth function) and derive their impulse responses.

■ 3.4.1 Scene Comprised of Only One Fronto-Parallel Plane

Here, we assume there is only one plane in the scene and it is parallel to the reference plane with $f(x) = \tau_0$. Therefore, using (3.16), the scene impulse response reduces to an impulse $h(t) = \int_0^1 \delta(t - \tau_0)\mathrm{d}x = \delta(t - \tau_0)$. This is what we refer to as the conventional model in the rest of this thesis because in the existing LiDAR literature, they consider constant depth for each pixel (scene) which is exactly equivalent to the model here where the impulse response of the scene is a time-delayed impulse. The left panels in Fig. 3.5 show an example for this case where the top panel depicts the scene (or equivalently, the depth function $z = f(x)$) and the bottom panel is a plot of the scene impulse response, $h(t)$, which is a Dirac delta function located at $t = \tau_0$.

■ 3.4.2 Scene Comprised of Only One Slanted Plane

In this section, we assume there is only one plane in the scene and it is tilted relative to the reference plane (for example, the scene in Fig. 3.2). Therefore, f is an affine function of the transverse coordinate x or in other words, depth variation in the scene is linear. Let us assume $f(0) = \tau_0 + \frac{\Delta_0}{2}$ and $f(1) = \tau_0 - \frac{\Delta_0}{2}$. Therefore, we get $f(x) = -\Delta_0 x + \tau_0 + \frac{\Delta_0}{2}$ (equivalently, we label the center of the interval that the scene depth occupies as τ_0 and label the depth spread in the scene Δ_0). Now let us plug the affine equation for $f(x)$ derived above into (3.16) to compute $h(t)$ as the following:

$$\begin{aligned} h(t) &= \int_0^1 \delta\left(t - (\tau_0 + \frac{\Delta_0}{2} - \Delta_0 x)\right) \mathrm{d}x \\ &= \frac{1}{\Delta_0}\int_0^{\Delta_0} \delta\left(t - (\tau_0 + \frac{\Delta_0}{2} - x')\right) \mathrm{d}x' \\ &= \mathrm{rect}\left(\frac{t - \tau_0}{\Delta_0}\right), \end{aligned} \tag{3.18}$$

where we use change of variable $x' = \Delta_0 x$ to derive the second line from the first line and $\mathrm{rect}(t)$ is a box function with unit amplitude for t's between $-\frac{1}{2}$ and $\frac{1}{2}$ and zero everywhere else. By inspecting (3.18), we realize $h(t)$ in this case is a box function between $t = \tau_0 - \frac{\Delta_0}{2}$ and $t = \tau_0 + \frac{\Delta_0}{2}$ with an amplitude equal to $\frac{1}{\Delta_0}$ and zero everywhere else. Therefore, as opposed to the scenario discussed in Section 3.4.1, if the scene is made up of only one plane which is tilted relative to the reference plane, the impulse response of the scene is not an impulse and is a box function instead. Clearly, if we consider all such impulse responses by considering all possible choices of τ_0 and Δ_0, we have a parametric class of functions for the scene impulse response which is characterized by

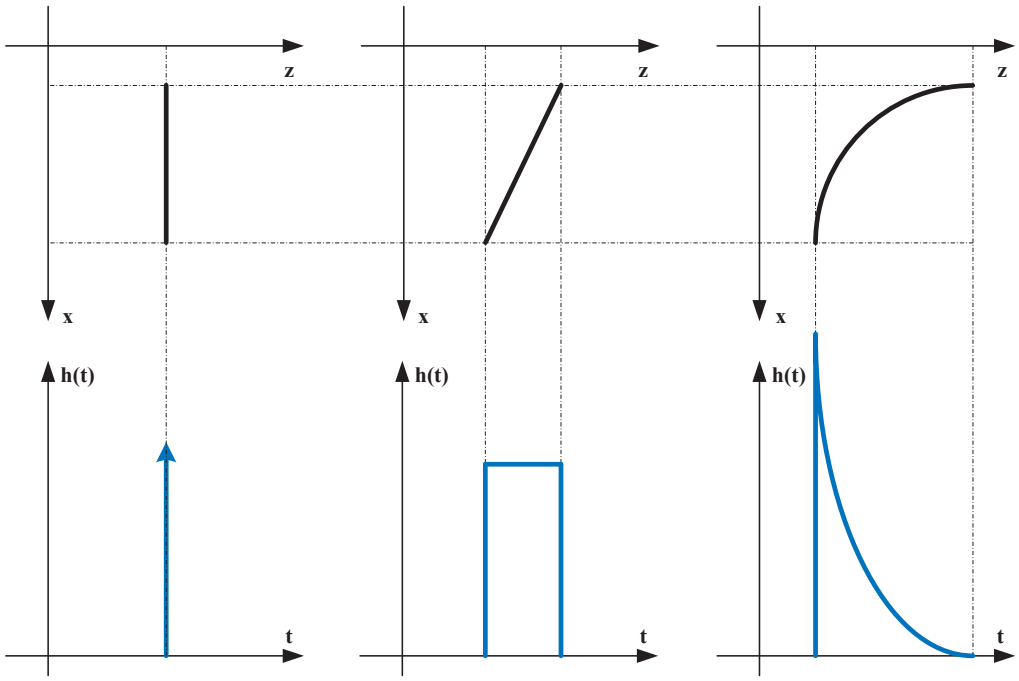


Figure 3.5. Three different models for the scene (top row) and their corresponding impulse responses (bottom row). The left panels depict the case where the scene is a single plane which is fronto-parallel to the reference plane. The middle panels represent an example for when the scene is a single slanted plane and the right belong to the scenario where the scene depth function is nonlinear.

two independent parameters, τ_0 and Δ_0. The middle panels in Fig. 3.5 represent an example for this scenario where the top panel shows the scene (or the depth function $z = f(x)$) and the bottom panel is a plot of the scene impulse response, $h(t)$, which is a box function.

■ 3.4.3 Beyond One Linear Plane

In this part, we briefly talk about the situations where the scene consists of multiple planes or the scene depth function is a nonlinear function (i.e., the scene is not piecewise planar and there is curvature in the scene).

Piecewise Planar Scene

In this part, the scene is assumed to be piecewise planar and therefore, its depth function can be considered as $f(x) = \sum_{k=1}^{K} f_k(x)$, where K is the number of planes (linear pieces) in the scene and each $f_k(\cdot)$ represents a linear piece. Let us denote the domain of f_k

by $\mathcal{D}_k$ and note that the measure (length) of the interval $\mathcal{D}_k$ is equal to the area of the orthogonal projection of k^{th} plane onto the $x-y$ plane (reference plane) for $k = 1, \ldots, K$. Also, f_k's are assumed to have disjoint domains and the union of their domains is the domain of the original depth function f which is $\mathcal{D} = [0, 1]$. Mathematically speaking, $\{\mathcal{D}_k\}_{k=1}^{K}$ partition the original domain $\mathcal{D}$ and we have $\sum_{k=1}^{K} a_k = 1$, where a_k is the length of the interval $\mathcal{D}_k$, $k = 1, \ldots, K$ and hence, positive. Now, using (3.16), we get

$$\begin{aligned} h(t) &= \int_0^1 \delta\left(t - \sum_{k=1}^{K} f_k(x)\right) \mathrm{d}x \\ &= \sum_{k=1}^{K} \int_{\mathcal{D}_k} \delta\left(t - f_k(x)\right) \mathrm{d}x \\ &= \sum_{k=1}^{K} a_k \int_0^1 \delta\left(t - f_k(x')\right) \mathrm{d}x' \\ &= \sum_{k=1}^{K} a_k h_k(t), \end{aligned} \tag{3.19}$$

where, in the second line, we are allowed to pull the sum out of the Dirac delta function because of the disjointness of $\mathcal{D}_k$'s. The third line is a direct result of change of integration variable such that each $\mathcal{D}_k$ gets mapped to $[0, 1]$ and hence, we have the coefficient a_k in front of the new integral. Using our results from Sections 3.4.1 and 3.4.2, each $h_k(\cdot)$ is either an impulse if its corresponding plane is fronto-parallel to the reference plane or a box function if the piece is tilted relative to the reference plane. It should also be noted that each $h_k(\cdot)$ integrates to one according to (3.17). Therefore, the integral of $h(\cdot)$ is also equal to one which is not surprising due to (3.17). Also, recalling that $\sum_{k=1}^{K} a_k = 1$, the scene impulse response $h(t)$ for a piecewise planar scene is a convex combination of the impulse responses of planar pieces in the scene which are either impulses or box functions according to our results from Sections 3.4.1 and 3.4.2.

Generic Nonlinear Scene

In this part, we derive the impulse response for a scene with an arbitrary depth function f. First, let us exclude intervals with nonzero length (measure) where the derivative of f is zero. In other words, let us exclude the intervals (with nonzero length) in which the depth function is constant. This assumption is equivalent to assuming the derivative of

the depth function with respect to x is nonzero over the domain of f, which is $\mathcal{D} = [0, 1]$, except on possibly a set of points with zero measure. This is because the intervals with nonzero measure in which the derivative of f is zero correspond to fronto-parallel planes in the scene and we already know how to calculate their contribution to the impulse response from the previous part. Also, because in the previous part, we learned how to deal with discontinuities in the depth function f, let us assume $f(x)$ is continuous over its domain $\mathcal{D}$. Furthermore, as we stated before, let us assume f is non-increasing or equivalently $\frac{\mathrm{d}f}{\mathrm{d}x}(x) \leq 0$ for $x \in \mathcal{D}$. Now, we can use (3.16) to derive the impulse response of the scene as the following

$$\begin{aligned} h(t) &= \int_0^1 \delta\left(t - f(x)\right) \mathrm{d}x \\ &= \int_{z_{\min}}^{z_{\max}} \frac{1}{|f'(f^{-1}(z))|} \delta(t - z) \mathrm{d}z \\ &= \begin{cases} \dfrac{1}{|f'(f^{-1}(t))|} & z_{\min} \leq t \leq z_{\max} \\ 0 & \text{o.w.}, \end{cases} \end{aligned} \tag{3.20}$$

where, in the second line, we change the integration variable to $z = f(x)$ and f' is the derivative of f with respect to x and f^{-1} is the inverse of the function f. Also, we denote the smallest and largest depths in the scene by $z_{\min}$ and $z_{\max}$, respectively. Here, we should note that f^{-1} is well defined because we have assumed the depth function to be monotonically decreasing over its domain and therefore, it is invertible. If the depth function f belongs to a parametric family of functions, say positive second order polynomials (parabola), then, according to (3.20), the scene impulse response is in a parametric family too. An example for a nonlinear f and its corresponding impulse response is presented in the right panels of Fig. 3.5. The top panel is the nonlinear scene (depth function) and the bottom panel shows its impulse response which goes to infinity at $t = z_{\min}$, however, the impulse response is integrable according to (3.17).

■ 3.5 How to Modify the Model for High Range Uncertainty

In the previous section, we assumed the depth acquisition problem to have low range uncertainty and the radial fall-off factor $\frac{1}{(L+f(x))^2}$ was therefore approximated as L^{-2} which does not depend on where $h(t)$ is located in the depth uncertainty interval $[0, T]$. Intuitively speaking, this makes sense because the depth uncertainty interval for that situation is assumed to be really small (relative to L) and therefore, where $h(t)$ is

located in that interval is immaterial for the radial fall-off calculation. However, this is not the case when the length of the depth uncertainty interval $[0, T]$ is comparable to the offset distance L (this situation is referred to as high range uncertainty). In this case, we should use (3.15) to find the impulse response of the scene. The only difference between (3.15) and (3.16)–(3.20) is the $\frac{1}{(L+f(x))^2}$ factor in (3.15) which we dropped in (3.16)–(3.20). Therefore, by inspecting (3.20), it is straightforward to conclude that the scene impulse response for the situation with high range uncertainty is simply $h(t)$ for the same scene derived in the previous section (with low range uncertainty assumption) multiplied by $(L + t)^{-2}$ over its support. In other words, if we denote the scene impulse response for a high range uncertainty situation by $h_{\text{HRU}}(t)$, we have: $h_{\text{HRU}}(t) = (L + t)^{-2}h(t)$, where $h(t)$ is calculated using (3.16)–(3.20) depending on what the scene model is. For example, for a scene comprised of only one tilted plane, we get $h_{\text{HRU}}(t) = \frac{(L+t)^{-2}}{\Delta_0} \operatorname{rect}\left(\frac{t-\tau_0}{\Delta_0}\right)$ which is not a box function on contrary to what we observed for the low range uncertainty situation. Now, if we assume that the depth spread of the scene, Δ_0, is small enough, we can approximate $(L+t)^{-2}$ as $(L+\tau_0)^{-2}$ and get $h_{\text{HRU}}(t) = \frac{(L+\tau_0)^{-2}}{\Delta_0} \operatorname{rect}\left(\frac{t-\tau_0}{\Delta_0}\right)$ which is a box function with its amplitude depending on the position of the center of the box, i.e., τ_0. This situation happens when we have a problem with high range uncertainty and the scene is known to have small depth spread (the range of the depth function f is assumed to be small).

Having seen different models for the scene impulse response, we are ready to use two of those models (the Dirac delta function and the box function) in the next chapter to form a parametric estimation problem in order to estimate the scene depth.

Chapter 4

Depth Estimation: Algorithm and Analysis

In the previous chapter, we formulated the relation between the photon arrival times at the output of direct detection photodetector and the scene impulse response $h(t)$ and summarized that relation in the block diagram of Fig. 3.4. We also derived two simple parametric models for the scene impulse response in Sections 3.4.1 and 3.4.2 among other more complicated models. Furthermore, we know from Chapter 2 that when the rate of the photon arrival times is parametric, we have a parametric estimation problem which can be formulated as a maximum likelihood problem. Therefore, in the following section, we employ the maximum likelihood method of Chapter 2 with the two simple parametric models derived for the scene impulse response in Chapter 3 to formulate optimization problems which will then be solved to estimate the depth of the scene. We present numerical analysis of our method (maximum likelihood with the parametric models for the scene impulse response) using simulation results in Section 4.2.

■ 4.1 Inference Algorithm

We divide this section into two parts; in the first part, we consider the scenario where the scene consists of only one fronto-parallel plane as was assumed in Section 3.4.1 and solve an optimization problem to estimate the distance of that fronto-parallel plane from the reference plane. We call this model the conventional model throughout the rest of this work because that is what the existing raster scanning LiDAR technologies use as we explained in Chapter 3. Then, in the second part of this section, we consider the introduced model for the scene impulse response in Section 3.4.2 which has two parameters and we formulate another optimization problem to estimate those two

parameters and eventually the scene depth.

■ 4.1.1 Only One Fronto-Parallel Plane in the Scene

Here, we assume the scene is made up of only one plane which is parallel to the reference plane and therefore, as was discussed in Chapter 3, the scene depth function is constant over its domain $f(x, y) = \tau_0$ and the scene impulse response is a time-delayed impulse $h(t) = \delta(t - \tau_0)$, where τ_0 is the distance of the scene from the reference plane. Now, in order to estimate the depth of the scene, τ_0, we use the block diagram of Fig. 3.4 and the maximum likelihood formulation of (2.7) with α replaced by τ. In other words, we have $\lambda_0(t;\tau) = s(t) * \delta(t - \tau) = s(t - \tau)$, where $*$ denotes the convolution operator and $s(t)$ is the source signal[1]. Also, from Chapter 2, we recall that we are allowed to use (2.7) only if the received energy $E_s = \int_0^T \lambda_0(t;\tau)\mathrm{d}t$ does not change by varying τ in its feasible range. As we discussed in the previous chapter, this is the case if we are considering a depth acquisition problem with low range uncertainty. Therefore, we have

$$\begin{aligned}\widehat{\tau} &= \underset{\tau\in(0,T)}{\operatorname{argmax}} \sum_{i=1}^{M} \log\left(\gamma + s(t_i - \tau)\right) \\ &= \underset{\tau\in(0,T)}{\operatorname{argmax}} \sum_{i=1}^{M} \log\left(1 + \frac{1}{\gamma}s(t_i - \tau)\right), \quad M > 0, \end{aligned} \tag{4.1}$$

where by subtracting $M \log\gamma$ from the first line and noticing that $M \log\gamma$ is constant with respect to τ, the second line follows. Also, $\widehat{\tau}$ denotes the time shift estimate when there are M photons in the observation interval $(0, T]$ with arrival times $\{t_1, \ldots, t_M\}$. In the case of observing no photons, a number is picked uniformly randomly from the interval $(0, T)$ which is the feasible set for τ_0.

Let us denote the Poisson arrival process over the observation interval by $P_{(0,T]}$, where the subscript is the observation interval and according to our model for the arrival times at the output of detector, $P_{(0,T]} \sim Poi\left(\gamma + s(t - \tau_0)\right)$. Now, because the background noise, γ, is assumed to be independent of the source signal $s(t)$ and the scene depth τ_0, the rate of $P_{(0,T]}$ is the sum of two independent terms and from the probability theory, it is known that this point process can be constructed by merging two independent Poisson processes with rates γ and $s(t - \tau_0)$. To be more precise, we

[1] If there is timing jitter in the detector, we simply substitute $s(t)$ with $\tilde{s}(t)$ which is the convolution of the source signal and the jitter pdf; however, in this chapter, to avoid carrying the tilde sign, we work only with $s(t)$.

have

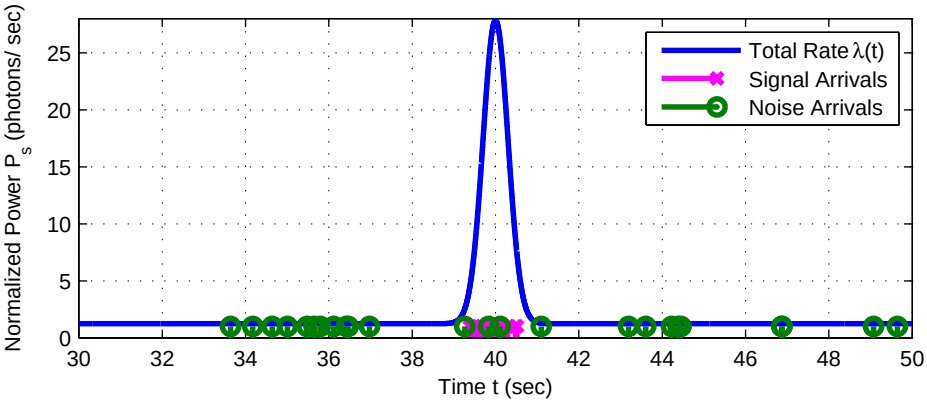
$$P_{(0,T]} \sim \Big\{ Poi(\gamma) \cup Poi(s(t-\tau_0)) \Big\}, \tag{4.2}$$

where the arrival times in (4.1), $\{t_i\}_{i=1}^{M}$, are a realization of $P_{(0,T]}$. Using (4.2), one can say the arrival times have labels "γ" and "s" based on what Poisson process they have come from; however, these labels are not known at the detector. Arrivals which come from $Poi(\gamma)$ are uniformly randomly spread in the whole observation interval while the arrivals coming from $Poi(s(t-\tau_0))$ are clustered around $t=\tau_0$ and they are spread in an interval whose length is equal to the width of the source signal. For example, if the source signal is a Gaussian pulse with variance parameter σ^2, the interval in which the arrivals labeled "s" are located has an approximate length of $6\,\sigma$ (neglecting the tails of Gaussian pulse) and is centered at $t=\tau_0$.

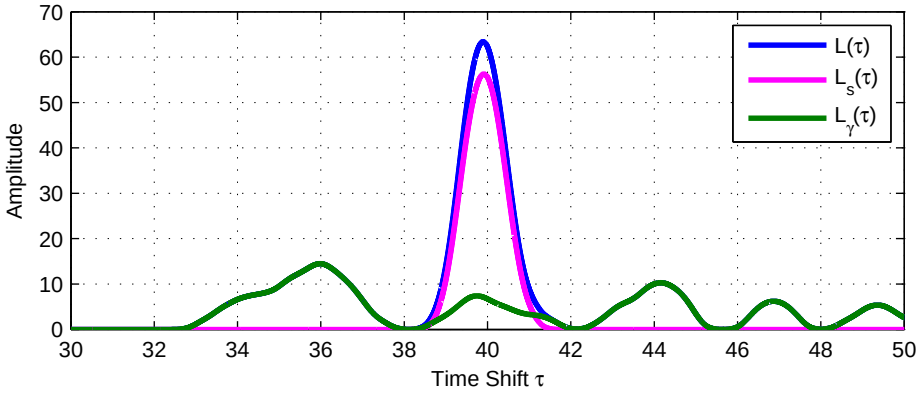
Using the argument in the previous paragraph about arrivals with different labels, we can reorder the arrival times $\{t_1,\ldots,t_M\}$ to get $\left\{t_1^{(s)},\ldots,t_{M_s}^{(s)},t_1^{(\gamma)}\ldots,t_{M_\gamma}^{(\gamma)}\right\}$, where we have assumed there are M_s arrivals with label "s" and $M_\gamma = M - M_s$ arrivals from the background noise with label "γ". The arrival times with superscript (s) are the photon arrival times due to the source signal while superscript (γ) means the arrivals are from the background noise. Also, without loss of generality, let us assume the arrival times within each group $\left\{t_i^{(s)}\right\}_{i=1}^{M_s}$ and $\left\{t_i^{(\gamma)}\right\}_{i=1}^{M_\gamma}$ are in an ascending order. However, there is no specific relation between two arrival times one from each set. Now, let us take a closer look at the cost function in (4.1) and see if we can use this notion of having two different sets of arrivals to get some insight into how solving (4.1) results in an estimate for the time shift. We have the following decomposition for the cost function in (4.1) which we call $L(\tau)$:

$$\begin{aligned} L(\tau) &\triangleq \sum_{i=1}^{M} \log\left(1+\frac{1}{\gamma}s(t_i-\tau)\right) \\ &= \sum_{i=1}^{M_s} \log\left(1+\frac{1}{\gamma}s(t_i^{(s)}-\tau)\right) + \sum_{i=1}^{M_\gamma} \log\left(1+\frac{1}{\gamma}s(t_i^{(\gamma)}-\tau)\right) \\ &= L_s(\tau) + L_\gamma(\tau), \end{aligned} \tag{4.3}$$

where the first line is just the definition of the cost function in (4.1). The first term in the second line is called L_s and is constructed using the arrivals from the signal. The second summation in the second line, called L_γ, is the noise component of the cost



(a)



(b)

Figure 4.1. (a) The rate of $P_{(0,T]}$ which consists of $\gamma = 1.25$ and $s(t - \tau_0)$ where $s(t)$ is a Gaussian pulse with $E_s = 20, \sigma = 0.3$ and the scene depth is $\tau_0 = 40$. The arrival times belong to a sample function of $P_{(0,T]}$ and are colored according to their labels where the noise arrivals are in green and the signal arrivals are in magenta. (b) The cost function $L(\tau)$ for the arrivals depicted in (a) and its signal and noise components, $L_s(\tau)$ and $L_\gamma(\tau)$, respectively.

function because it is evaluated using the arrivals caused by the background noise. It should be noted that we are not going to use the decomposition of (4.3) to estimate the time shift because the labels cannot be known at the detector. However, we introduced that decomposition to analyze how the log likelihood function looks like and what the effect of background noise on the estimation problem is.

The two panels in Fig. 4.1 represent an example where the signal is a Gaussian pulse with standard deviation $\sigma = 0.3$ and normalized energy of 20 photons per second. The scene depth is $\tau_0 = 40$ and there is assumed to be background noise, $\gamma = 1.25$ photons

per second. Therefore, the arrival times are from the Poisson process $Poi(\gamma + s(t-\tau_0))$. Here, it should be noted that the arrivals are generated by merging the arrivals of independent Poisson processes $Poi(\gamma)$ and $Poi(s(t-\tau_0))$ which implies the merged arrivals are from $Poi(\gamma + s(t-\tau_0))$ as was stated in (4.2). In Fig. 4.1(a), we have drawn the rate of this Poisson process which is $\gamma + s(t-\tau_0)$ with the observed arrivals in the interval $(30, 50)$ where the arrivals due to the signal component of the rate (referred to as "signal" arrivals) are marked in magenta and the arrivals caused by the background noise (referred to as "noise" arrivals) are in green. It is observed that the noise arrivals are present in the whole observation interval while the signal arrivals are only within the signal pulse boundaries. Also, in Fig. 4.1(b), the cost function $L(\tau)$ marked in blue and its signal and noise components, $L_s(\tau)$ (marked in magenta) and $L_\gamma(\tau)$ (marked in green), respectively are shown. As it can be seen in Fig. 4.1(b), the signal component of the cost function has a peak which is very close to the true underlying value of depth and its noise component has a random shape with a lot of peaks which are smaller than the signal component peak. Therefore, using maximum likelihood of (4.1), we are actually using the kernel $\log\left(1 + \frac{1}{\gamma}s(-\tau)\right)$ which with high probability has a totally different behavior for the set of signal arrivals (resulting in $L_s(\tau)$) than for the set of noise arrivals (resulting in $L_\gamma(\tau)$). Heuristically, this discrimination between the two types of arrivals is what is expected from a good estimator and is the core of designing an estimator for this problem. One may want to start from this point and devise a more computationally-efficient estimation algorithm than the maximum likelihood of 4.1, however, we do not investigate other estimators in this work.

■ 4.1.2 Only One Tilted Plane in the Scene

In the previous part, we introduced the maximum likelihood framework to estimate the depth of the scene where the scene was assumed to consist of only one fronto-parallel plane to the reference plane. In this section, we consider the second case, where the scene is made up of only one slanted plane and therefore, from Section 3.4.2, its depth function is $f(x, y) = \tau_0 - \Delta_0\left(x - \frac{1}{2}\right)$ for $(x, y) \in [0, 1]^2$ where the two parameters τ_0 (referred to as "time shift") and Δ_0 (referred to as "depth spread") are unknown. Also, we recall that the scene impulse response in this case is a box function centered at $t = \tau_0$ and has a width of Δ_0. To be more precise, from Section 3.4.2, the impulse response is as the following

$$h(t) = \frac{1}{\Delta_0}\,\mathrm{rect}\left(\frac{t-\tau_0}{\Delta_0}\right). \tag{4.4}$$

In this case, the scene impulse response $h(t)$ integrates to one regardless of the values of τ_0 and Δ_0 and we have $\lambda_0\left(t;\{\tau_0,\Delta_0\}\right) = s(t) * h(t)$ and it is straightforward to observe

$$\begin{aligned} E_s &= \int \lambda_0\left(t;\{\tau_0,\Delta_0\}\right)\mathrm{d}t = \int s(t) * h(t)\mathrm{d}t \\ &= \int s(t)\mathrm{d}t \int h(t)\mathrm{d}t = \int s(t)\mathrm{d}t, \end{aligned} \tag{4.5}$$

where we have used the identity that the area under the convolution of two nonnegative signals is the same as the product of the areas under each signal to get the second line. And, in the end, we have used the fact that $h(t)$ integrates to one regardless of the values of its parameters. Therefore, from (4.5), we know that the normalized received energy is the same as the energy of the source signal and hence, we call it E_s. Now, because this normalized energy E_s does not vary with changing the value of the impulse response parameters, we can again use the simplified maximum likelihood of (2.7) with α being replaced by (τ,Δ). Before formulating the maximum likelihood problem, let us define the received signal, for the case where the depth spread of the scene is assumed to be Δ, as the following:

$$r_\Delta(t) = s(t) * \frac{1}{\Delta}\,\mathrm{rect}\left(\frac{t}{\Delta}\right), \tag{4.6}$$

where we can use $r_\Delta(t)$ to formulate the maximum likelihood problem as the following:

$$\begin{aligned} \left(\widehat{\tau},\widehat{\Delta}\right) &= \underset{\substack{\tau\in(0,T)\\ \Delta\in[0,\Delta_{\max}]}}{\operatorname{argmax}} \sum_{i=1}^{M} \log\left(\gamma + r_\Delta(t_i-\tau)\right) \\ &= \underset{\substack{\tau\in(0,T)\\ \Delta\in[0,\Delta_{\max}]}}{\operatorname{argmax}} \sum_{i=1}^{M} \log\left(1+\frac{1}{\gamma}r_\Delta(t_i-\tau)\right), \quad M>0, \end{aligned} \tag{4.7}$$

where we have subtracted $M\log\gamma$ from the first line to get the second line because that term is constant with respect to (τ,Δ) and can be omitted from the cost function. $\Delta_{\max}$ is the maximum depth spread we consider for the scene. Like before, if there is no arrival observed, $(\widehat{\tau},\widehat{\Delta})$ is a pair randomly picked from the feasible set for (τ,Δ).

The first point to note is that $r_\Delta(t)$ given in (4.6) reduces to $s(t)$ and (4.7) becomes (4.1) in the case where $\Delta = 0$ which makes the model of (4.4) contain the model introduced in the previous section which we called the conventional model. This is a key part of the present work where we have introduced a richer model that contains

the conventional model and therefore, it is expected to perform at least as well as the conventional model. This is going to be observed in Section 4.2.2 where we compare the estimates obtained using our model with the results obtained by employing the conventional model.

The second important note to make is that one might ask how using a maximum likelihood framework as in (4.7) is justified while our ultimate goal is estimating the depth function (or depth map)[2]. In other words, how do we want to justify the solution in (4.7) yields a good estimate for depth map (i.e., it minimizes the MSE of the reconstructed depth map)? Let us first write the MSE of the depth map reconstruction, denoted as $\mathrm{MSE}_{\mathrm{DM}}$, as the following:

$$\mathrm{MSE}_{\mathrm{DM}} \triangleq E\left[\int_0^1 \int_0^1 \left(\widehat{b}(x,y) - b(x,y)\right)^2 \mathrm{d}x\,\mathrm{d}y\right], \tag{4.8}$$

where the expectation is taken over all possible arrival times (the sample space of arrival times) and $b(x,y)$ and $\widehat{b}(x,y)$ represent the brightness of the point (x,y) on the depth map and its estimate, respectively and are constructed as below:

$$b(x,y) = 1 - \frac{1}{T}\left[\tau_0 - \Delta_0\left(x - \frac{1}{2}\right)\right], \quad (x,y) \in [0,1]^2, \tag{4.9}$$

$$\widehat{b}(x,y) = 1 - \frac{1}{T}\left[\widehat{\tau} - \widehat{\Delta}\left(x - \frac{1}{2}\right)\right], \quad (x,y) \in [0,1]^2. \tag{4.10}$$

Now, by plugging (4.9) and (4.10) into (4.8) we get

$$\mathrm{MSE}_{\mathrm{DM}} = \frac{1}{T^2} E\left[(\widehat{\tau} - \tau_0)^2 + \frac{1}{12}(\widehat{\Delta} - \Delta_0)^2\right] = \frac{1}{T^2}\left(\mathrm{MSE}_\tau + \frac{1}{12}\mathrm{MSE}_\Delta\right) \tag{4.11}$$

where $\mathrm{MSE}_\tau \triangleq E\left[(\widehat{\tau} - \tau_0)^2\right]$ and $\mathrm{MSE}_\Delta \triangleq E\left[(\widehat{\Delta} - \Delta_0)^2\right]$. By inspecting (4.11), we see that minimizing MSE_τ and MSE_Δ results in minimizing $\mathrm{MSE}_{\mathrm{DM}}$. Therefore, if one wants to minimize the MSE in the reconstruction of the depth map, they should employ an algorithm which minimizes MSE_τ and MSE_Δ. It is very well known that the posterior mean is the optimal estimator when the cost function is the squared error. Also, it is known that the maximum likelihood problem in (4.7) is equivalent to the maximum a posteriori problem if the prior joint distribution for τ_0 and Δ_0 is uniform over the feasible set. Furthermore, for high enough E_s, the mean of the posterior

[2]Because there is a one-to-one relation between the depth function and depth map as was discussed in Section 3.1, from now on, we only consider the depth map.

distribution is very close to its maximum. Therefore, loosely speaking, the maximum likelihood framework of (4.7) gives an optimal estimate for high enough normalized received energies which will be observed in Section 4.2.2.

■ 4.2 Performance Analysis of the Estimation Algorithms

In this section, we analyze the performance of the maximum likelihood (ML) methods proposed in the previous section. In the following section, we consider the scenario where the scene has only one front-parallel plane and hence, the ML problem is a one dimensional optimization problem and in Section 4.1.2, the situation where both time shift and the scene spread are unknown is investigated.

■ 4.2.1 Only Time Shift Is Unknown

In this part, we consider the maximum likelihood problem of (4.1) and its solution $\widehat{\tau}$ to which we refer as the ML estimate of time shift. Two of the most common performance measures for the estimate $\widehat{\tau}$ are its bias and mean square error defined as the following:

$$\text{Bias}_\tau \triangleq E[\widehat{\tau} - \tau_0], \tag{4.12}$$

$$\text{MSE}_\tau \triangleq E\left[(\widehat{\tau} - \tau_0)^2\right], \tag{4.13}$$

where the expectations are over the Poisson arrival times whose joint probability distribution has the true underlying time shift τ_0 as its parameter.

Now, let us start with assuming there is no background noise in the experiment, i.e., $\gamma = 0$, and the source signal, $s(t)$, is a Gaussian pulse with parameter σ. Therefore, we get the following maximum likelihood problem:

$$\begin{aligned}\widehat{\tau} &= \underset{\tau\in(0,T)}{\text{argmax}} \sum_{i=1}^{M} \log s(t_i - \tau) = \underset{\tau\in(0,T)}{\text{argmax}} \sum_{i=1}^{M} \log\left(\frac{E_s}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t_i-\tau)^2}{2\sigma^2}}\right) \\ &= \underset{\tau\in(0,T)}{\text{argmax}} \sum_{i=1}^{M} -(t_i - \tau)^2 = \frac{1}{M}\sum_{i=1}^{M} t_i,\end{aligned} \tag{4.14}$$

where the second line follows directly from the first line by omitting the constant terms in the first line. It should also be noted that the optimization problem in the second line is straightforward to solve and its analytical solution is the average of arrival times. Because we are assuming a Gaussian source signal here and the scene impulse response is an impulse and there is no background noise, we have $\lambda(t) = s(t - \tau_0)$ which is

a Gaussian pulse centered at τ_0 and from probability theory [10], it is known that given the number of arrivals in the observation interval is m, the arrival times can be considered as m independent samples from a distribution whose pdf is $\lambda(t)$ multiplied by a normalization constant so it integrates to one. Therefore, in this case, given that the number of arrivals is $m > 0$, the arrival times are independent Gaussian random variables with mean equal to τ_0 and a variance of σ^2. Consequently, the estimator of (4.14) (the average of arrival times) has a mean equal to τ_0 and a variance of $\frac{\sigma^2}{m}$. Now, we can easily calculate the bias of $\widehat{\tau}$ given in (4.14) as the following:

$$\begin{aligned}\text{Bias}_\tau &\triangleq E\left[\widehat{\tau} - \tau_0\right] \overset{(a)}{=} E_M\left[E[\widehat{\tau} - \tau_0|M]\right] \\ &= (\frac{T}{2} - \tau_0)e^{-E_s} + \sum_{m=1}^{\infty} e^{-E_s}\frac{E_s^m}{m!}(\tau_0 - \tau_0) \\ &= (\frac{T}{2} - \tau_0)e^{-E_s}, \end{aligned} \tag{4.15}$$

where (a) is simply because of the law of "iterated expectations" from the probability theory [10] and M is the random variable representing the number of arrivals which is Poisson with mean equal to E_s. Now, if there is no arrival $M = 0$, a number is picked randomly from the interval $(0, T)$. This event has a probability of e^{-E_s} and the mean of the estimate in this case is $\frac{T}{2}$ and hence, we have the first term in the second line. For nonzero number of arrivals, as we discussed earlier, the mean of the estimate is τ_0, hence, we have the term $\tau_0 - \tau_0$ in the summation. From (4.15), for the case where $\tau_0 > \frac{T}{2}$ (as in our simulation results presented later in this section), we conclude that the bias of the estimator is always negative and decays to zero exponentially with increasing E_s. Having calculated the bias of the estimate given in (4.14), let us now evaluate its MSE as the following:

$$\begin{aligned}\text{MSE}_\tau &\triangleq E\left[(\widehat{\tau} - \tau_0)^2\right] \overset{(a)}{=} E_M\left[E\left[(\widehat{\tau} - \tau_0)^2|M\right]\right] \\ &\overset{(b)}{=} \nu^2 e^{-E_s} + \sum_{m=1}^{\infty} e^{-E_s}\frac{E_s^m}{m!}\frac{\sigma^2}{m} \\ &= e^{-E_s}\left\{\nu^2 + \sigma^2\int_0^{E_s}\frac{e^x - 1}{x}\mathrm{d}x\right\}, \end{aligned} \tag{4.16}$$

where (a) is iterated expectations, (b) is simply the expectation over M where the first term $(\nu^2 e^{-E_s})$ corresponds to the case where there is no arrival and the summation corresponds to the estimate in (4.14). If there is no arrival observed, $\nu^2 = \frac{T^2}{12} + (\frac{T}{2} - \tau_0)^2$

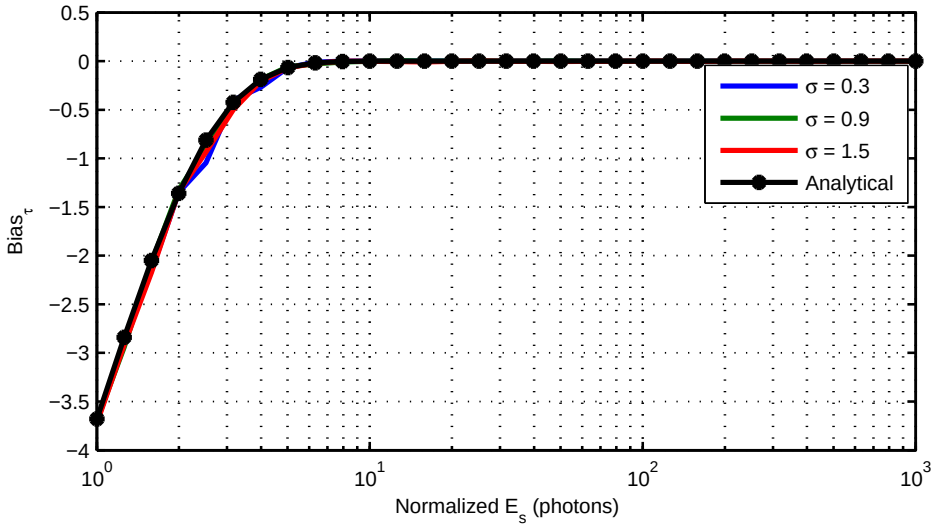
where the first term is the variance of a uniform variable with support $(0, T)$ and the second term is because the mean of that uniform random variable is not the same as τ_0 (i.e., nonzero bias). The last line in (4.16) follows from the second line and using the identity $\sum_{m=1}^{\infty} \frac{1}{m} \frac{E_s^m}{m!} = \int_0^{E_s} \frac{e^x - 1}{x} \mathrm{d}x$. Now, by taking $10 \log_{10}$ of the MSE_τ in (4.16), we get MSE_τ in dB as the following:

$$\mathrm{MSE}_\tau(dB) = \frac{10}{\log 10} \left\{ \log \nu^2 - E_s + \log \left(1 + \frac{\sigma^2}{\nu^2} \int_0^{E_s} \frac{e^x - 1}{x} \mathrm{d}x \right) \right\}, \tag{4.17}$$

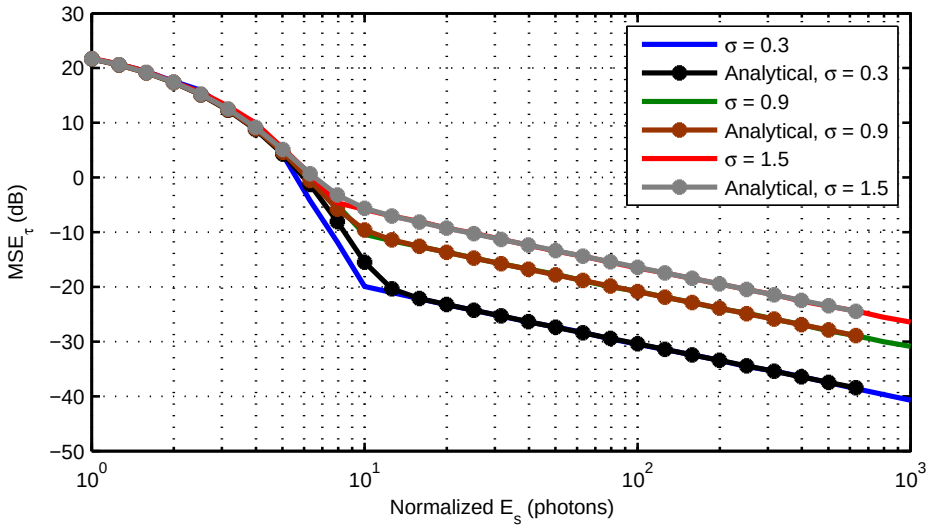
where log denotes the natural logarithm.

In Fig. 4.2, we have compared the numerical bias and mean square error of the time shift estimate to the analytical results of (4.15) and (4.17) when the source signal $s(t)$ is Gaussian with parameter σ. The numerical results are obtained using Monte-Carlo simulations with $10,000$ independent trials where $\tau_0 = 40$. Also, we have considered three different values $(0.3, 0.9, \text{and } 1.5)$ for the source pulse width, σ. It should be noted that because the bias in (4.15) is independent of σ, we have drawn the analytical curve only once in Fig. 4.2(a). The first observation is that the analytical formulas of (4.15) and (4.17) are completely precise for the case where there is no background noise as in Fig. 4.2. In Fig. 4.2(b), two different regions for the MSE_τ are observed. For small E_s's, MSE_τ is exponential in $\log E_s$ because the first two terms in (4.17) dominate the third term. However, mean square error has a linear behavior in $\log E_s$ after a threshold around $E_s = 10$ and this is because for large enough E_s's, the third term in (4.17) is the dominant term which is linear in $\log E_s$. To be more precise, for large enough E_s, $\mathrm{MSE}_\tau \approx \frac{\sigma^2}{E_s}$ which results in a linear $\mathrm{MSE}_\tau(dB)$ with respect to $\log E_s$. It should also be mentioned this linear behavior of MSE_τ is the same as the asymptotic performance that Bar-David [5] derived for the ML estimate and we presented in (2.10). We should also note that the Cramér-Rao lower bound (CRB) is the same as the asymptotic behavior in (2.10). Therefore, when there is no background noise, the ML estimate in (4.14) reaches the CRB for high enough E_s (E_s's larger than the threshold). Also, as was stated above and Bar-David pointed out in [5], the linear region is proportional to σ^2 which justifies the intuition behind using a narrower source signal to get a better time shift estimate.

It is also observed that the bias of the time shift estimate in Fig. 4.2(a) goes to zero at an energy level ($E_s \approx 7$) smaller than the energy required to reach the linear region in Fig. 4.2(b) which is $E_s \approx 10$. Now, let us analyze why this phenomenon happens. For the signal energies between these two thresholds ($7 < E_s < 10$), e^{E_s} is large enough



(a) Bias of the ML estimate for the time shift.



(b) MSE of the ML estimate for the time shift.

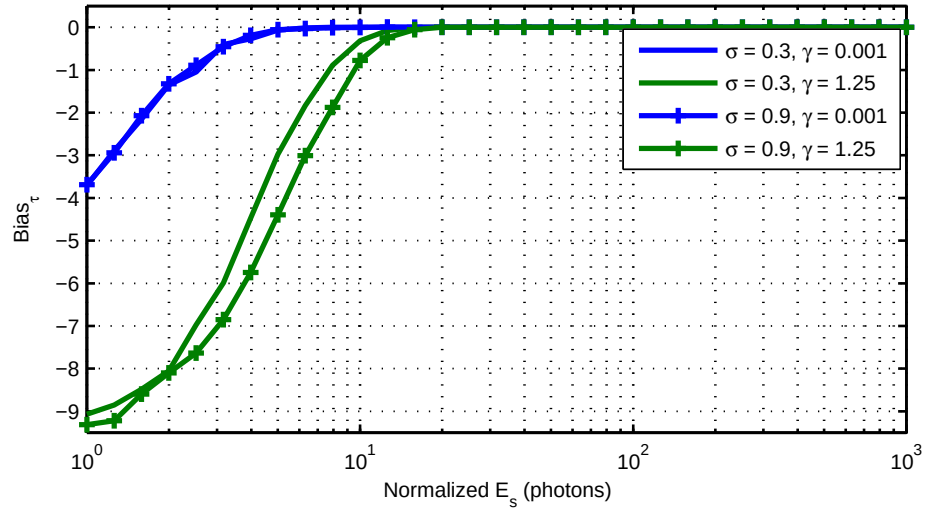
Figure 4.2. Numerical bias and mean square error (MSE) in dB of the time shift estimate for three different pulse widths ($\sigma = 0.3, 0.9, 1.5$) using Monte-Carlo simulations with $10,000$ independent trials along with the analytical bias and MSE. The source signal is a Gaussian pulse with parameter σ and we have $\tau_0 = 40$, $T = 60$ and $\gamma = 0.001$.

relative to $\left|\frac{T}{2} - \tau_0\right|$ that makes the bias in (4.15) almost zero while for MSE_τ in (4.16), $\sigma^2 \int_0^{E_s} \frac{e^x - 1}{x} \mathrm{d}x$ is not significantly larger than $\nu^2 = (\frac{T}{2} - \tau_0)^2 + \frac{T^2}{12}$ and therefore, we see

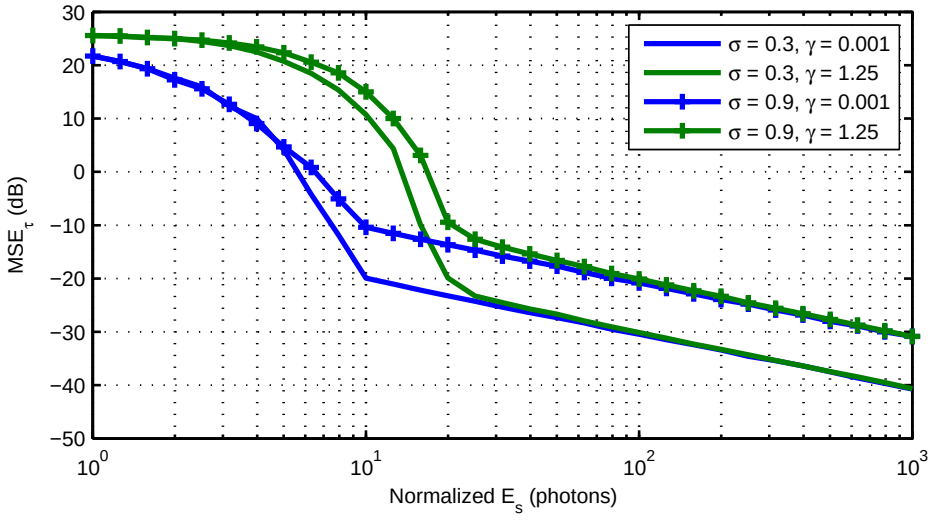
a nonlinear behavior in MSE_τ for those intermediate signal energies in Fig. 4.2(b). In other words, for the intermediate signal energies, although the probability of getting an estimate far from the true time shift is very small, in MSE_τ, the squared error due to that estimate is big and cannot be ignored.

Now, let us consider the maximum likelihood estimate in (4.1) where there is background noise, γ. Like before, let us consider a Gaussian source signal with parameter σ and analyze how the ML estimate of the time shift behaves for different source signal energies. The bias and MSE of the time shift estimate are depicted in Fig. 4.3 where we have used Monte-Carlo simulations with $10,000$ independent trials to calculate the bias and MSE. In Fig. 4.3, we have considered two values for the σ parameter of the Gaussian source ($\sigma = 0.3, 0.9$) and two values for the background noise level ($\gamma = 0.001, 1.25$), where $\gamma = 0.001$ is the same case as in Fig. 4.2 which we refer to as the noiseless case. First of all, it is observed in Fig. 4.3(a) that the bias for the noisy case like in the noiseless case is always negative and decays to zero for large enough E_s's. This is because with some probability, say β, $L_\gamma(\tau)$ in (4.3) has a maximum larger than the maximum of $L_s(\tau)$ in (4.3) and this results in an estimate which has a uniform distribution over the observation interval $(0, T)$. Therefore, with probability equal to β, the mean of the time shift estimate is $\frac{T}{2}$ and with probability $1-\beta$, the peak of the signal component of the log likelihood function, i.e., $L_s(\tau)$, is larger than the peaks in $L_\gamma(\tau)$ and therefore, the mean of the time shift estimate is τ_0 (because the source signal is symmetric around τ_0). Assuming $\tau_0 = 40$ and $T = 60$ as in Fig. 4.3, we get a negative bias equal to $\beta(\frac{T}{2} - \tau_0)$ where β decreases with E_s. In the noisy problem, as opposed to the noiseless problem, β depends on the width of the source signal and is a larger number for wider signals when the signal energy is held constant. This is why, in Fig. 4.3(a), we observe a bigger bias for $\sigma = 0.9$ than the $\sigma = 0.3$ scenario when there is noise in the experiment ($\gamma = 1.25$). Also, β is surely larger than e^{-E_s} because we see larger bias for the noisy case than the bias in the noiseless case in Fig. 4.3(a).

The second difference between the noiseless and noisy problem is that the source signal energy for which the bias goes to zero in the noisy problem is bigger than the threshold in the noiseless problem and it is because β is bigger than e^{-E_s} and therefore, we have a situation where $e^{-E_s}(\frac{T}{2}-\tau_0)$ is very close to zero while $\beta(\frac{T}{2}-\tau_0)$ is significantly nonzero. The same phenomenon happens for MSE_τ in Fig. 4.3(b). Also, the same relation between the threshold for MSE_τ and the threshold for Bias_τ holds for the noisy case as we saw for the noiseless case; i.e., the bias threshold happens at smaller E_s while the threshold for MSE_τ happens for larger E_s.



(a) Bias of the ML estimate for the time shift.



(b) MSE of the ML estimate for the time shift.

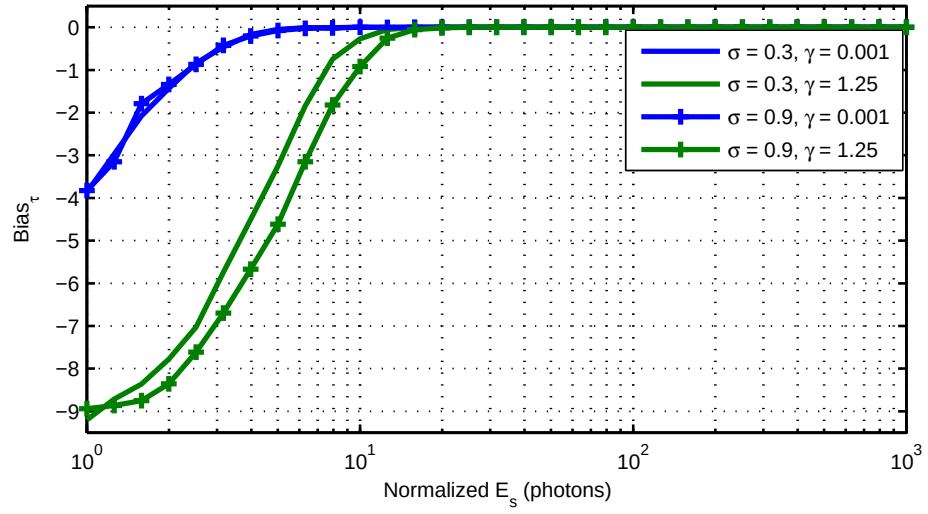
Figure 4.3. Bias and MSE in dB for the time shift estimate for two different pulse widths ($\sigma = 0.3,\ 0.9$) and two different background noise levels ($\gamma = 0.001,\ 1.25$) using Monte-Carlo simulations with $10,000$ independent trials. The source signal is a Gaussian pulse with parameter σ and we have $\tau_0 = 40$, $T = 60$.

The last point to make about the MSE_τ is that for large enough E_s's, the existence of background noise γ in the experiment does not matter and the performance of the ML estimate of the time shift is equivalent to the performance of the noiseless ML

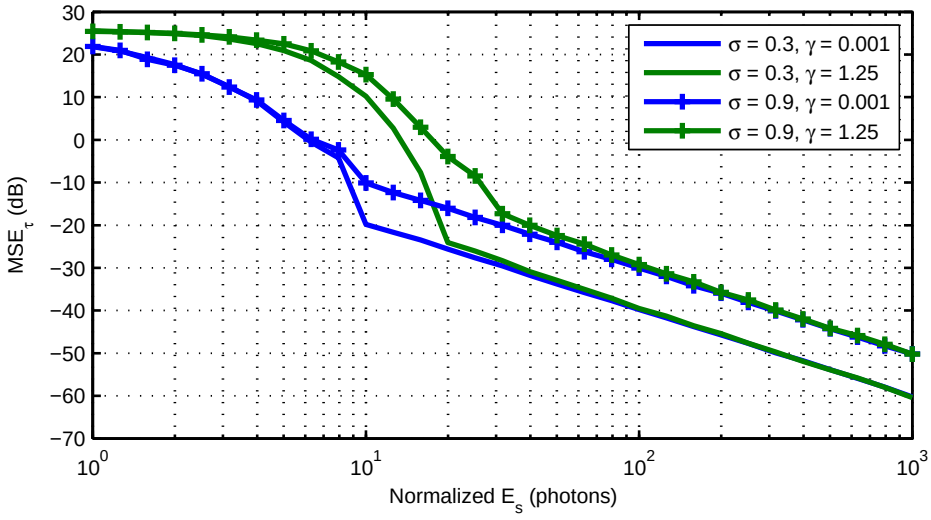
estimate which is the Cramér-Rao lower bound as well and for a Gaussian source is $\text{MSE}_\tau \approx \frac{\sigma^2}{E_s}$ as stated earlier in this section. Therefore, the ML estimator for the time shift reaches the CRB in both noiseless and noisy cases for large enough signal energy, E_s. Finally, we should note that because for low E_s, the probability of getting large errors (as we call it β) increase with σ and for large E_s, MSE_τ is proportional to σ^2, it is desired to have a narrower source pulse as the intuition suggests.

Having investigated different aspects of the estimation problem where the source signal is a Gaussian pulse, we can now discuss the case where the source signal is a rectangular signal. As Israel Bar-David in [8] pointed out, the MSE_τ dependence on the signal energy in the linear region of MSE_τ is different than what we observed in Fig. 4.3 where the source signal was a Gaussian pulse. For signals which do not have a flat top (by flat top, we mean having zero derivative over an interval like in a box function which has a "flat" top), MSE_τ for large E_s is proportional to E_s^{-1} as we derived in (2.10) which is the Cramér-Rao lower bound. However, as Bar-David discussed in [8], for a rectangular source signal, the mean square error obtained in the time shift estimation using maximum likelihood framework is proportional to E_s^{-2} for large enough E_s's. In Fig. 4.4, we have shown the numerically evaluated bias and the mean square error for two different background noise levels ($\gamma = 0.001,\ 1.25$) and two different width parameters ($\sigma = 0.3,\ 0.9$) where the source signal is a rectangular signal and is nonzero over an interval with length $\sigma\sqrt{12}$. We ran $10,000$ independent iterations for the results shown in Fig. 4.4 where the true time shift is $\tau_0 = 40$ and the length of the uncertainty interval for the time shift is $T = 60$. It is clearly observed that MSE_τ is proportional to E_s^{-2} in the linear region. Everything else that we discussed for the Gaussian case is the same in the rectangular pulse case.

It should be noted that according to the simulations we have presented up until now, the ML estimator in (4.1) can be thought of as a denoising algorithm followed by an optimal estimator to find the mean of a distribution where the pdf associated to the distribution has the same shape as the source signal. If a Gaussian signal is being utilized, the efficient estimator of the mean is the empirical average as we derived in (4.14) and the empirical average has a variance inversely proportional to the number of samples (arrivals here). The number of arrivals is a Poisson random variable with mean E_s and therefore, the E_s^{-1} behavior in the MSE_τ is observed. However, if a rectangular pulse is being employed as the source signal, the optimal estimator for the mean of the distribution (which is uniform in this case) is the mid range of the arrival times whose variance decays proportional to M^{-2} where M is the number of arrivals



(a) Bias of the ML estimate for the time shift.



(b) MSE of the ML estimate for the time shift.

Figure 4.4. Bias and MSE in dB for the time shift estimate for two different pulse widths ($\sigma = 0.3,\ 0.9$) and two different background noise levels ($\gamma = 0.001,\ 1.25$) using Monte-Carlo simulations with $10,000$ independent trials. The source signal is a rectangular pulse with parameter σ and we have $\tau_0 = 40$, $T = 60$.

(samples) and therefore, we observe the E_s^{-2} behavior in the linear region of MSE_τ. Interpreting the ML estimate as a denoising algorithm followed by the optimal mean estimator, we can easily see why the ML estimate reaches the CRB which is the lowest

possible MSE one can hope for. Using the same interpretation, one can argue that if the signal $s(t)$ we formulate the maximum likelihood problem in (4.1) with is different from the true underlying source signal which has generated the arrivals, we have a situation where we use a suboptimal mean estimator for a distribution. For example, if the real source signal is Gaussian and we use a maximum likelihood framework with $s(t)$ being a rectangular pulse, we expect to have MSE_τ proportional to $(\log E_s)^{-1}$ rather than E_s^{-1} because when the maximum likelihood uses a rectangular $s(t)$, we are taking the mid range of denoised arrival times as the estimate for their mean and for Gaussian arrival times, it is known that the variance of the mid range is inversely proportional to $\log M$ where M is the number of samples (arrivals).

■ 4.2.2 Time Shift and Depth Spread Are Unknown

In this section, we analyze the performance of the maximum likelihood algorithm of (4.7) for the joint estimation of time shift and depth spread of the scene impulse response. We also recall that the maximum likelihood problem in (4.7) is the direct result of using the parametric model for the scene impulse response given in (4.4) which we originally derived in Section 3.4.2 to model a possibly tilted planar scene. Clearly, if the planar scene is fronto-parallel to the reference plane, the conventional model, which assumes the scene impulse response is a time-delayed impulse, is accurate. In that case, our model reduces to the conventional model because in our model, when the depth spread Δ_0 approaches to zero, the box function reduces to an impulse (the depth spread is a measure of the plane tilt and zero depth spread corresponds to fronto-parallel plane). Therefore, for the case where the scene consists of only one plane parallel to the reference plane, we expect to get the same performance using either of these two models. However, when the scene is composed of a tilted plane, we expect to get a better performance by using our model than the conventional model because the conventional model in this case is mismatched to the true underlying scene impulse response which is a box function as in (4.4). Here, using simulation results, we investigate this difference between the conventional and our model.

Before, getting to simulation results for the algorithm, let us first define $L(\tau, \Delta)$ as the ML cost function in (4.7):

$$L(\tau, \Delta) \triangleq \sum_{i=1}^{M} \log\left(1 + \frac{1}{\gamma} r_\Delta(t_i - \tau)\right), \tag{4.18}$$

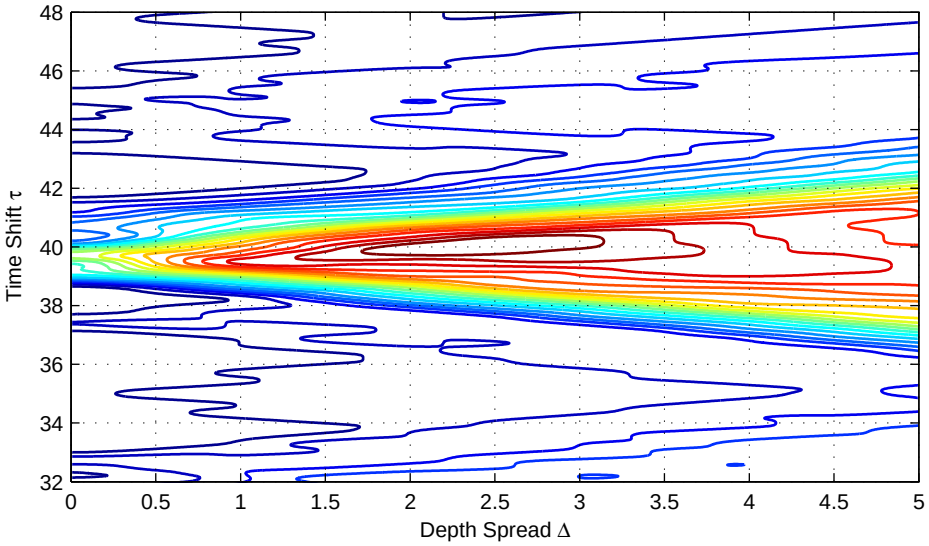


Figure 4.5. Contour plot of the log-likelihood function of the Poisson arrival times where $\tau_0 = 40$ and $\Delta_0 = 2$.

where $r_\Delta(t)$ is defined in (4.6) and is the convolution of the source signal $s(t)$ and a box function with width of Δ. It is worth mentioning that $L(\tau, \Delta)$ is highly non-convex with a lot of peaks and valleys and therefore, we could not use any gradient algorithms to solve the optimization problem in (4.7). We instead evaluated the cost function in (4.18) over a dense two dimensional grid (one dimension representing τ and the other denoting Δ) and picked the maximum by searching over that grid. This is the downside of our model which makes the optimization problem more computationally expensive because of introducing a new parameter (i.e., Δ) into the ML problem. However, if one can devise a more computationally efficient estimation algorithm than the brute-force search over the aforementioned grid, the proposed model in this work will greatly improve the depth estimation results as we will see later in this section. An example for the contour plot of the cost function $L(\tau, \Delta)$ is shown in Fig. 4.5, where a Gaussian source signal with $E_s = 20$ and $\sigma = 0.1$ is used and the parameters of the scene are $\tau_0 = 40$, $\Delta_0 = 2$. Also, background noise in the experiment is assumed to be $\gamma = 1.25$. The red color represents the largest function value and the blue color denotes the smallest value; it is clear that $L(\tau, \Delta)$ has a maximum close to the true underlying parameters, however, a lot of local minimums and maximums are observed in Fig. 4.5 as well.

In the previous section, we interpreted the ML estimator as a hypothetical denoising block followed by the optimal estimator for the mean of the distribution of arrival times.

We also recall that the ML estimate of (4.14) was exactly the optimal estimate for the mean of the distribution in the noiseless problem. Now, it should be noted that in the noiseless scenario (i.e., $\gamma = 0$) in this section, the pdf of the Poisson arrival times is $r_\Delta(t-\tau_0) = s(t) * h(t)$ multiplied by a normalization factor so the pdf integrates to one. Therefore, because the pdf of arrival times is the convolution of two signals, the arrival times can be considered as the sum of two independent random variables[3]; one of them has a pdf equal to source signal $s(t)$ multiplied by a normalization constant and the other random variable has a pdf equal to $h(t)$ which implies that random variable is a uniform random variable with mean τ_0 and with variance $\frac{\Delta_0^2}{12}$. Consequently, the mean of the arrival times is τ_0 because the mean of the random variable with pdf $s(t)$ is zero. Also, because the variance of sum of two independent random variables is equal to the sum of their variances, the variance of the arrival times is $\sigma^2 + \frac{\Delta_0^2}{12}$ assuming the source signal is Gaussian with a known width parameter σ. Using this interpretation, we can now say that the estimator we should use to solve our problem has to reduce to the optimal estimator for the mean and the variance of the Poisson arrival times (i.e., τ_0 and $\sigma^2 + \frac{\Delta_0^2}{12}$, respectively) in the noiseless problem. Also, the algorithm, in the noisy scenario, should be able to denoise the arrival times first and then find the mean and variance of the time of arrivals. If there exists such an estimator, then, it should achieve the Cramér-Rao lower bound for large enough signal energies which means enough number of samples from the underlying distribution. The good news is that solving the two dimensional ML problem in (4.7) gives us an estimate satisfying all the above requirements; i.e, the ML estimate of τ_0 and Δ_0 achieves the CRB for MSE_τ and MSE_Δ for large enough E_s's as shown in Figs. 4.6–4.9.

Now, let us consider the maximum likelihood problem of (4.7) with true underlying parameters $\tau_0 = 40$ and $\Delta_0 = 2$. Also, let us assume the depth uncertainty interval is $(0, T)$ where $T = 60$. We have run 1000 independent trials of Monte-Carlo simulations to numerically evaluate MSE_τ for the noiseless and noisy problems in Figs. 4.6 and 4.7, respectively, where the noisy problem has background noise with $\gamma = 1.25$. MSE_Δ is also plotted in Figs. 4.8 and 4.9 for the noiseless and noisy problem, respectively. In each of Figs. 4.6–4.9, a Gaussian pulse with parameter σ is used as the source signal and each panel corresponds to a different σ. The top panel of those figures represents the results where $\sigma = 0.1$ which means the source signal is much narrower than the scene impulse response (with $\Delta_0 = 2$). This yields to a huge gap between the performance of

[3]From probability theory [10], it is known that the pdf of the sum of two independent random variables is the convolution of their pdf's.

the conventional model and our model. As we observe in Figs. 4.6(a) and 4.7(a), the mean square error, obtained by using the conventional model, changes very slowly by increasing E_s and the reason behind this is that although increasing E_s gives us more samples, because the conventional model is a mismatched model to the true underlying impulse response, we cannot benefit from more samples. On the other hand, our model achieves the best possible performance predicted by CRB for large enough E_s's. We have also presented the results of another algorithm which we have called "oracle (Δ_0 known)". That is basically using maximum likelihood to estimate the time shift when the depth spread of the scene is exactly known at the estimator; i.e., we have solved the ML problem of (4.1) with $s(t)$ replaced by $r_{\Delta_0}(t)$. As we can see in all the panels of Figs. 4.6 and 4.7, the performance of our model, for high enough E_s, matches the performance of the ML algorithm using oracle and also the CRB prediction of the best performance.

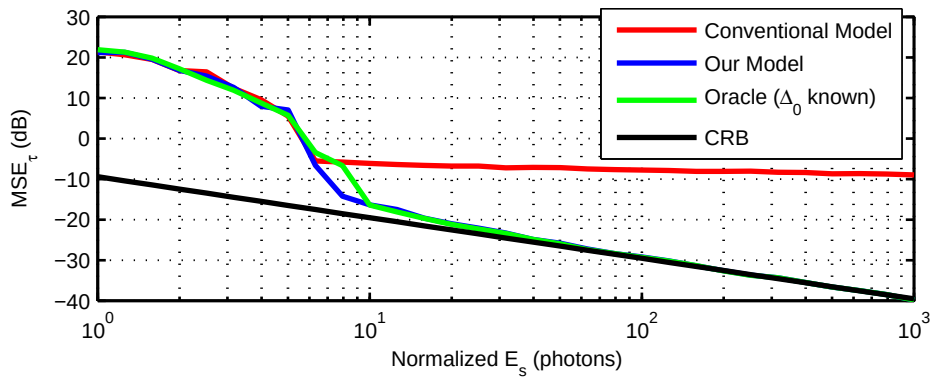
In the middle panels of Figs. 4.6 and 4.7, we have increased the σ parameter of the Gaussian source to 0.4 which implies the source is approximately 2.4 units wide which is almost the same as the depth spread of the scene, $\Delta_0 = 2$. In the noiseless problem in Fig. 4.6(b), the performance of all methods and CRB are almost the same up to some intermediate energy level ($E_s \approx 100$) and then the performance of conventional method starts deviating from the performance of other methods and CRB. This is because the Kullback-Leibler distance between the distributions of arrival times in the conventional model and our model is relatively small for $\sigma = 0.4$ and therefore, in order to distinguish these two distributions from each other, we need a lot of samples. Hence, as long as the signal energy which determines the number of samples is low or intermediate, there is no difference between the conventional model and our model. However, for large number of samples (corresponding to large E_s), we start distinguishing between the distribution obtained by using the conventional model and the distribution obtained by our model. In the noisy problem in Fig. 4.7(b), the performance of the conventional model is uniformly worse than the other models and CRB because the denoising part using the conventional model is not performed as well as it is done using the correct model.

In the bottom panels of Figs. 4.6 and 4.7, the σ of the Gaussian source has been increased to 1.6 which results in a very wide source signal relative to the true depth spread. Therefore, as it is observed, in both noiseless and noisy problems, there is no performance gap between the conventional model and our model and both match the performance of ML algorithm using oracle and the CRB. Therefore, as we have

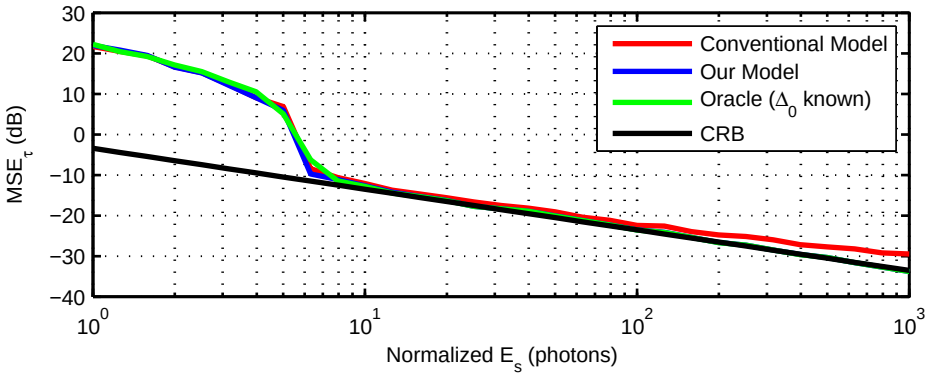
seen in Figs. 4.6 and 4.7, when the width of the Gaussian source (approximately $6\,\sigma$) is small relative to the depth spread, the conventional model has a very poor performance in comparison to our model which achieves the Cramér-Rao lower bound. As σ increases, the difference between the conventional model and our model gets more and more indistinguishable and therefore, we need more samples (large E_s) to get a better performance using our model relative to the conventional model. At some point, when the σ of Gaussian source is very large, we basically cannot distinguish between the conventional model and our model for the signal energies we have considered in our simulations. Therefore, the parameter σ of the source signal plays a key role in our estimation problem and puts a fundamental limit on how small Δ_0 can be in order to be distinguishable from $\Delta = 0$ which corresponds to the conventional model.

In Figs. 4.8 and 4.9, we have plotted the MSE_Δ for the noiseless and noisy problems, respectively. Everything is the same as for the time shift estimation in the sense that we lose precision as σ gets larger and our model achieves the CRB for large enough E_s's. However, in the depth spread estimation problem when the σ parameter of the source signal is small ($\sigma = 0.1$, Figs. 4.8(a) and 4.9(a)) and the signal energy is low or intermediate, our method is slightly worse than using the ML algorithm with the exact value of τ_0 (referred to as "oracle (τ_0 known)"). This was not the case in the time shift estimation problem where our method was always as well as having oracle regardless of the signal energy level. Also, when $\sigma = 1.6$, because $\Delta = 2$ and $\Delta = 0$ are indistinguishable, the mean square error of the depth spread estimate is relatively poor even for large signal energies as shown in Figs. 4.8(c) and 4.9(c). It should also be noted that the scale of the vertical axis in each of Figs. 4.6–4.9 is the same among the panels to emphasize that the estimate MSE gets worse as we increase σ in both time shift and depth spread estimation problems.

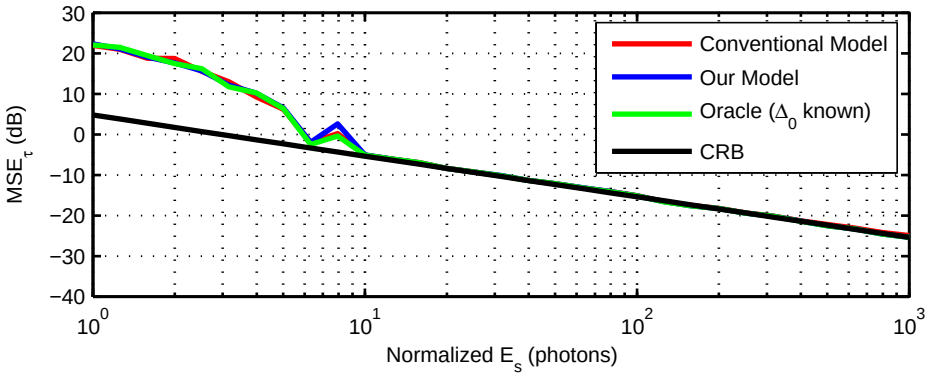
As the last remark in this chapter, we should note that the off-diagonal entries in the Fisher information matrix are zero (see Appendix A), therefore, the time shift and depth spread parameters, in theory, can be estimated independently which could help reduce the computation complexity by a large factor. However, because we do not have time to investigate this thoroughly, we consider it as a future work which we discuss more in the next chapter.



(a) Source signal is a Gaussian pulse with $\sigma = 0.1$.

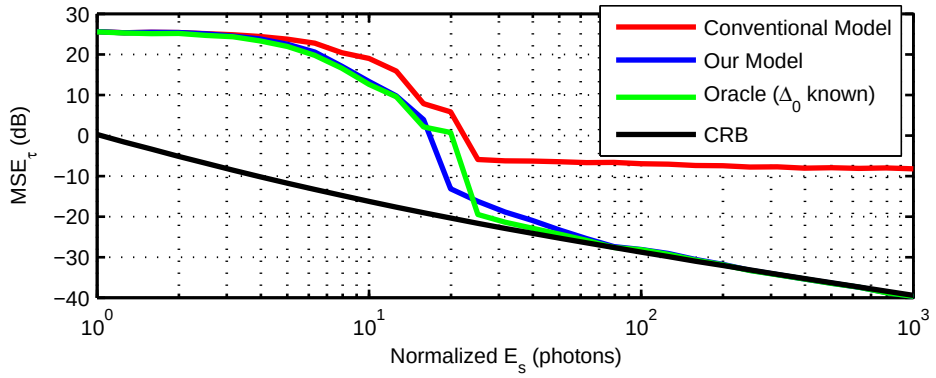


(b) Source signal is a Gaussian pulse with $\sigma = 0.4$.

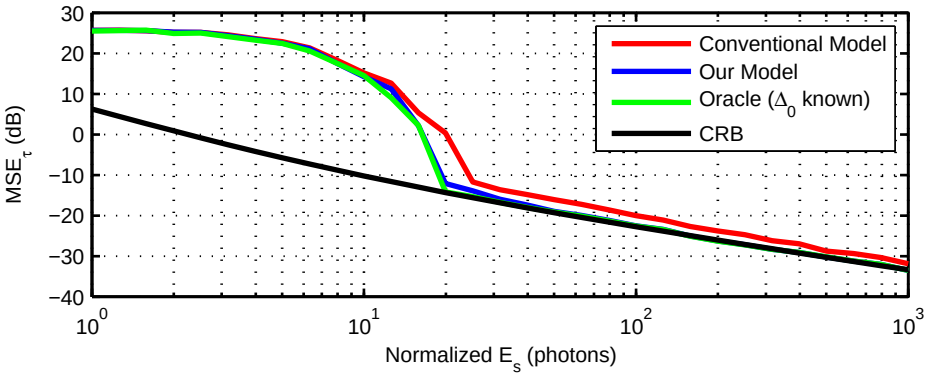


(c) Source signal is a Gaussian pulse with $\sigma = 1.6$.

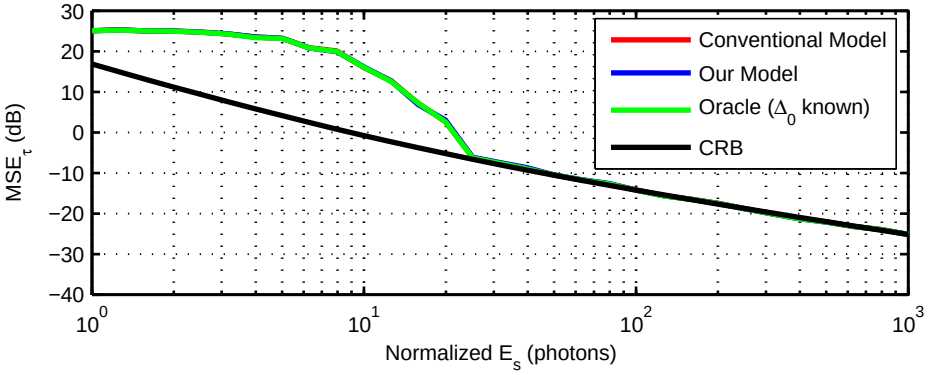
Figure 4.6. Empirical mean square error in dB for the ML estimate of the time shift for three different models evaluated using Monte-Carlo simulations with 1000 independent trials. The conventional model assumes $\Delta = 0$; oracle (Δ_0 known) is the ML estimate where the depth spread of the scene is exactly known. Here, we have $\Delta_0 = 2$, $\tau_0 = 40$, $T = 60$, $\gamma = 0.001$.



(a) Source signal is a Gaussian pulse with $\sigma = 0.1$.

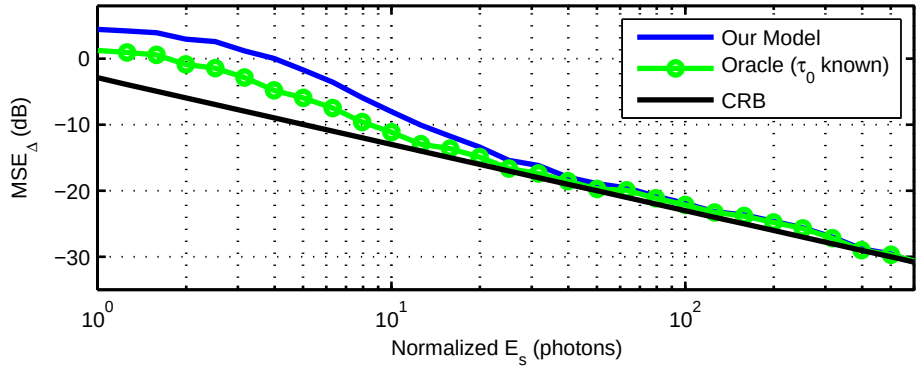


(b) Source signal is a Gaussian pulse with $\sigma = 0.4$.

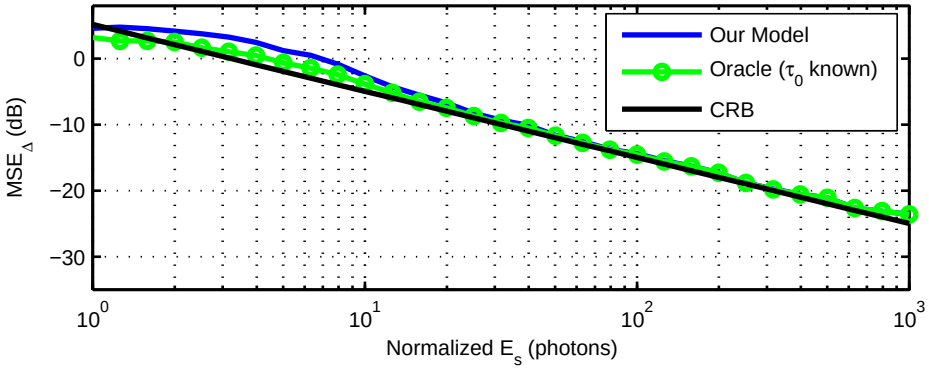


(c) Source signal is a Gaussian pulse with $\sigma = 1.6$.

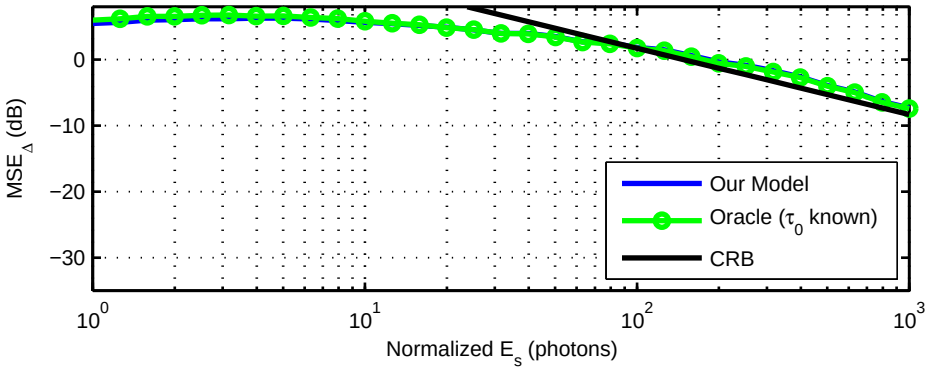
Figure 4.7. Empirical mean square error in dB for the ML estimate of the time shift for three different models evaluated using Monte-Carlo simulations with 1000 independent trials. The conventional model assumes $\Delta = 0$; oracle (Δ_0 known) is the ML estimate where the depth spread of the scene is exactly known. Here, we have $\Delta_0 = 2$, $\tau_0 = 40$, $T = 60$, $\gamma = 1.25$.



(a) Source signal is a Gaussian pulse with $\sigma = 0.1$.

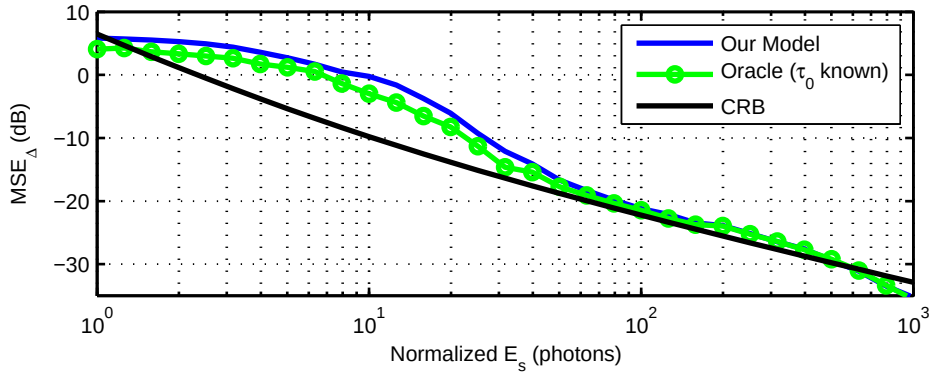


(b) Source signal is a Gaussian pulse with $\sigma = 0.4$.

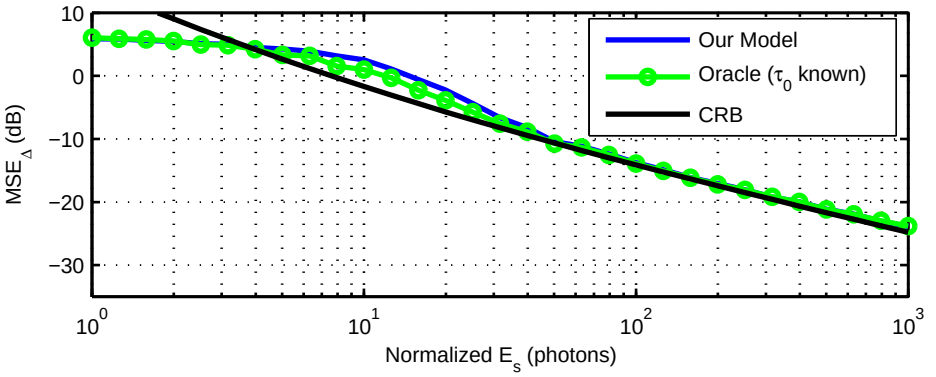


(c) Source signal is a Gaussian pulse with $\sigma = 1.6$.

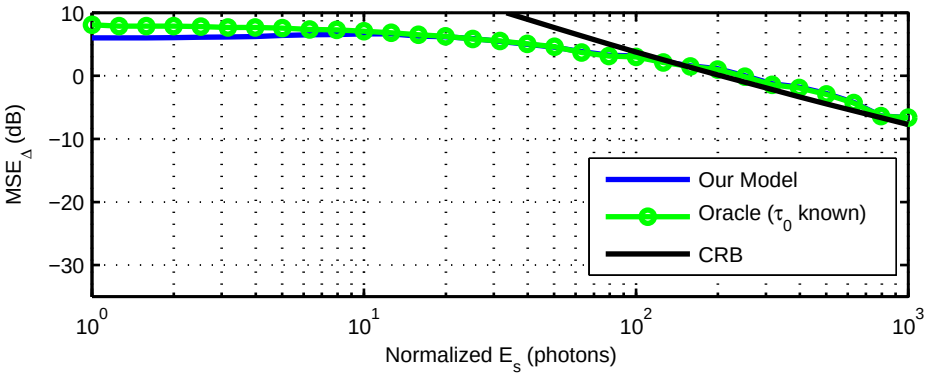
Figure 4.8. Empirical mean square error in dB for the ML estimate of Δ for two different models evaluated using Monte-Carlo simulations with 1000 independent trials. Oracle (τ_0 known) is the ML estimate where the time shift of the scene center is exactly known in advance. Here, we have $\Delta_0 = 2$, $\tau_0 = 40$, $T = 60$, $\gamma = 0.001$.



(a) Source signal is a Gaussian pulse with $\sigma = 0.1$.



(b) Source signal is a Gaussian pulse with $\sigma = 0.4$.



(c) Source signal is a Gaussian pulse with $\sigma = 1.6$.

Figure 4.9. Empirical mean square error in dB for the ML estimate of Δ for two different models evaluated using Monte-Carlo simulations with 1000 independent trials. Oracle (τ_0 known) is the ML estimate where the time shift of the scene center is exactly known in advance. Here, we have $\Delta_0 = 2$, $\tau_0 = 40$, $T = 60$, $\gamma = 1.25$.

Chapter 5

Discussion and Conclusion

In this thesis, we considered a generic time-of-flight depth acquisition problem with low photon count, where there could be ambient light present in the experiment and the photodetector could have dark count and timing jitter. However, we assumed the photon arrival times at the output of detector to be precisely known. We then decomposed the received light intensity at the detector as the convolution of the source intensity and a signal which we called the impulse response of the scene. After that, we used tools from calculus to formulate a general framework where given the scene, one can calculate the scene impulse response. Therefore, if the scene belongs to a parametric class of scenes, one gets a parametric impulse response. We specifically considered the two most basic parametric classes for the scene impulse response where one of them is the model being used in the existing ranging technologies and we refer to it as the conventional model and the other model is new and a key part of this work. Our proposed model is accurate for planar scenes and can be thought of as a better approximation for non-planar scenes than the conventional model. We then in Chapter 4, analyzed the performance of the conventional and the proposed models and compared the scene parameter estimates obtained by using those two models. It was observed that the proposed model of this work performs much better than the conventional model if the true underlying scene is a tilted plane with large tilt (relative to the width of the source signal used in the experiment). The reason behind this performance gap is the fact that the conventional model in that case is a totally mismatched model to the true underlying scene and therefore, the estimation of parameters is poor if one wants to use the conventional model in the estimation problem. Also, it should be noted that our proposed model is basically the same as the conventional model if the source signal with which we illuminate the scene is very wide in time domain and hence, both models have the same performance. Therefore, the proposed model in this work always has a uniformly better or equal performance to what one can get by using the conventional

model which emphasizes the importance of this work. However, we also noted that the computational complexity of the maximum likelihood algorithm with our model is quite high and the performance improvement is at the expense of more computation.

As future work, one may want to develop a more computationally efficient estimation algorithm than the maximum likelihood we used in this thesis. Specifically, because the off-diagonal entries in the Fisher information matrix in our estimation problem are zero, in theory, we should be able to estimate the two parameters of the proposed model separately by solving two independent one-dimensional problems instead of a two-dimensional problem. For example, one may want to work with the empirical moments of the arrival times to form an overdetermined system of equations to find the closest match for the parameters of the model. This may lead to huge computational reduction while the solution may remain close to the optimal solution.

Also, one can use the framework we derived in Chapter 3 to devise more complicated models to get better approximation to the true scene impulse response if the scene is known to belong to a specific class of scenes. Furthermore, the methodology employed in this thesis with the right model for the scene impulse response (e.g., the convex combination of box functions which corresponds to piecewise linear scenes) can be easily applied to the experiments where the scene is flood-lighted. Moreover, the framework introduced in this work can be extended to more accurate models of the photodetector (e.g., the shot noise models we introduced in Chapter 2 to take into account the Gaussian noise and finite electrical bandwidth of the photodetector) and the resulting problems can be solved by the expectation-maximization (EM) algorithm with our formulation as its foundation.

One may also consider solving a bigger problem where using the spatial correlation of depths in adjacent pixels, it is possible to get *super resolution* in the reconstruction step. This can be achieved using some framework like Generalized Approximate Message Passing (GAMP) of [37] where the estimator investigated in this work will be the scalar estimator of the GAMP formulation.

Finally, we should note that, in this thesis, we treated the problem of photon-limited depth acquisition only from a parametric perspective. However, one may consider using non-parametric inference algorithms which can sometimes be implemented using fast iterative methods.

Appendix A

Cramér-Rao Lower Bounds

In this appendix, we derive the Cramér-Rao lower bounds for the mean square error of any estimate for the time shift parameter in the time shift estimation problem and the mean square error of the time shift and depth spread in the joint estimation problem. In the following section, we derive the CRB for the time shift estimation problem and in Section A.2, we calculate the CRB for the joint estimation of the time shift and depth spread. First, let us prove the following lemma which will come in handy when we derive the CRB's.

Lemma A.0.1. *Suppose the arrivals of a Poisson process with rate $\lambda(t)$ are being observed over the interval $(0, T]$ and the arrival times are $\{t_i\}_{i=1}^{M}$ where M, the number of arrivals, is a Poisson random variable with mean $E = \int_0^T \lambda(t)\mathrm{d}t$. The expected value of the function $\tilde{g}\left(\{t_i\}_i; \alpha\right) = \sum_i g(t_i; \alpha)$, where α is a vector of deterministic parameters, is as follows ($\tilde{g} = 0$ if there is no arrival):*

$$G(\alpha) \triangleq E\left[\tilde{g}(\{t_i\}_i; \alpha)\right] = \int_0^T \lambda(t) g(t; \alpha)\, \mathrm{d}t. \tag{A.1}$$

Proof. From probability theory, it is known that conditioned on the number of arrivals, the Poisson arrival times can be considered as independent and identically distributed (iid) random variables with their marginal pdf equal to $\frac{\lambda(t)}{E}$. Therefore, to calculate the expectation of (A.1), we can first condition on the number of arrivals and then take the expectation with respect to the Poisson distribution for the number of arrivals as the following (basically, iterated expectations):

$$
\begin{aligned}
G(\alpha) &\triangleq E\left[\tilde{g}(\{t_i\}_i;\alpha)\right] \\
&\overset{(a)}{=} \sum_{m=1}^{\infty} \frac{e^{-m}E^m}{m!} E\left[\sum_{i=1}^{m} g(t_i;\alpha)\big|M=m\right] \\
&\overset{(b)}{=} \sum_{m=1}^{\infty} \frac{e^{-m}E^m m}{m!} E\left[g(t_1;\alpha)\big|M=m\right] \\
&\overset{(c)}{=} \sum_{m=1}^{\infty} \frac{e^{-m}E^m m}{m!} \int_0^T g(t_1;\alpha)\frac{\lambda(t_1)}{E}\,\mathrm{d}t_1 \\
&\overset{(d)}{=} \left(\sum_{m=1}^{\infty} \frac{e^{-E}E^{m-1}}{(m-1)!}\right) \int_0^T g(t_1;\alpha)\lambda(t_1)\,\mathrm{d}t_1 \\
&\overset{(e)}{=} \int_0^T \lambda(t)g(t;\alpha)\,\mathrm{d}t, \qquad \text{(A.2)}
\end{aligned}
$$

where (a) is the law of iterated expectations ($\tilde{g}$ is zero if there is no arrival and therefore, the summation starts from $m=1$) and (b) is due to the fact that the expectation is linear and given the number of arrivals, the arrival times are iid. (c) follows from (b) by replacing the expectation by its definition. Getting (d) from (c) is straightforward and (e) results directly from (d) because the summation is one. ■

■ A.1 CRB for Estimation of Time Shift

In this section, we consider the scenario where the scene is assumed to be only one fronto-parallel plane with depth τ_0. Therefore, assuming the background noise is γ and the source signal is $s(t)$, the rate of the Poisson arrival times is $\lambda(t) = s(t-\tau_0)+\gamma$ and the log likelihood function is as follows:

$$
L(\tau) = \sum_i \log\left(1+\frac{1}{\gamma}s(t_i-\tau)\right). \qquad \text{(A.3)}
$$

Now, from the estimation theory [43], it is known that the Fisher information can be derived as the following:

$$
\begin{aligned}
\mathcal{I}(\tau_0) &\triangleq -E\left[\left.\frac{\partial^2 L(\tau)}{\partial \tau^2}\right|_{\tau=\tau_0}\right] \\
&\overset{(a)}{=} -E\left[\sum_i \frac{\frac{1}{\gamma}\ddot{s}(t_i-\tau_0)}{1+\frac{1}{\gamma}s(t_i-\tau_0)} - \frac{\left(\frac{1}{\gamma}\dot{s}(t_i-\tau_0)\right)^2}{\left(1+\frac{1}{\gamma}s(t_i-\tau_0)\right)^2}\right] \\
&\overset{(b)}{=} -E\left[\sum_i \frac{\frac{1}{\gamma}\ddot{s}(t_i-\tau_0)}{1+\frac{1}{\gamma}s(t_i-\tau_0)}\right] + E\left[\sum_i \frac{\left(\frac{1}{\gamma}\dot{s}(t_i-\tau_0)\right)^2}{\left(1+\frac{1}{\gamma}s(t_i-\tau_0)\right)^2}\right] \\
&\overset{(c)}{=} -\int_0^T \ddot{s}(t-\tau_0)\,\mathrm{d}t + \int_0^T \frac{\left(\dot{s}(t-\tau_0)\right)^2}{\gamma+s(t-\tau_0)}\,\mathrm{d}t \\
&\overset{(d)}{=} \int_0^T \frac{\left(\dot{s}(t-\tau_0)\right)^2}{\gamma+s(t-\tau_0)}\,\mathrm{d}t \\
&\overset{(e)}{=} \int_0^T \frac{\left(\dot{s}(t)\right)^2}{\gamma+s(t)}\,\mathrm{d}t,
\end{aligned}
\tag{A.4}
$$

where $\dot{s}(t)$ and $\ddot{s}(t)$ are the derivative and second derivative of $s(t)$ with respect to t. In (a), we have used (A.3) to calculate $\left.\frac{\partial^2 L(\tau)}{\partial \tau^2}\right|_{\tau=\tau_0}$ and (b) directly follows (a) because the expectation is linear. (c) is the result of applying Lemma A.0.1 to the expectations in (b) and finally, we get (d) because we assume the derivative of $s(t-\tau_0)$ at the interval boundaries is zero. Because we always assume the feasible set for τ_0 is much smaller than the observation interval $(0,T)$, we can shift the integrand in (d) by τ_0 to get (e). Now, we can use the Fisher information in (A.4) to derive the Cramér-Rao lower bound as the following:

$$
\mathrm{MSE}_\tau \geq \frac{1}{\mathcal{I}(\tau_0)} = \left(\int_0^T \frac{\left(\dot{s}(t)\right)^2}{\gamma+s(t)}\,\mathrm{d}t\right)^{-1}. \tag{A.5}
$$

■ A.2 CRB for Joint Estimation of Time Shift and Depth Spread

In this section, we consider the case where the scene is a tilted plane and therefore, there are two parameters to be estimated (time shift, τ_0, and depth spread, Δ_0). Assuming background noise level is γ and the source signal is $s(t)$, we can write the log likelihood

function as the following:

$$L(\tau,\Delta)=\sum_i \log\left(1+\frac{1}{\gamma}r_\Delta(t_i-\tau)\right), \tag{A.6}$$

where

$$r_\Delta(t)=s(t)*\frac{1}{\Delta}\,\mathrm{rect}\left(\frac{t}{\Delta}\right). \tag{A.7}$$

Also, in this case, the rate of the Poisson arrival times is $\lambda(t)=\gamma+r_{\Delta_0}(t-\tau_0)$. Here, because we have two parameters which should be estimated, the Fisher information takes a matrix form as the following:

$$\mathcal{I}(\tau_0,\Delta_0)=\begin{bmatrix}\mathcal{I}_{11} & \mathcal{I}_{12}\\ \mathcal{I}_{21} & \mathcal{I}_{22}\end{bmatrix}, \tag{A.8}$$

where we have

$$\begin{aligned}
\mathcal{I}_{11} &\triangleq -E\left[\frac{\partial^2 L}{\partial\tau^2}(\tau,\Delta)\Big|_{(\tau_0,\Delta_0)}\right]\\
&\overset{(a)}{=} -E\left[\frac{\frac{\partial^2 r_{\Delta_0}(\tau)}{\partial\tau^2}\big|_{\tau=t_i-\tau_0}}{\gamma+r_{\Delta_0}(t_i-\tau_0)}\right]+E\left[\frac{\left(\frac{\partial r_{\Delta_0}(\tau)}{\partial\tau}\big|_{\tau=t_i-\tau_0}\right)^2}{\left(\gamma+r_{\Delta_0}(t_i-\tau_0)\right)^2}\right]\\
&\overset{(b)}{=} E\left[\frac{\left[\frac{1}{\Delta_0}\left(s(t_i-\tau_0+\frac{\Delta_0}{2})-s(t_i-\tau_0-\frac{\Delta_0}{2})\right)\right]^2}{\left(\gamma+r_{\Delta_0}(t_i-\tau_0)\right)^2}\right]\\
&\overset{(c)}{=} \int_0^T \frac{\left[\frac{1}{\Delta_0}\left(s(t-\tau_0+\frac{\Delta_0}{2})-s(t-\tau_0-\frac{\Delta_0}{2})\right)\right]^2}{\gamma+r_{\Delta_0}(t-\tau_0)}\,\mathrm{d}t\\
&\overset{(d)}{=} \int_0^T \frac{\left[\frac{1}{\Delta_0}\left(s(t+\frac{\Delta_0}{2})-s(t-\frac{\Delta_0}{2})\right)\right]^2}{\gamma+r_{\Delta_0}(t)}\,\mathrm{d}t,
\end{aligned} \tag{A.9}$$

where in (a), we have expressed the second derivative of L in terms of the derivatives of r_{Δ_0}. It should also be noted that the first term in (a) is zero because it is equivalent to the integral of the second derivative of r_{Δ_0} which is zero as we saw in the one-dimensional case. Also, in (b), we have replaced the derivative of r_{Δ_0} by the numerator there. By using Lemma A.0.1, (c) directly follows from (b) and (d) is just a time-shifted copy of (c).

$$
\begin{aligned}
\mathcal{I}_{22} &\triangleq -E\left[\frac{\partial^2 L}{\partial \Delta^2}(\tau,\Delta)\Big|_{(\tau_0,\Delta_0)}\right] \\
&\overset{(a)}{=} -E\left[\frac{\frac{\partial^2 r_{\Delta_0}(\tau)}{\partial \Delta^2}|_{\tau=t_i-\tau_0}}{\gamma + r_{\Delta_0}(t_i-\tau_0)}\right] + E\left[\frac{\left(\frac{\partial r_{\Delta_0}(\tau)}{\partial \Delta}|_{\tau=t_i-\tau_0}\right)^2}{\left(\gamma + r_{\Delta_0}(t_i-\tau_0)\right)^2}\right] \\
&\overset{(b)}{=} E\left[\frac{\left[\frac{1}{\Delta_0}\left(r_{\Delta_0}(t_i-\tau_0) - \frac{1}{2}s(t_i-\tau_0+\frac{\Delta_0}{2}) - \frac{1}{2}s(t_i-\tau_0-\frac{\Delta_0}{2})\right)\right]^2}{\left(\gamma + r_{\Delta_0}(t_i-\tau_0)\right)^2}\right] \\
&\overset{(c)}{=} \int_0^T \frac{\left[\frac{1}{\Delta_0}\left(r_{\Delta_0}(t-\tau_0) - \frac{1}{2}s(t-\tau_0+\frac{\Delta_0}{2}) - \frac{1}{2}s(t-\tau_0-\frac{\Delta_0}{2})\right)\right]^2}{\gamma + r_{\Delta_0}(t-\tau_0)}\,\mathrm{d}t \\
&\overset{(d)}{=} \int_0^T \frac{\left[\frac{1}{\Delta_0}\left(r_{\Delta_0}(t) - \frac{1}{2}s(t+\frac{\Delta_0}{2}) - \frac{1}{2}s(t-\frac{\Delta_0}{2})\right)\right]^2}{\gamma + r_{\Delta_0}(t)}\,\mathrm{d}t,
\end{aligned}
\tag{A.10}
$$

where as before in (a), the first expectation is zero because $\frac{\partial^2 r_{\Delta_0}(\tau)}{\partial \Delta^2}$ is antisymmetric around $\tau=\tau_0$ and the second expectation is expanded in (b). Also, using Lemma A.0.1, (b) can be rewritten as (c) and as before, (d) is the result of shifting the integrand in (c) by τ_0.

$$
\begin{aligned}
\mathcal{I}_{21} = \mathcal{I}_{12} &\triangleq -E\left[\frac{\partial^2 L}{\partial \tau \partial \Delta}(\tau,\Delta)\Big|_{(\tau_0,\Delta_0)}\right] \\
&\overset{(a)}{=} -E\left[\frac{\frac{\partial^2 r_{\Delta_0}(\tau)}{\partial \tau \partial \Delta}|_{\tau=t_i-\tau_0}}{\gamma + r_{\Delta_0}(t_i-\tau_0)}\right] + E\left[\frac{\left(\frac{\partial r_{\Delta_0}(\tau)}{\partial \tau}\frac{\partial r_{\Delta_0}(\tau)}{\partial \Delta}\right)\Big|_{\tau=t_i-\tau_0}}{\left(\gamma + r_{\Delta_0}(t_i-\tau_0)\right)^2}\right] \\
&\overset{(b)}{=} -\int_0^T \frac{\partial^2 r_{\Delta_0}(\tau)}{\partial \tau \partial \Delta}|_{\tau=t-\tau_0}\,\mathrm{d}t + \int_0^T \frac{\left(\frac{\partial r_{\Delta_0}(\tau)}{\partial \tau}\frac{\partial r_{\Delta_0}(\tau)}{\partial \Delta}\right)\Big|_{\tau=t-\tau_0}}{\gamma + r_{\Delta_0}(t-\tau_0)}\,\mathrm{d}t \\
&\overset{(c)}{=} -\frac{\partial^2}{\partial \tau \partial \Delta}\int_0^T r_{\Delta_0}(\tau)\,\mathrm{d}\tau + 0 \\
&\overset{(d)}{=} 0,
\end{aligned}
\tag{A.11}
$$

where (a) is just the expansion of the second derivative and using Lemma A.0.1, we get (b) from (a) and the integrand in the second integral in (b) is antisymmetric around τ_0. Also, the derivative in the first integral of (b) can be pulled out of the integral and because r_{Δ_0} always integrates to E_s regardless of the values of τ_0 and Δ_0, we get zero. Therefore, the Fisher information matrix is $\mathcal{I}(\tau_0, \Delta_0) = \begin{pmatrix} \mathcal{I}_{11} & 0 \\ 0 & \mathcal{I}_{22} \end{pmatrix}$ where $\mathcal{I}_{11}$ and $\mathcal{I}_{22}$ are given in (A.9) and (A.10). Now, because the Fisher information matrix is a diagonal matrix, using (A.9) and (A.10), we have the following Cramér-Rao lower bounds for the mean square error of time shift and depth spread:

$$\mathrm{MSE}_\tau \geq \frac{1}{\mathcal{I}_{11}} = \left(\int_0^T \frac{\left[\frac{1}{\Delta_0} \left(s(t + \frac{\Delta_0}{2}) - s(t - \frac{\Delta_0}{2}) \right) \right]^2}{\gamma + r_{\Delta_0}(t)} \, \mathrm{d}t \right)^{-1}, \tag{A.12}$$

$$\mathrm{MSE}_\Delta \geq \frac{1}{\mathcal{I}_{22}} = \left(\int_0^T \frac{\left[\frac{1}{\Delta_0} \left(r_{\Delta_0}(t) - \frac{1}{2} s(t + \frac{\Delta_0}{2}) - \frac{1}{2} s(t - \frac{\Delta_0}{2}) \right) \right]^2}{\gamma + r_{\Delta_0}(t)} \, \mathrm{d}t \right)^{-1}. \tag{A.13}$$

Bibliography

[1] K. Abend. Optimum photon detection. *IEEE Trans. Inform. Theory*, 12(1):64–65, Jan. 1966.

[2] A. Antoniadis and J. Bigot. Poisson inverse problems. *Ann. Stat.*, 34(5):2132–2158, 2006.

[3] N. Antoniadis and A. O. Hero. Time-delay estimation for filtered Poisson processes using an EM-type algorithm. *IEEE Trans. Signal Process.*, 42(8):2112–2123, Aug. 1994.

[4] I. Bar-David. Statistics of nonstationary shot processes and the fluctuations of detected signals. *Proc. IEEE*, 56(12):2167–2168, Dec. 1968.

[5] I. Bar-David. Communication under the Poisson regime. *IEEE Trans. Inform. Theory*, 15(1):31–37, Jan. 1969.

[6] I. Bar-David. On the extension of the discrete model for optical detection. *Proc. IEEE*, 59(11):1612, Nov. 1971.

[7] I. Bar-David. Information in the time of arrival of a photon packet: Capacity of PPM channels. *J. Opt. Soc. Amer.*, 63(2):166–170, Feb. 1973.

[8] I. Bar-David. Minimum-mean-square-error estimation of photon pulse delay. *IEEE Trans. Inform. Theory*, 21(3):326–330, May 1975.

[9] T. Berger. *Rate Distortion Theory*. Prentice-Hall, New York, 1971.

[10] D. P. Bertsekas and J. N. Tsitsiklis. *Introduction to Probability*. Athena Scientific, Nashua, NH, 2008.

[11] F. Bretar, A. Chauve, C. Mallet, and B. Jutzi. Managing full waveform LiDAR data: A challenging task for the forthcoming years. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(Part B1): 415–420, 2008.

[12] J. Castorena and C. D. Creusere. Compressive sampling of LiDAR: Full-waveforms as signals of finite rate of innovation. In *Proc. 20th European Signal Process. Conf. (EUSIPCO 2012)*, pages 984–988, Bucharest, Romania, Aug. 2012.

[13] J. Castorena, C. D. Creusere, and D. Voelz. Using finite moment rate of innovation for LiDAR waveform complexity estimation. In *Conf. Rec. Asilomar Conf. on Signals, Syst. & Computers*, pages 608–612, Pacific Grove, CA, 2010.

[14] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. Plerrot-Deseilligny, and W. Puech. Processing full-waveform LiDAR data: Modeling raw signals. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 3/W52):102–107, Sep. 2007.

[15] N. H. Clinthorne, N. A. Petrick, W. L. Rogers, and A. O. Hero. A fundamental limit on timing performance with scintillation detectors. *IEEE Trans. Nucl. Sci.*, 37(2):658–663, Apr. 1990.

[16] T. F. Curran and M. Ross. Optimum detection thresholds in optical communications. *Proc. IEEE*, 53(11):1770–1771, Nov. 1965.

[17] B. I. Erkmen and B. Moision. Maximum likelihood time-of-arrival estimation of optical pulses via photon-counting photodetectors. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 1909–1913, Seoul, Korea, Jul. 2009.

[18] R. M. Gagliardi and S. Karp. M-ary Poisson detection and optical communications. *IEEE Trans. Commun. Tech.*, 17(2):208–216, Apr. 1969.

[19] J. W. Goodman. *Statistical Optics*. Wiley, 1985.

[20] A. O. Hero. Lower bounds on estimator performance for energy-invariant parameters of multidimensional Poisson processes. *IEEE Trans. Inform. Theory*, 35(4): 843–858, Jul. 1989.

[21] A. O. Hero. Timing estimation for a filtered Poisson process in Gaussian noise. *IEEE Trans. Inform. Theory*, 37(1):92–106, Jan. 1991.

[22] A. O. Hero, N. Antoniadis, N. H. Clinthorne, W. L. Rogers, and G. D. Hutchins. Optimal and sub-optimal post-detection timing estimators for PET. *IEEE Trans. Nucl. Sci.*, 37(2):725–729, Apr. 1990.

[23] M. A. Hofton, J. B. Minster, and J. B. Blair. Decomposition of laser altimeter waveforms. *IEEE Trans. Geoscience and Remote Sens.*, 38(4):1989–1996, 2000.

[24] E. V. Hoversten, D. L. Snyder, R. O. Harger, and K. Kurimoto. Direct-detection optical communication receivers. *IEEE Trans. Commun.*, 22(1):17–27, Jan. 1974.

[25] A. V. Jelalian. *Laser Radar Systems*. Artech House, Boston-London, 1992.

[26] S. Karp, E. L. O'Neill, and R. M. Gagliardi. Communication theory for the free-space optical channel. *Proc. IEEE*, 58(10):1611–1626, Oct. 1970.

[27] P. L. Kelley and W. H. Kleiner. Theory of electromagnetic field measurement and photoelectron counting. *Phys. Rev. A*, 136(2):316–334, Oct. 1964.

[28] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal. Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Opt. Expr.*, 19 (22):21485–21507, 2011.

[29] A. Kirmani, H. Jeelani, V. Montazerhodjat, and V. K. Goyal. Diffuse imaging: Creating optical images with unfocused time-resolved illumination and sensing. *IEEE Signal Process. Lett.*, 19(1):31–34, Jan. 2012.

[30] C. Mallet and F. Bretar. Full-waveform topographic LiDAR: State-of-the-art. *ISPRS J. Photogrammetry and Remote Sensing*, 64(1):1–16, 2009.

[31] L. Mandel. Phenomenological theory of laser beam fluctuations and beam mixing. *Phys. Rev.*, 138(3B):B753–B762, May 1965.

[32] L. Mandel and E. Wolf. Coherence properties of optical fields. *Rev. Mod. Phys.*, 37(2):231–287, Apr. 1965.

[33] P. N. Misra and H. W. Sorenson. Statistical theory of estimation in doubly-stochastic Poisson processes. *Information Sciences*, 6:343–358, 1973.

[34] P. N. Misra and H. W. Sorenson. Parameter estimation in Poisson processes. *IEEE Trans. Inform. Theory*, 21(1):87–90, Jan. 1975.

[35] M. Oren and S. K. Nayar. Generalization of the Lambertian model and implications for machine vision. *Int. J. Comput. Vis.*, 14(3):227–251, Apr. 1995.

[36] Å. Persson, U. Söderman, J. Töpel, and S. Ahlberg. Visualization and analysis of full-waveform airborne laser scanner data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 3/W19):103–108, Sep. 2005.

[37] S. Rangan. Generalized approximate message passing for estimation with random linear mixing. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 2168–2172, Saint-Petersburg, Russia, Aug. 2011.

[38] B. Reiffen and H. Sherman. An optimum demodulator for Poisson processes: Photon source detectors. *Proc. IEEE*, 51(10):1316–1320, Oct. 1963.

[39] B. Schwarz. LiDAR: Mapping the world in 3D. *Nature Photonics*, 4(7):429–430, Jul. 2010.

[40] D. L. Snyder. Filtering and detection for doubly stochastic Poisson processes. *IEEE Trans. Inform. Theory*, 18(1):91–102, Jan. 1972.

[41] D. L. Snyder and M. I. Miller. *Random Point Processes in Time and Space*. Springer Texts in Electrical Engineering. Springer-Verlag, New York, 2nd edition, 1991.

[42] H. L. Van Trees. *Detection, Estimation, and Modulation: Part I*. Wiley, New York, 1968.

[43] H. L. Van Trees and K. L. Bell. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. IEEE press, 2007.

[44] S. Verdú. Poisson communication theory. Invited talk in the International Technion Communication Day in honor of Israel Bar-David, 1999. Available on-line at http://www.princeton.edu/~verdu.

[45] W. Wagner, A. Ullrich, T. Melzer, C. Briese, and K. Kraus. From single-pulse to full-waveform airborne laser scanners: Potential and practical challenges. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(B3):201–206, 2004.

[46] J. Ziv and M. Zakai. Some lower bounds on signal parameter estimation. *IEEE Trans. Inform. Theory*, IT-15(3):386–391, May 1969.