

# **Low-Light Scene Reconstruction Using Hyper-Spectral Single-Photon LiDAR**

V Anirudh Puligandla

Université de Bourgogne · Universitat de Girona · Heriot Watt University

A Thesis Submitted for the Degree of  
MSc Erasmus Mundus in Vision and Robotics (VIBOT)

· 2018 ·

## Abstract

Recent introduction of the Single-Photon LiDAR technology has garnered wide interest to produce accurate 3D reconstructions from limited amounts of photons. Such devices have higher sensitivity over traditional laser scanning systems and are capable of imaging over multiple wavelengths from visible to near infra red simultaneously. With such advantages this technology has found many applications under low-light scenarios. Low-Light constraints may arise due to various factors such as short acquisition times (e.g, long range applications) or low light-flux of the emitted laser (e.g, bio-medical applications, where even low-powered lasers might damage the tissues). This work focuses on the reconstruction of Intensity images simultaneously, from measurements taken over multiple peak wavelengths, using a Poisson measurement model and Alternating Direction Method of Multipliers (ADMM) algorithm. Different sparsity promoting regularizers have been proposed under a compressed sensing framework. This work will also show that the proposed methods are capable of reconstruction with light flux as low as 1 or even 0.5 received photons per pixel over an image.

*Research is what I'm doing when I don't know what I'm doing. . . .*

Werner von Braun

# Contents

<b>Acknowledgments</b>	v
<b>1 Introduction</b>	1
1.1 Understanding Single Photon Lidar . . . . .	1
1.2 Problem Definition . . . . .	4
1.3 Outline . . . . .	5
<b>2 Background</b>	6
2.1 Statistical Model . . . . .	6
2.2 Related Work . . . . .	9
2.3 Observation Model . . . . .	10
2.3.1 Baseline Intensity . . . . .	13
2.3.2 Response Intensity . . . . .	13
2.4 Compressed Sensing . . . . .	14
2.5 ADMM Algorithm . . . . .	15
2.5.1 Varying Step Size . . . . .	17
<b>3 Methodology</b>	19
3.1 TV and Nuclear Norm Regularization Model . . . . .	20
3.1.1 Minimization problem . . . . .	21
3.1.2 Algorithmic Details . . . . .	21

3.2	Joint Sparsity Model . . . . .	22
3.2.1	Minimization Problem . . . . .	24
3.2.2	Algorithmic Details . . . . .	24
<b>4</b>	<b>Results</b>	<b>27</b>
4.1	Discussion . . . . .	33
<b>5</b>	<b>Conclusion</b>	<b>35</b>
5.1	Conclusion . . . . .	35
5.2	Future Work . . . . .	36
<b>A</b>	<b>Convex Analysis</b>	<b>37</b>
<b>B</b>	<b>Moreau Proximity Operator</b>	<b>38</b>
	<b>Bibliography</b>	<b>44</b>

# List of Figures

1.1	An example of forest cover mapping from an aircraft . . . . .	2
1.2	An example of forest floor or sea floor mapping . . . . .	2
1.3	ToF Lidar system . . . . .	3
1.4	SPL sensor array . . . . .	4
1.5	Problem Definition . . . . .	5
2.1	Photon histogram at each pixel . . . . .	7
2.2	Response signal structure . . . . .	8
2.3	Response signal structure at low measurements . . . . .	9
2.4	An illustration of the measurement model . . . . .	12
2.5	Standard ADMM algorithm . . . . .	15
3.1	Variant of the PIDAL-TV algorithm, [1] . . . . .	23
3.2	Variant of the PIDAL-FA algorithm, [1] . . . . .	26
4.1	Acquisition Setup . . . . .	28
4.2	Ground truth images . . . . .	29
4.3	High photon count results . . . . .	31
4.4	Results at low measurements . . . . .	32
4.5	Results at $\sim 0.5$ ppp . . . . .	34

# List of Tables

4.1 Under-sampling map . . . . .	29
4.2 SNR values for all reconstructions . . . . .	30

# Acknowledgments

I would like to take this opportunity to, firstly, thank my supervisors, Prof. Yves Wiaux, for providing me with the necessary motivation to carry on the task until the very end of the internship, and also for his novel and practical ideas, and Dr. Yoann Altmann, for the long meetings he spent with me to discuss the problems encountered during the project, down to the minute detail. I would also like to thank the other PhD. students at the BASP group for sharing their priceless knowledge and experience with me. I would also like to thank my friends for their contributions to the beginning of this project. And last but not the least, I would like to thank my parents without whose support i would not be in the position where I am today.

# Chapter 1

## Introduction

### 1.1 Understanding Single Photon Lidar

Photon counting techniques have been used for decades mainly to monitor the atmosphere [2]. For example, precipitation can be estimated by checking the amount of water present in the air [3], or, pollution can be estimated by checking the amount of aerosols present in the air [4]. Single-Photon Lidar (SPL) is an example of advanced photon counting device. The high sensitivity of SPL photo-receivers enable them to map large regions of interest from a device mounted on airplanes, Figure 1.1, [5]. Unlike traditional LiDAR systems, SPL is also capable to work in the visible spectrum of light. SPL has seen wide applications in archaeological survey. For example, recently, a sprawling ancient Mayan city was mapped from under the thick jungles of Guatemala using a lidar device mounted on an aircraft [6].

In addition to the above mentioned applications, traditional lidar systems have been widely used in *geographical information systems* [7,8], *underwater imaging* (mainly bathymetry) [9,10], Figure 1.2, and also in *forest monitoring* [11–13]. Some modification can be done to the lidar design for space imaging [14,15]. Similarly, lidar systems have also been used in molecular biology applications mainly for time-resolved fluorescence spectroscopy [16–18]. It is important to note that although the structural design of any lidar system is similar, the manufacturing process of the solid state devices varies with the application as detailed in this book [19].

The interest of this work lies in the context of Time-of-Flight systems [20]. Figure 1.3 shows an example setup of a Time-of-Flight Lidar system. The system consists of a pulsed light source, an optical diffuser to increase field-of-view, a collection lens to focus the reflected light onto the sensor, a solid state range sensor (array of photo-detectors) and a sensor interface to manage the system and the output data stream. Depending on the design of the SPL system,

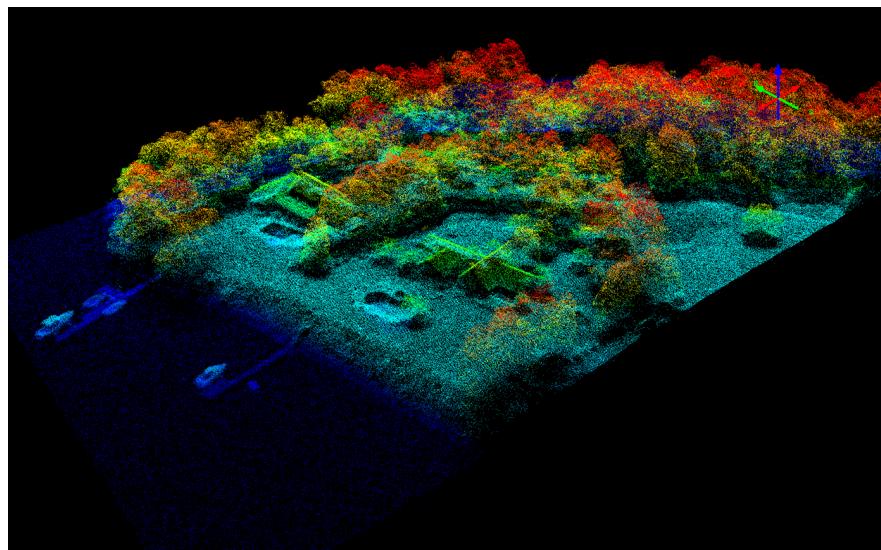


Figure 1.1: An example of forest cover mapping from an aircraft,  
*source: Sigma Space Corporation*

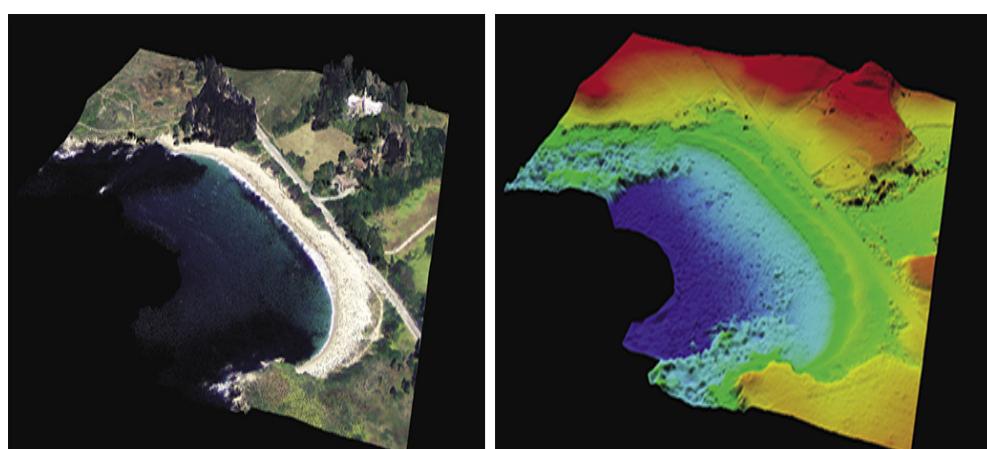


Figure 1.2: An example of forest floor or sea floor mapping, *source: NASA*

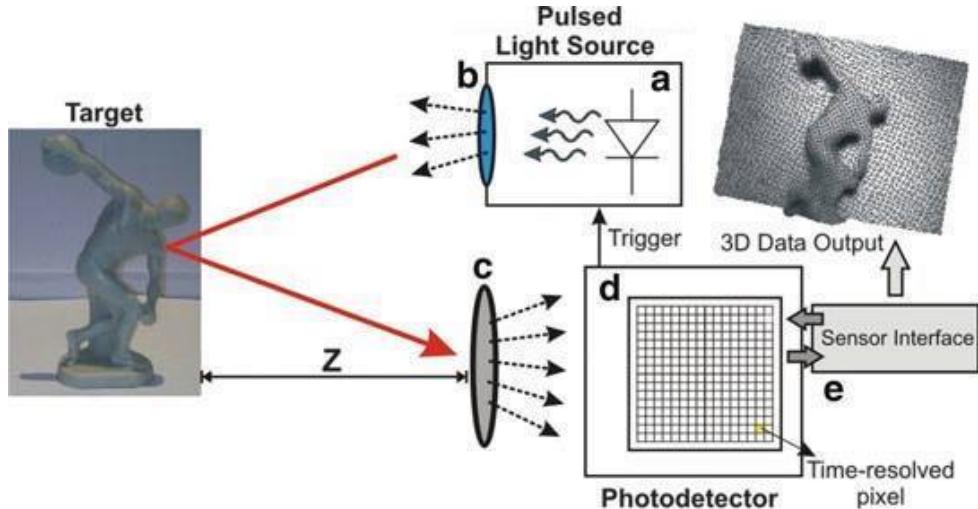


Figure 1.3: Schematic diagram of a typical Time-of-Flight SPL system, *source: [19]*

the scene can be imaged either simultaneously over all the wavelengths or sequentially over each wavelength.

An SPL system acquires the scene by recording all the pixels at the same time or by recording the scene in a raster scanning fashion. Devices with an architecture to scan the scene one pixel at a time are of primary interest in this context [20]. The reception of photons at each element (pixel) in the array can be seen to follow a statistical model known as *Poisson Process*. The next section describes the Poisson model, but for now, it is sufficient to know that a probability distribution can be associated to each pixel location and each pixel can be seen as an independent Poisson process. As mentioned in [19], *Single-photon imaging is the detection of two-dimensional patterns of low-intensity light (i.e., when the number of detected photons in each pixel is lower than 10)*.

Figure 1.4 depicts an illustration of the photon detection process. The round symbols in the images symbolize the positions where the reflected photons are imaged onto the image sensor. Full circles represent the positions where photons interact with the image sensor while open circles indicate no interaction of photons with the image sensor at those locations, either because there are no reflected photons or none were detected by the image sensor. The image on the left illustrates the distribution of photons interacting with the image sensor while the image on the right illustrates the subsequent electronic photo-charge detection and conversion process.

The noise in the detection process arises due to two primary factors. Firstly, the variance of photon detection is the sum of the variance of the number of photons interacting with the image

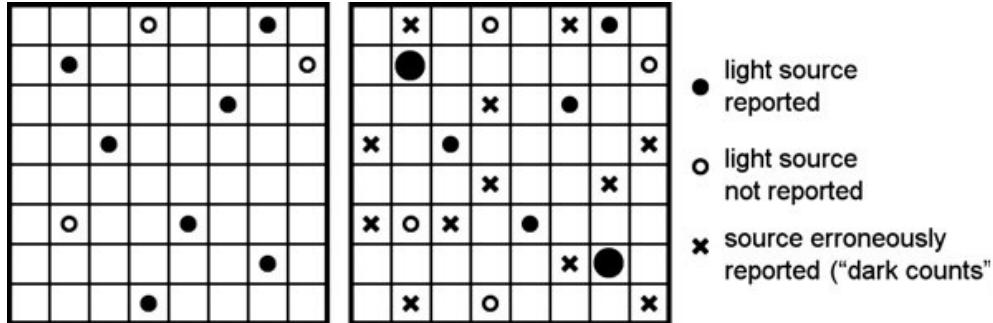


Figure 1.4: An illustration showing the noisy photon detection process, *source: [19]*

sensor, and the variance of the electronic photo-charge detection process. As a consequence, the reported number of photons becomes prone to errors, such as some interacting photons may not be detected anymore (full circles on the left become open circles on the right), or, the number of interacting photons may be exaggerated (larger full circles on the right).

Secondly, as the electronic noise is acting on all the pixels, photons may be reported even in locations where no photons are imaged onto the image sensor. These erroneously reported photons are indicated on the right-hand side image with a cross symbol, and are known as "dark counts". When these low-intensity reflected photons are sparsely distributed over the image sensor, it can become impossible to detect them when the electronic noise of the charge detection process is high.

## 1.2 Problem Definition

As seen in the previous section, the impact of noise is very high when the number of detected photons is low. Traditionally, the observation time is kept high (about a billion photons are collected at each pixel) to denoise using Gaussian noise assumptions. However, in the case of *single-photon imaging*, Poisson noise assumptions are more appropriate to minimize noise, as will be shown in the next chapter. In the rest of the document, we will see how an optimization framework can be constructed to maximize the likelihood of photon detection. The architecture of single-pixel ToF cameras has shown to reduce the required size, complexity and cost of the photon detector array down to a single unit [21]. Combining the single-pixel camera with random Compressed Sensing measurements, a trade-off between space and time can be achieved during image acquisition. Moreover, since the camera compresses as it images, it will be able to efficiently handle high-dimensional data sets from applications like video and hyper-spectral imaging [22].

However, combining single-photon imaging and compressed sensing results in the formulation of an inverse problem consisting of an ill-posed or ill-conditioned observation operator [1]. In the following chapters, we will also see how some regularizers (as known as priors from a Bayesian point of view) can help in hyper-spectral intensity image reconstruction when the observation operator is part of an under-determined system. Lastly, we will see how a minimization problem can be framed that follows the conditions mentioned above and that can be optimized using a convex optimization algorithm such as the Alternating Direction Method of Multipliers (ADMM) algorithm.

Figure 1.5 depicts an illustration of the problem stated above. It might be interesting to note that both, Figure 1.5a and Figure 1.5b, have the same average number of photons (observations) even though the latter consists of data only on  $1/16^{th}$  of the pixels. The sub-sampled image actually consists of an average 16 photons/pixel but when only  $1/16^{th}$  of the pixels are considered, the overall average will be 1 photon/pixel because  $15/16^{th}$  of the pixels will have 0 detected photons (no recorded data).

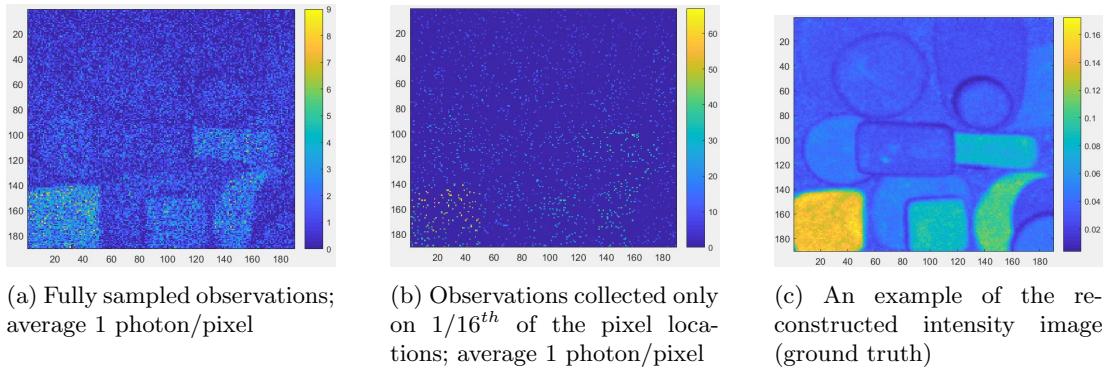


Figure 1.5: An illustration explaining the problem. This figure shows observations taken uniformly at all pixels (a), under-sampled observations on random pixels (b), expected result from the algorithm (c)

### 1.3 Outline

The rest of the document is organized as follows. Chapter 2 introduces related work and the background information that is required to understand the methods proposed in Chapter 3. Chapter 4 will describe the results obtained using the proposed methods while the conclusions and possibilities for future work will be presented in Chapter 5.

# Chapter 2

## Background

This chapter will introduce the concepts that are necessary to understand the methods. Precisely, we will first see the statistics of photon detection at each pixel, followed by the observation model, and finally, we will conclude this chapter after understanding the theory behind Compressed Sensing and convex optimization.

### 2.1 Statistical Model

SPL uses the Time Correlated Single Photon Counting technique (TCSPC) to count the number of photons imaged onto the sensor [23]. TCSPC is a standard statistical sampling technique used in ToF imaging, with single photon detection sensitivity, capable of pico-second timing resolution [24]. TCSPC exploits the fact that for signals with high repetition rate, the light intensity is so low that the probability of detecting one photon in one signal period is far less than one [25]. The output signal is constructed by recording the photons, measuring their time of arrival in the signal period and building a histogram of the photon times. Figure 2.1 shows an illustration of the histogram construction.

The output signal from the detector is a histogram of distributed pulses over predefined number of bins corresponding to the detection of individual photons. It can be noticed from Figure 2.1 that there exist many signal periods without any photons while the others have one photon per period. The probability of detecting more than one photon per period is close to zero [25]. Therefore, the probability of detection of photons for each signal period can be seen as a Bernoulli trial<sup>1</sup>, as there can be either 1 or 0 photons per signal period. The photon detection at each pixel location is a summation of multiple signal periods over the per pixel

---

<sup>1</sup>A Bernoulli trial is an experiment whose outcome is random and can only be of two possible outcomes, either success or failure.

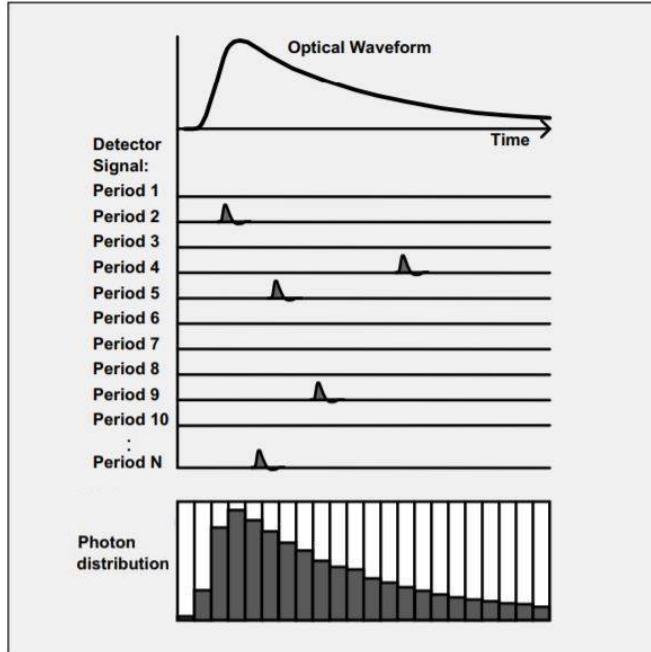


Figure 2.1: An illustration of the general principle of TCSPC. *source:* [25]

acquisition time, i.e, the detection of photons at each pixel is a sequence of consecutive Bernoulli trials. We know that the summation of consecutive Bernoulli trials (binary terms) leads to a binomial distribution [19]. The binomial distribution is similar to a Poisson distribution when the number of detected photons is very low, [26]. Precisely, the Poisson distribution is a special case of the binomial distribution when the number of signal periods are very high while the overall number of detected photons is very low. Moreover, the book [19] clearly explains why Poisson statistics are appropriate for SPL signals.

The probability  $p_N(k)$  that a number  $k \geq 0$  of events is observed during a fixed observation period  $T$  if these events occur with a known average rate  $N/T$ , and independent of time since the last event is given by,

$$p_N(k) = \frac{N^k e^{-N}}{k!} \quad (2.1)$$

where,  $k$  is also known as the mean intensity of the Poisson probability distribution. The following section will explain how Eq.(2.1) can be used to model the observations.

Now, if we assume that there exists an object within the field of view of the SPL system, the histogram at each pixel will contain a peak whose position depends on the depth/range of the target. The peak intensity value of the response histogram depends on the reflectivity

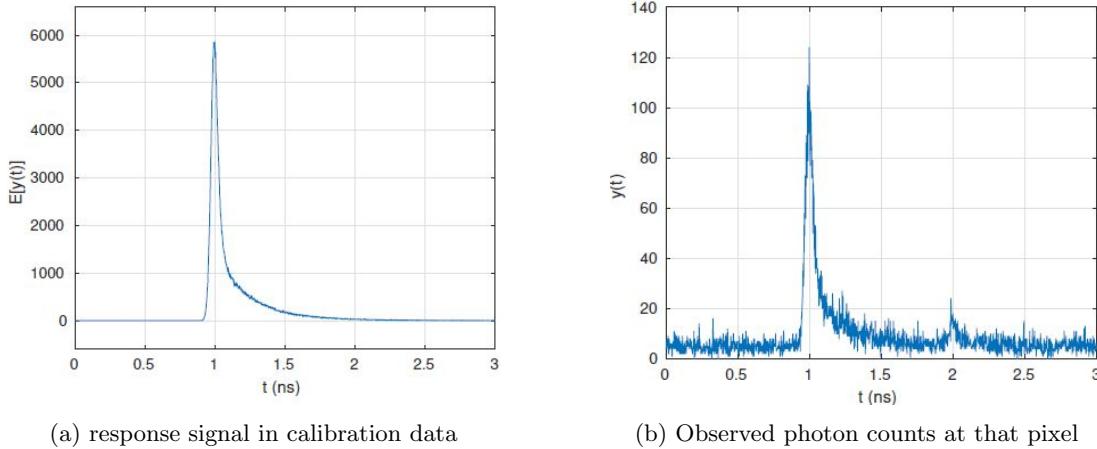


Figure 2.2: Illustration of the response signal structure for one pixel location when the number of detected photons is high.

of the object at the wavelength of the emitted pulse. We assume that some calibration data is available, that is nothing but the response of an object with known reflectivity placed at a known position with respect to the SPL device. These calibration data contain the shape and intensity of the response signals for that material, for a given acquisition time, repeated over all the considered wavelengths. The calibration data used here have been collected over an acquisition time of  $10ms$ . Figure 2.2 shows an illustration of the calibration data and the observations for one pixel location. The following two assumptions have been made in the context of low number of detected photons:

1. The shape of the response signal does not vary with the depth of the imaged surface
2. The angle between the light pulse and the surface of the target affects only the intensity of the response and not its shape

These assumptions are adequate in the case of narrow pulse width with very small beam divergence. The received response can be affected by other light sources in real environments. The passive background light, known as baseline response, can influence the response intensity reconstruction. This background light intensity usually appears in the response signal as a vertical shift (positive) of the intensities (see Figure 2.3a). However, the structure of the response signal becomes complicated when the number of detected photons is very low, i.e, when the light flux or the acquisition time are very low. One can observe from Figure 2.2 that the shape of the response signals from the calibration data and the real data with high observations, is similar. This similarity does not hold when the number of detected photons is

very low. Figure 2.3 shows an example of the response signal extracted from data containing very few observed photons (average number of observed photons <10 per pixel).

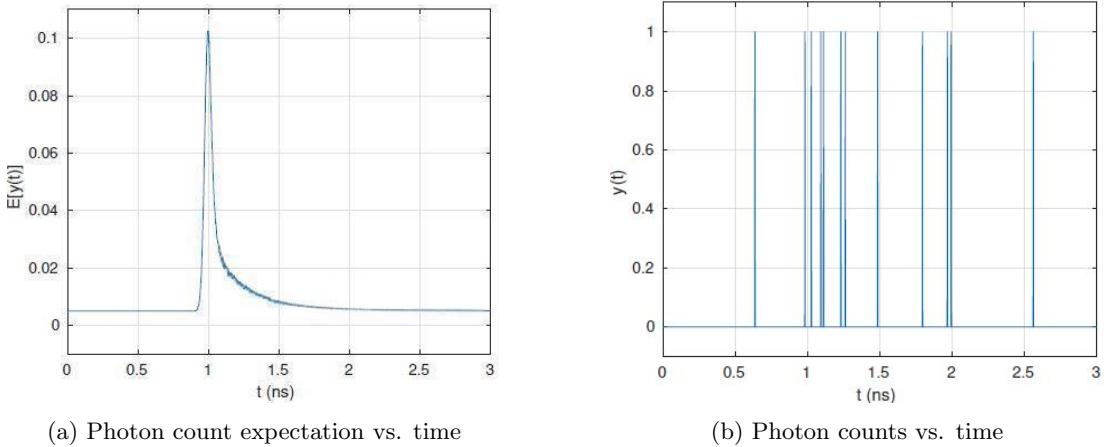


Figure 2.3: Illustration of the response signal at each pixel when very few photons are received.

## 2.2 Related Work

The challenges encountered during SPL signal reconstruction and the diverse applications are the reasons for ample amount of existing literature in this context. Despite the vast literature, there still exist possibilities of improvement, that will be evident in the following sections. A recent methodology to obtain 3D reconstructions from SPL data was proposed by Shin et al. in [27, 28]. This method assumes that at a given pixel, the response signal can be the sum of responses from limited number of objects with different reflectivities at different depths. With these assumptions the method is shown to be able to model translucent objects with a better resolution at the pixels containing edges. But the union-of-subspaces model limits it only to a few pre-defined responses. Moreover, this method does not exploit the spatial correlation among the pixels and thus, works only for the cases with sufficient amount of photons at each pixel location. Additionally, this method was proposed for only single band SPL and has no obvious extensions for multi-spectral SPL.

Recently, Altmann et al. had proposed an algorithm under the Bayesian context for simultaneous reconstruction and spectral unmixing of multi-spectral SPL data using the Markov Chain Monte Carlo (MCMC) method, [29]. This method works by estimating the abundances of each material in each pixel. Although the algorithm is efficient, it requires previous knowledge about the spectral response of all the materials present in the scene. In similar lines, they had also

proposed another method to use convex optimization with an aim to minimize the estimated number of materials in each pixel while maintaining the data fidelity or the regularizer terms low, [30]. However, this method also requires prior knowledge about the spectral responses of all the materials present in the scene.

Shin et al. proposed another method in [31, 32], where a known baseline noise is assumed over all the measured pixels. These assumptions are not adequate as the baseline noise is pixel dependent. These assumptions also do not hold for applications where the Lidar system is mobile. Despite these assumptions, the method achieves good reconstructions for data with low photon counts because of the exploitation of spatial correlation among the pixels through the use of Total Variation (TV) regularization. From prior work it is evident that convex optimization techniques provide reconstructions with high accuracy under low-light (low photon reception) settings. As mentioned in [33], photons reaching a detector can be seen as counts of discrete events, and thus, cannot be modelled effectively using additive bounded or Gaussian noise but instead require a Poisson noise model. This implies that the reconstruction of an intensity image from Poisson data cannot be achieved by minimizing a conventional penalized least-squares objective function.

In addition to the above mentioned methods, some comprehensive reviews of Poisson image restoration algorithms can be found in [34, 35]. Regularized variants of the classical Richardson-Lucy (RL) method have been proposed in [36, 37] that use Total Variation (TV) and Wavelet-based Regularization, respectively. Some multiscale approaches that handle image inverse problems can be found in [38–41].

Having described the prior work and the motivation for the chosen method, one can head to formulate the inverse problem of interest. The problem that will be formulated in this work considers the work done in [31, 32] as a base while proposing some improvements in extending it to multi-spectral SPL data. The following sections will introduce the necessary theoretical concepts that are required to understand the method used in this work. The inverse problem along with the objective function, appropriate penalty terms, and the algorithm used are described in detail in chapter 3, that is then followed by the results.

## 2.3 Observation Model

As we have seen why a Poisson model fits best with the data at hand, we will formulate a prior model that can be feasibly optimized using convex optimization. Inverse Problems have been widely used in all situations that require estimation of some physical quantities from a set of observations or measurements. Precisely, they have been widely used in various inpainting problems such as, denoising, deblurring, image reconstruction, etc.. The general theory says

that the unknown physical quantity can be linearly related to the observations by Eq.(2.2).

$$y = \Phi(x) \quad (2.2)$$

Where,  $\Phi$  is called the measurement operator, i.e, the parameters that linearly map the unknown quantity  $x$  to the measurements. In this situation, the unknown quantity can be directly estimated by simple matrix inversion of  $\Phi$ . However, in real-scenarios, this matrix inversion is not straight forward. Moreover, most of the practical linear systems are under-determined leading to an ill-posed inverse problem.

One of the most common approaches to model Poisson image restoration is to consider the likelihood of the observations. Using Poisson statistics, we can establish a linear relationship between the measurements and the intensity image that needs to be reconstructed as shown in Eq.(2.3).

$$y_p \sim Pr(Fx_p + b_p) \quad (2.3)$$

Where,  $Pr(\lambda)$  represents the Poisson Distribution with mean intensity  $\lambda$ ,  $y_p$  are the measurements at each pixel,  $y_p = (y_{p1}, \dots, y_{pm}) \in \mathbb{N}_0^m$  (where,  $\mathbb{N}_0$  represents the set of non-negative integers and  $m$  is the number of bins in the constructed histogram),  $x_p$  and  $b_p$  are the response and baseline intensities, respectively.  $F \in \mathbb{R}^{D \times M}$  (where,  $D$  is the range of the SPL system in discrete steps and  $M$  is the number of bins) is known as the calibration matrix. This calibration data is collected prior to the measurements with each column of  $F$  containing the response signal at discrete locations over the range of the lidar. It is collected by imaging a reference object with unitary reflectivity over a long acquisition time (100s in this case) to minimize noise. These calibrated response signals can be scaled during reconstruction according to the desired per-pixel acquisition by using Eq.(2.4).

$$F = F_{calib} \left( \frac{\Delta t_0}{100} \right) \quad (2.4)$$

where,  $\Delta t_0$  is the acquisition time in seconds. Assuming that  $y_p$  is drawn from a random vector  $Y_p = (Y_{p1}, \dots, Y_{pm})$  of  $m$  independent Poisson variables will lead us to the likelihood of our observation vector being the mean intensity of the underlying image using Eq.(2.1). 3D reconstruction from SPL data is usually conducted in three sequential steps, i.e, estimation of the baseline intensity  $b_p$ , followed by estimation of response intensity  $x_p$  followed by the estimation of depth  $d_p$  [31,32]. The estimation of  $b_p$  and  $x_p$  is similar except that  $b_p$  is modelled without considering the calibration data.  $b_p$  can be seen simply as an additive intensity value that can be estimated from the first few bins of the histogram. Moreover, the estimation of  $b_p$  facilitates the estimation of  $x_p$  because the system is now described more appropriately. The

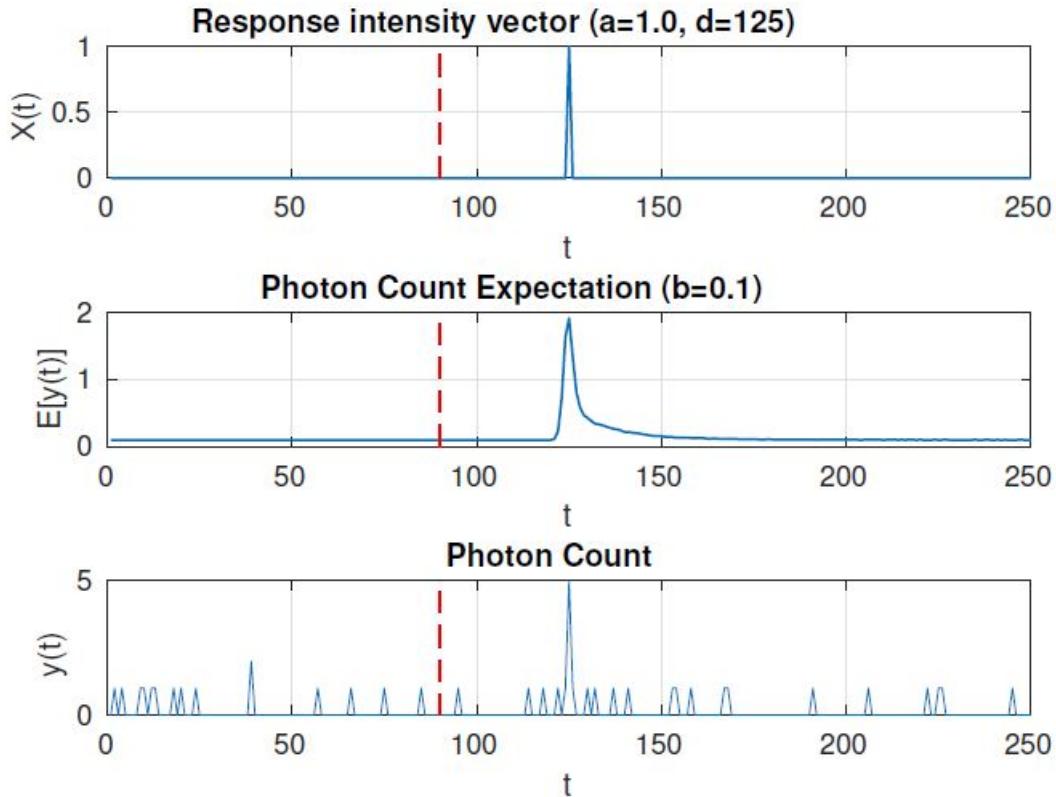


Figure 2.4: An illustration of the measurement model. Top: example of true response intensity per depth. Middle: expected photon count for each time bin that depends on  $a, b$  and  $d$ . Bottom: simulated photon count  $y$ , at each pixel, using the Poisson model

estimation of  $b_p$  and  $x_p$  is convex while the estimation of the depth  $d_p$  is non-convex. Although, non-convexity of  $d_p$  poses a serious challenge in its estimation, the reconstructed baseline and response intensities add more information to the system, thus, facilitating the estimation of  $d_p$ . This is the reason the estimation process follows the sequential steps mentioned above.

The following sub-sections will explain the two different prior likelihood models. This particular Poisson likelihood model forms the base for any minimization problem involving Poisson images. We will see that this model fits this scenario better than the Euclidean distance measure commonly used in compressed sensing and convex optimization problems. This observation has been used in all of the methods mentioned at the beginning of this chapter along with other methods such as multiscale reconstruction, [42], and deblurring, [1]. However, the models proposed in the following sub-sections differ slightly from the ones mentioned in those works in the sense that this model takes into account, the calibration data, a pixel-wise multiplicative

factor compensating for the wavelength vignetting effect and lastly, the background intensity is estimated more efficiently. Figure 2.4 shows an illustration of the proposed observation model. The red dotted line in the images indicates the last time bin that will be considered for baseline intensity estimation.

### 2.3.1 Baseline Intensity

The intensity estimation follows two sequential steps of which, the first step is to estimate the baseline intensity. Provided the data has a histogram of received photons over time, the baseline intensity can be estimated from the first few bins. Only the bins with  $t \leq T_b$  are sufficient to model the background intensity. The inverse problem can be constructed as shown in Eq.(2.5).

$$y_{p,t,l} \sim Pr(b_p) \quad (2.5)$$

Where,  $y_{p,t}$  is the photon count at pixel  $p$ , bin  $t$  and wavelength  $l$ . The baseline intensity  $b_p$  has been modelled here as pixel dependent value, unlike a constant global value as mentioned in [27, 28, 31, 32]. Using, Eq.(2.1), the likelihood can be written as follows, Eq.(2.6).

$$P(y_{p,t,l} = k) = \frac{b_{p,l}^k e^{-b_{p,l}}}{k!} \quad (2.6)$$

Supposing that under the compressed framework, we randomly select a set of pixels  $S_{\alpha,l}$  from the  $P$  pixels for each wavelength  $l$ , and taking logarithm will lead us to the negative log likelihood function as in Eq.(2.7).

$$\mathcal{L}_{Y,\alpha}(B) = \sum_{l=1}^L \sum_{p \in S_{\alpha,l}} [-\log(b_{p,l})\sigma_{p,l} + T_b b_{p,l}] \quad (2.7)$$

Where,  $B$  represents the concatenation of  $L$  matrices  $[b_1, \dots, b_l, \dots, b_L]$  and  $B \in \mathbb{R}^{P \times L}$  (where,  $P$  is the total number of pixels in each wavelength  $l$ ) and  $\sigma_{p,l} = \sum_{t=1}^{T_b} y_{p,t,l}$  is the sum of the response signal for each pixel location. Eq.(2.7) will be known as the observation model for baseline intensity estimation. This will form the base for the minimization problem as it will become evident in the following chapter.

### 2.3.2 Response Intensity

Similarly, the per-pixel inverse problem for reconstructing response intensity can be written as in Eq.(2.8). This will constitute the second step in the reconstruction process that will constitute the true reflectivity of the target without any background noise.

$$y_{p,l} \sim Pr(r_{p,l} F_l x_{p,l} + b_p) \quad (2.8)$$

Where,  $F_l$  is each column of the concatenated calibration matrix  $F = [F_1 \dots F_l \dots F_L]$ , and,  $r_{p,l}$  is a pixel-wise multiplicative factor compensating for the wavelength vignetting effect that is the same as the reduction in brightness at the peripheries of the image often seen in photography or optics. Similarly, the observation model (negative log-likelihood) for this case can be written as shown below, Eq.(2.9).

$$\mathcal{L}_{Y,\alpha}(A|\hat{B}) = \sum_{l=1}^L \sum_{p \in S_{\alpha,l}} [r_{p,l} \sigma_f a_{p,l} + T_a b_{p,l} - \sigma_{p,l} \log(r_{p,l} \sigma_f a_{p,l} + T_a b_{p,l})] \quad (2.9)$$

Where,  $\hat{B}$  is the reconstructed solution matrix of the baseline intensity images over all the images,  $T_a$  is the total number of bins (time-steps) in the photon count histogram and again,  $\sigma_{p,l} = \sum_{t=t_b+1}^{T_a} y_{p,t,l}$ . Similarly to the previous case,  $A$  is considered to be a concatenation of  $L$  column-vector images,  $A \in \mathbb{R}^{P \times L}$ . Eq.(2.7) and Eq.(2.9) will be, respectively, the so-called "data fidelity" terms that will be minimized using appropriate regularizers and ADMM algorithm.

## 2.4 Compressed Sensing

The image estimation inverse problem that was constructed above describes an under-determined system under low-light conditions as many pixels might end up without any photon reception due to various factors. According to the compressed sensing strategy, the reconstruction of an image (vectorized) of dimensions  $N \times 1$  will be from a measurement vector of dimensions  $M \times 1$  with  $M \ll N$ . This tells us that the architecture of the SPL device to record one pixel at a time can be exploited to collect data randomly on fewer pixel locations (upto  $1/8^{th}$  or  $1/16^{th}$  of the total number of pixels in the image). A solution to such highly under-determined linear system can be achieved through optimization by using prior information on the sparsity of the signal in some known basis. Compressed sensing is useful to reduce acquisition time, costs or even health risks, such as, in the cases of MRI, [43–45] and radio astronomy, [46, 47].

If we consider a complex vector  $x \in \mathbb{C}^N$  and an orthogonal basis  $\Psi \in \mathbb{C}^{M \times N}$ , and, if we assume that vector  $x$  is linearly related to another vector  $y \in \mathbb{C}^M$  through a measurement operator  $\Phi \in \mathbb{C}^{M \times N}$ , such that,  $y = \Phi x + n$ , where  $n$  is additive noise, then, the minimization problem can be formulated under the Tikhonov formulation, [1, 28, 31], as shown in Eq.(2.10).

$$\hat{x} = \arg \min_x C(y, \Phi x) + \lambda \|\Psi^\dagger x\|_1 \quad (2.10)$$

1. set  $K = 0$ , choose  $\mu > 0$ ,  $u_0$  and  $d_0$
2. **repeat**
3.  $z_{k+1} \in \arg \min_z f_1(z) + \frac{\mu}{2} \|z - u_k - d_k\|_2^2$
4.  $u_{k+1} \in \arg \min_u f_2(u) + \frac{\mu}{2} \|z_{k+1} - u - d_k\|_2^2$
5.  $d_{k+1} \leftarrow d_k - (z_{k+1} - u_{k+1})$
6.  $k \leftarrow k + 1$
7. **until** convergence or stopping criterion is satisfied

Figure 2.5: Standard ADMM algorithm

where,  $C(\cdot)$  represents any function that can appropriately relate  $x$  to the measurements  $y$ . The traditional approach in compressed sensing framework is to minimize the squared distance between  $y$  and  $\Phi x$ . However, the Euclidean distance is not an appropriate measure for noise statistics of Poisson images. The term  $C(\cdot)$  will be replaced by the Poisson likelihood, Eq.(2.7) and Eq.(2.9), in subsequent chapters.

## 2.5 ADMM Algorithm

As calculating an exact solution of the formulated inverse problem is computationally impossible, optimization can be employed to approximate the solution iteratively. Convex optimization algorithms try to reach the global minimum of a convex cost function. The cost function consists of a convex (see APPENDIX A) and smooth data-fidelity term along with convex but not necessarily smooth regularizer terms. Although, non-convex cost functions can better model the problem, but, there exist no algorithms that can accurately approximate the global minimum for such functions. The ADMM algorithm belongs to the primal-dual class of convex optimization algorithms. The ADMM algorithm is not new but it is still widely used due to its accuracy and versatility. The goal is now to frame a cost function along with appropriate regularizer terms that can be minimized using ADMM algorithm. The standard minimization problem under the ADMM framework can be written as follows

$$\min f(z) + g(u); S.T.: Az + Bu = c \quad (2.11)$$

where,  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g: \mathbb{R}^m \rightarrow \mathbb{R}$ ,  $z \in \mathbb{R}^n$ ,  $u \in \mathbb{R}^m$ ,  $A \in \mathbb{R}^{p \times n}$ ,  $B \in \mathbb{R}^{p \times m}$  and  $c \in \mathbb{R}^p$ . This minimization problem can then be solved using ADMM algorithm as shown in Figure 2.5.

However, our cost function, as stated in the previous section, consists of more than two terms including the Poisson likelihood function and regularizer terms. [1] has a detailed explanation on a variant of the ADMM algorithm that is able to model cost functions with more than two terms. Under the framework of the ADMM variant, a minimization problem can be written as follows

$$\min_{z \in \mathbb{R}^N} \sum_{j=1}^J g_j H^{(j)}(z) \quad (2.12)$$

where,  $g_j(\cdot)$  are closed, proper, convex functions and  $H^{(j)} \in \mathbb{R}^{m \times n}$  are arbitrary matrices. According to [1], this problem can be written in the form of Eq.(2.11) with the following considerations

$$f_1 = 0$$

$$G = \begin{bmatrix} H^{(1)} \\ \vdots \\ H^{(J)} \end{bmatrix} \in \mathbb{R}^{m \times n} \quad (2.13)$$

where,  $m = m_1 + \dots + m_J$ , and  $f_2 : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$  given by,

$$f_2(u) = \sum_{j=1}^J g_j(u^{(j)}) \quad (2.14)$$

where,  $u^{(j)} \in \mathbb{R}^{m_j}$  and  $u = [(u^{(1)})^T, \dots, (u^{(J)})^T]^T \in \mathbb{R}^m$ .

With this structure, our minimization problem can still be solved using the ADMM algorithm shown in Figure 2.5 with some small changes. The variables  $u_k$  and  $d_k$  will now be a collection of matrices as shown below

$$d_k = \begin{bmatrix} d_k^{(1)} \\ \vdots \\ d_k^{(J)} \end{bmatrix} \quad u_k = \begin{bmatrix} u_k^{(1)} \\ \vdots \\ u_k^{(J)} \end{bmatrix}. \quad (2.15)$$

Since  $f_1 = 0$ , step 3 of the algorithm can now be seen as a simple quadratic minimization problem that has a unique solution if  $G$  is a full column rank matrix, i.e,

$$\begin{aligned} \arg \min_z \|Gz - \zeta_k\|_2^2 &= (G^T G)^{-1} G^T \zeta_k \\ &= \left[ \sum_{j=1}^J (H^{(j)})^T H^{(j)} \right]^{-1} \sum_{j=1}^J (H^{(j)})^T \zeta_k^{(j)} \end{aligned} \quad (2.16)$$

where,  $\zeta_k = u_k + d_k$  (and  $\zeta_k^{(j)} = u_k^{(j)} + d_k^{(j)}$ ) and the second equation has been written using Eq.(2.13). By decoupling step 4 into  $J$  independent steps, it can be written as

$$\begin{bmatrix} u_{k+1}^{(1)} \\ \vdots \\ u_{k+1}^{(J)} \end{bmatrix} \leftarrow \arg \min_{u^{(1)}, \dots, u^{(J)}} g_1(u^{(1)}) + \dots + g_J(u^{(J)}) + \frac{\mu}{2} \left\| \begin{bmatrix} H^{(1)} \\ \vdots \\ H^{(J)} \end{bmatrix} z_{k+1} - \begin{bmatrix} u^{(1)} \\ \vdots \\ u^{(J)} \end{bmatrix} - \begin{bmatrix} d_k^{(1)} \\ \vdots \\ d_k^{(J)} \end{bmatrix} \right\|_2^2 \quad (2.17)$$

In other words, the solution for each term can be computed sequentially at each iteration given by the following equation

$$u_{k+1}^{(j)} \leftarrow \arg \min_{v \in \mathbb{R}^{m_j}} g_j(v) + \frac{\mu}{2} \|v - s_k^{(j)}\|_2^2 \quad (2.18)$$

for  $j = 1, \dots, J$ , where,  $s_k^{(j)} = H^{(j)} z_{k+1} - d_k^{(j)}$ .

This chapter has explained clearly the structure of the cost function that needs to be minimized. Along with this variant of the ADMM algorithm shown above, we can now move on to construct our minimization problem using the likelihood functions, Eq.(2.7) & Eq.(2.9), and some appropriate regularizer terms. In the next chapter, some regularizer terms will be proposed and we will see how they impact the reconstruction, specially, at low-light ( $<1$  mean photons per pixel) conditions.

### 2.5.1 Varying Step Size

Before moving on to the next chapter, we can discuss how to vary the step size  $\mu$  at each iteration to achieve convergence faster. Given a minimization problem of the form Eq.(2.11), the primal and dual residuals for the primal and dual feasibility conditions, respectively, can be defined as follows [48]. The dual residual at iteration  $k + 1$  is given by

$$s^{k+1} = \mu A^T B(z^{k+1} - z^k) \quad (2.19)$$

while, the primal residual at iteration  $k + 1$  is given by

$$r^{k+1} = Ax^{k+1} + Bz^{k+1} - c \quad (2.20)$$

Now, a standard approach is to use possibly different penalty parameters  $\mu^k$  for each iteration, to improve convergence, as well as making performance less dependant on the initial choice of the penalty parameter. As mentioned in [48], a simple scheme to vary the penalty parameter at each iteration can be given as follows,

$$\mu^{k+1} := \begin{cases} \tau^{incr} \mu^k & \text{if } \|r^k\|_2 > \rho \|s^k\|_2 \\ \mu^k / \tau^{decr} & \text{if } \|s^k\|_2 > \rho \|r^k\|_2 \\ \mu^k & \text{otherwise} \end{cases} \quad (2.21)$$

where,  $\rho > 1$ ,  $\tau^{incr} > 1$ , and  $\tau^{decr} > 1$  are predefined parameters. The idea behind varying this penalty parameter is to try to keep the primal and dual residual norms within a factor of  $\rho$  of one another as they both converge to zero. The adjustment scheme Eq.(2.21) inflates  $\mu$  by  $\tau^{incr}$  when the primal residual appears large compared to the dual residual, and deflates  $\mu$  by  $\tau^{decr}$  when the primal residual seems very small relative to the dual residual.

Additionally, when a varying penalty parameter is used in the scaled form of ADMM, the scaled dual variable  $u^k = (1/\mu)y^k$  must also be re-scaled after updating  $\mu$ . For example, if  $\mu$  is halved,  $u^k$  should be doubled before proceeding.

# Chapter 3

## Methodology

In this chapter we will see what regularizer terms are appropriate for image reconstruction from Hyper-spectral lidar data. The regularizer terms are some prior information about the scene that are imposed on the observation model as penalty terms. The prior information that is generally used under compressed sensing framework can be some sparsity coefficients in the gradient domain or some wavelet basis. Two novel minimization problems have been proposed in this work. The following sections will explain in detail, the impact, accuracy and computational complexity of both the proposed minimization problems. As the interest of work is image reconstruction, a positivity constraint is also enforced on the minimization problem to obtain a solution that is never negative. Precisely, one problem consists of TV norm and nuclear norm regularization, while, the other problem consists of a new joint sparsity regularization method that is more accurate and at the same time computationally less expensive. However, both the methods work similarly, as in, they both promote sparsity, smoothness and spatial and spectral correlation that are very important in the context of hyper-spectral data.

In most of the cases, the peak wavelengths at which the scene is imaged are close to one another. For example, for the data used in this work, the peak wavelengths range from 500nm to 820nm at steps of 10nm. Moreover, the surface responsible for generating response at each pixel location is the same over all the wavelengths. This implies that it is sensible to assume a high correlation between the intensity images over different wavelengths. Additionally, the change in reflectivity of the scene with the imaging wavelength can be used for material classification. Although, material classification is out of the scope of this work, the primary goal is to obtain accurate intensity reconstructions at all the wavelengths simultaneously.

### 3.1 TV and Nuclear Norm Regularization Model

This section will describe the construction of the minimization problem using the log-likelihood functions, described under Section (2.3) along with Total Variation (TV) and Nuclear Norm (NN) regularization terms. The resulting minimization problem will have a structure of the Tikhonov regularization problem, whose solution can be efficiently approximated with the ADMM algorithm variant described previously.

The first use of Total Variation for image denoising and reconstruction can be traced back to 1992 in a celebrated paper by Rudin, Osher and Fatemi, [49]. However, more efficient yet simple algorithms for Total Variation minimization algorithms were developed in the 2000s, [50, 51]. These algorithms are fast and fall under the category of gradient-search algorithms. The concept of regularization in minimization problems arises from a Bayesian perspective. An appropriate regularizer should impart useful prior information to the system. Total Variation is one such convex regularization term that describes spatial correlation (smoothness) over an image while preserving the edges. Due to its convexity, a solution maximizing the likelihood of the image given the observation can be accurately estimated using convex optimization. [52], contains a detailed overview on the working and convexity of TV regularization in a convex minimization problem.

Total variation describes a phenomenon that is similar to the energy of a signal. Mathematically, it is the integral of the absolute gradient of the signal (or image in this context). We know that the gradient of an image describes the edges. Exactly in the same way Total Variation describes the detail present in an image. For example, images with excessive or spurious detail will have high TV. Noise can also be seen as an unwanted detail in the image that can increase the TV of the image. Thus, minimizing a problem while keeping the TV low (subject to the observed signal) removes unwanted detail while preserving important details such as edges. TV is effective in denoising applications by smoothing away the noise in flat regions while preserving the edges. So, with TV regularization one can achieve denoising, image in-painting over smooth regions while preserving the edges, simultaneously. This also explains why TV can promote both sparsity and spatial correlation.

Another regularization term that is popular in reconstruction from multi-spectral images is the Nuclear Norm Regularization. Let us assume the images corresponding to each peak wavelength vectorized and stacked along the columns as a data cube  $X \in \mathbb{R}^{N \times L}$ , where  $N$  is the total number of pixels in each image imaged over  $L$  wavelengths. The nuclear norm over the data cube is represented by  $\|X\|_*$ , where  $X = U\Sigma V^T$  is the SVD decomposition of  $X$ . Assuming that  $X$  is full column rank, what we are actually doing by the SVD decomposition is to represent our  $N$  row vectors with an approximation of a linear combination of the first few singular vectors, thus, promoting low rank matrices (since  $L \ll N$ ). It is well known

that an image can be accurately represented by the first few singular vectors from it's SVD decomposition.

The nuclear norm works by estimating from the sum of absolute singular values (we have  $L$  absolute singular values in this case) over all the imaged wavelengths  $L$ . In other words, nuclear norm enhances the spectral correlation present in each row of our data cube  $X$ . As the same scene is imaged over all the wavelengths, correlation over the spectra can be expected to be very high and nuclear norm imparts information about this spectral correlation into the minimization problem. Moreover, this term is also convex and can be added to the minimization problem as an additional regularization term. NN regularization has been shown to be very efficient in the context of multi-spectral radio-interferometry imaging, [46, 53].

### 3.1.1 Minimization problem

it is important to recollect that the reconstruction process follows two sequential steps of estimating the baseline and response intensities  $\hat{B}$  and  $\hat{A}$  one after the other. The minimization problem under this context will consist of the Poisson log-likelihood observation model (Eq.(2.7) and Eq.(2.9), respectively) along with the two regularization terms mentioned above. The TV regularization will be performed individually over each image. For simplicity, it will be called Spectral TV Regularization, STV, and it can be represented as follows

$$STV(X) = \sum_{l=1}^L TV(x_l) \quad (3.1)$$

where,  $X$  will correspond to  $B$  or  $A$  depending on whether we want to estimate the baseline intensity or the response intensity. Combining all this information, the minimization problem to estimate the intensity images can be written as follows

$$\hat{X} = \arg \min_X [\mathcal{L}_{Y,\alpha}(X) + \tau_1 STV(X) + \tau_2 \|X\|_* + i_{\mathbb{R}^+}(X)] \quad (3.2)$$

where,  $X = B$  and  $\mathcal{L}_{Y,\alpha}(X) = \mathcal{L}_{Y,\alpha}(B)$  for baseline intensity reconstruction and  $X = A$  and  $\mathcal{L}_{Y,\alpha}(X) = \mathcal{L}_{Y,\alpha}(A|\hat{B})$  for response intensity reconstruction, and,  $i_{\mathbb{R}^+}(X)$  is an indicator function enforcing non-negative values on the solution  $X$ . It might be interesting to note that Eq.(3.2) follows the same structure as that of Eq.(2.12) with  $J = 4$ .

### 3.1.2 Algorithmic Details

Now that the minimization problem has been framed, we will look into how to solve it using the ADMM algorithm. ADMM, in general, guarantees convergence and does not depend on initialization as long as the problem is convex. The algorithm to solve the minimization problem

stated in Eq.(3.2) is shown in Figure 3.1. One might note the smaller minimization problems over one variable in steps 14, 16 and 18 of the algorithm. The solution to these steps is given by the so called "Moreau proximity operator" (see APPENDIX B). Following the definition of the proximity operator, the solution to step 14 can be given as follows

$$\begin{aligned} prox_{\mathcal{L}_{Y,\alpha/\mu}}(N) &= \arg \min_A [\mathcal{L}_{Y,\alpha}(A|\hat{B}) + \frac{\mu}{2} \|A - N\|_2^2] \\ &= \frac{1}{2} \left( N - \frac{\sigma_f R}{\mu} - \frac{T_a \hat{B}}{\sigma_f R} + \sqrt{\left( N - \frac{\sigma_f R}{\mu} - \frac{T_a \hat{B}}{\sigma_f R} \right)^2 + \frac{4\Sigma}{\mu} + \frac{4T_a \hat{B} \odot N}{\sigma_f R} - \frac{4T_a \hat{B}}{\mu}} \right) \end{aligned} \quad (3.3)$$

where,  $\odot$  represents element-wise product and all the divisions are element-wise,  $R$  is the wavelength vignetting effect and  $\Sigma = \sum_{t=1}^{T_1} y_t$ ,  $\Sigma \in \mathbb{R}^{P \times L}$ , where  $P$  is the number of pixels in each image and  $L$  is the total number of imaged wavelengths. Computing the proximal value of TV operator (step 16) at each iteration is a sub-iterative process. [1] contains detailed information on the implementation of this operator. The solution to step 18 is the soft thresholding operator on the singular values of the matrix, that can be given by

$$prox_{\|\nu\|_*} = \text{soft}\left(\Sigma, \frac{2\tau_2}{\mu}\right) \quad (3.4)$$

where,  $\|\nu\|_* = \|\Sigma\|_{1,1} = \sum_i |\sigma_i|$ , with  $\nu = U\Sigma V^T$  the SVD decomposition of  $\nu$ . The algorithm will be complete when these proximal solutions to the corresponding terms are replaced in the steps 14, 16 and 18 respectively.

This particular model involves the SVD decomposition of a large matrix at each iteration because of the presence of the nuclear norm. Computing the SVD decomposition of such large matrices at each iteration can be computationally expensive. Moreover, this model requires manual parameter tuning of two regularization parameters,  $\tau_1$  and  $\tau_2$ . The reconstruction and the results shown in the next section depend on the fine tuned parameter values for both the regularization parameters. Although, there exist methods to automatically tune the regularization parameter during subsequent iterations, they apply only to the case of a single regularization parameter. Automatic tuning of regularization parameter cannot be applied here because of the presence of two parameters.

## 3.2 Joint Sparsity Model

A new regularization term that promotes joint sparsity and low-rankness on the data matrix was proposed recently in the context of radio astronomy, [46]. The term 'joint' refers to sparsity

1. **Input:**  $M_\alpha, \mu, \tau_1, \tau_2, \Sigma, \sigma_f, \hat{B}$  (only if estimating  $A$ )
2. **Result:**  $\hat{B}$  or  $\hat{A}$
3. Initialize  $u_0^{(1)}, u_0^{(2)}, u_0^{(3)}, u_0^{(4)}, d_0^{(1)}, d_0^{(2)}, d_0^{(3)}, d_0^{(4)}$
4. Pre-compute  $(M_\alpha^T M_\alpha + 3I)^{-1}$
5. Set  $k = 0, \mu > 0, \tau_1 > 0, \tau_2 > 0$
6. **while**  $\|z_k - z_{k-1}\|_2 / (\min(\|z_k\|_2, \|z_{k-1}\|_2) + \epsilon)$  AND  $k \leq k_{max\_iter}$  **do**

  7.  $\zeta_k^{(1)} \leftarrow u_k^{(1)} + d_k^{(1)}$
  8.  $\zeta_k^{(2)} \leftarrow u_k^{(2)} + d_k^{(2)}$
  9.  $\zeta_k^{(3)} \leftarrow u_k^{(3)} + d_k^{(3)}$
  10.  $\zeta_k^{(4)} \leftarrow u_k^{(4)} + d_k^{(4)}$
  11.  $\gamma_k \leftarrow M_\alpha^T \zeta_k^{(1)} + \zeta_k^{(2)} + \zeta_k^{(3)} + \zeta_k^{(4)}$
  12.  $z_{k+1} \leftarrow (M_\alpha^T M_\alpha + 3I)^{-1} \gamma_k$
  13.  $\nu_k^{(1)} \leftarrow M_\alpha z_{k+1} - d_k^{(1)}$
  14.  $u_{k+1}^{(1)} \leftarrow \arg \min_v \frac{\mu}{2} \|v - \nu_k^{(1)}\|_2^2 + \mathcal{L}_{Y,\alpha}(v)$
  15.  $\nu_k^{(2)} \leftarrow z_{k+1} - d_k^{(2)}$
  16.  $u_{k+1}^{(2)} \leftarrow \arg \min_v \frac{\mu}{2} \|v - \nu_k^{(2)}\|_2^2 + \tau_1 STV(v)$
  17.  $\nu_k^{(3)} \leftarrow z_{k+1} - d_k^{(3)}$
  18.  $u_{k+1}^{(3)} \leftarrow \arg \min_v \frac{\mu}{2} \|v - \nu_k^{(3)}\|_2^2 + \tau_2 \|v\|_*$
  19.  $\nu_k^{(4)} \leftarrow z_{k+1} - d_k^{(4)}$
  20.  $u_{k+1}^{(4)} \leftarrow \max(0, \nu_k^{(4)})$
  21.  $d_{k+1}^{(1)} \leftarrow d_k^{(1)} - (M_\alpha z_{k+1} - u_{k+1}^{(1)})$
  22.  $d_{k+1}^{(2)} \leftarrow d_k^{(2)} - (z_{k+1} - u_{k+1}^{(2)})$
  23.  $d_{k+1}^{(3)} \leftarrow d_k^{(3)} - (z_{k+1} - u_{k+1}^{(3)})$
  24.  $d_{k+1}^{(4)} \leftarrow d_k^{(4)} - (z_{k+1} - u_{k+1}^{(4)})$
  25.  $k \leftarrow k + 1$
  26. **end**
  27.  $\hat{B} \leftarrow z_k$  or  $\hat{A} \leftarrow z_k$

Figure 3.1: Variant of the PIDAL-TV algorithm, [1]

in both the spatial and spectral domains simultaneously. This model is based on the assumption that the photons received over the image can be seen to be reflected from a finite number of sources,  $\rho$ , each with a different spectral signature. Under this assumption, a linear mixture model  $X = SH^\dagger$  can be adopted, where the columns of the matrix  $S \in \mathbb{C}^{N \times \rho}$  represent the responsible sources present in the image and the columns of the matrix  $H \in \mathbb{C}^{\rho \times L}$  are their corresponding spectral signatures.

This assumption is sensible because the image would actually be composed of only a few objects that have distinctive spectral signatures. For example, the data used in this work has images consisting of about 15 objects of different colors. The rank of the matrix  $X$  is given by  $\rho$ , that implies low-rankness. Suppose if none of the sources are active at a given pixel location, then one whole row of  $X$  will be zero. This property implies joint-sparsity over the model. For this work we assume that  $X$  is the image to be reconstructed that is already composed in the above mentioned form. This reduces complexity as solving explicitly for  $S$  and  $H$  would comprise of a non-linear and non-convex problem. Moreover, source/material classification is not of interest in this work.

### 3.2.1 Minimization Problem

Similarly, as in the previous case, a minimization problem can be framed by combining the log-likelihood observation model with the joint-sparsity regularization term. The regularization term can be defined by encapsulating joint-sparsity in some adequate basis  $\Psi$  along with an analysis prior based on the  $l_{2,1}$  norm. The minimization problem can then be defined as

$$\hat{X} = \arg \min_X [\mathcal{L}_{Y,\alpha}(X) + \tau_1 \|\Psi^\dagger X\|_{2,1} + i_{\mathbb{R}^+}(X)] \quad (3.5)$$

where,  $i_{\mathbb{R}^+}$  is again an indicator function enforcing non-negativity on the solution. Similarly as in the previous model,  $X = B$  when reconstructing baseline intensity and  $X = A$  along with  $\mathcal{L}_{Y,\alpha}(X) = \mathcal{L}_{Y,\alpha}(A|\hat{B})$  = when estimating response intensity. In Eq.(3.5) above,  $\|\Psi^\dagger X\|_{2,1}$  stands for the component-wise  $l_{2,1}$  norm. This regularization term promotes smoothness of the spectral lines along with joint-sparsity in basis  $\Psi$ .  $\Psi$  can be any wavelet basis in which the data can be assumed to be sparse.

### 3.2.2 Algorithmic Details

As, the joint-sparsity minimization problem, Eq.(3.5) has the same structure as Eq.(3.2), the resulting algorithm will also be similar to the one shown previously. Again, the proximal solutions to each of the terms must be calculated at each iteration. The proximal solution to the log-likelihood term will be the same as Eq.(3.3) and so is the case for the solution of the

indicator function that is, again, simply a projection of the variable onto the first orthant, as shown below,

$$u_{k+1}^{(3)} = \max \{ \nu_k^{(3)}, 0 \}. \quad (3.6)$$

The  $\|\cdot\|_{2,1}$  operator is nothing but the  $l_1$  norm of the row-wise  $l_2$  norm of the matrix. The proximal solution to the regularization term is given by the soft-thresholding operation, defined for each row  $k$  as,

$$(\mathcal{S}_\alpha^{l_{2,1}}(Z))_{k,:} \triangleq \begin{cases} \bar{z} \frac{\|\bar{z}\|_{l_2} - \alpha}{\|\bar{z}\|_{l_2}} & \|\bar{z}\|_{l_2} > \alpha \\ 0 & \|\bar{z}\|_{l_2} \leq \alpha \end{cases} \quad \forall k \quad (3.7)$$

where,  $\bar{z} = z_k$ , is the row  $k$  of the input matrix  $Z$  and  $\alpha$  can be chosen equal to  $\frac{2\tau}{\mu}$ . The algorithm is shown in Figure 3.2. Eq.(3.6) and Eq.(3.7) correspond to the solutions to the steps 17 and 15 in the algorithm, respectively.

This model is simpler and efficient than the TVNN model shown in the previous section in various aspects :

- There is only one regularization parameter that needs to be tuned
- Absence of the nuclear norm term dramatically increases speed
- This model promotes both spatial and spectral correlation through a single regularization term

By choosing a simple and appropriate wavelet domain for  $\Psi$ , this algorithm can also be run in real time. Now that two different minimization problems have been defined in detail, we can move on to the next section to see a comparison between the two methods in terms of performance and accuracy.

1. **Input:**  $M_\alpha$ ,  $\mu$ ,  $\tau$ ,  $\Sigma$ ,  $\sigma_f$ ,  $\hat{B}$  (only if estimating  $A$ )
2. **Result:**  $\hat{B}$  or  $\hat{A}$
3. Initialize  $u_0^{(1)}$ ,  $u_0^{(2)}$ ,  $u_0^{(3)}$ ,  $d_0^{(1)}$ ,  $d_0^{(2)}$ ,  $d_0^{(3)}$ ,
4. Pre-compute  $(M_\alpha^T M_\alpha + 2I)^{-1}$
5. Set  $k = 0, \mu > 0, \tau > 0$
6. **while**  $\|z_k - z_{k-1}\|_2 / (\min(\|z_k\|_2, \|z_{k-1}\|_2) + \epsilon)$  AND  $k \leq k_{max\_iter}$  **do**
7.      $\zeta_k^{(1)} \leftarrow u_k^{(1)} + d_k^{(1)}$
8.      $\zeta_k^{(2)} \leftarrow u_k^{(2)} + d_k^{(2)}$
9.      $\zeta_k^{(3)} \leftarrow u_k^{(3)} + d_k^{(3)}$
10.     $\gamma_k \leftarrow M_\alpha^T \zeta_k^{(1)} + P^T \zeta_k^{(2)} + \zeta_k^{(3)}$
11.     $z_{k+1} \leftarrow (M_\alpha^T M_\alpha + 2I)^{-1} \gamma_k$
12.     $\nu_k^{(1)} \leftarrow M_\alpha z_{k+1} - d_k^{(1)}$
13.     $u_{k+1}^{(1)} \leftarrow \arg \min_v \frac{\mu}{2} \|v - \nu_k^{(1)}\|_2^2 + \mathcal{L}_{Y,\alpha}(v)$
14.     $\nu_k^{(2)} \leftarrow P z_{k+1} - d_k^{(2)}$
15.     $u_{k+1}^{(2)} \leftarrow \arg \min_v \frac{\mu}{2} \|v - \nu_k^{(2)}\|_2^2 + \tau \|\Psi^\dagger v\|_{2,1}$
16.     $\nu_k^{(3)} \leftarrow z_{k+1} - d_k^{(3)}$
17.     $u_{k+1}^{(3)} \leftarrow \max(0, \nu_k^{(3)})$
18.     $d_{k+1}^{(1)} \leftarrow d_k^{(1)} - (M_\alpha z_{k+1} - u_{k+1}^{(1)})$
19.     $d_{k+1}^{(2)} \leftarrow d_k^{(2)} - (P z_{k+1} - u_{k+1}^{(2)})$
20.     $d_{k+1}^{(3)} \leftarrow d_k^{(3)} - (z_{k+1} - u_{k+1}^{(3)})$
21.     $k \leftarrow k + 1$
22. **end**
23.  $\hat{B} \leftarrow z_k$  or  $\hat{A} \leftarrow z_k$

Figure 3.2: Variant of the PIDAL-FA algorithm, [1]

# Chapter 4

## Results

The two methods proposed above were tested on real data consisting of 32 images, each of dimension  $190 \times 190$  pixels, imaged at peak wavelengths of light ranging from 500nm to 810nm in steps 10nm. Figure 4.1 shows the setup that was used to acquire this real data. As it can be seen the scene consists of about 15 differently colored and shaped pieces of clay to generate varied spectral responses. This data was acquired for 10ms for each pixel location over each wavelength step. As the reflectivity of the materials depends upon the wavelength of the light pulse, the average photon counts of the measurements over the data varies from  $\sim 230$  to  $\sim 1800$  photons per pixel depending on the wavelength used for the image. The methods have been tested only on real data as generating multi-spectral synthetic data is impossible when the spectra are unavailable.

To evaluate the results quantitatively, the reconstructions, from data at full measurements, obtained using the TVNN method were used as the ground truth images. Figure 4.2 shows an example of the ground truth images at 6 peak wavelengths (500nm, 560nm, 620nm, 680nm, 740nm and 800nm). These reconstructions were obtained with very low values for the two regularization parameters (see Eq.(3.2)) to keep their impact low. Two metrics, namely, Mean Squared Error (MSE) and Signal to Noise Ratio (SNR) were used to compare the new reconstructions with the shown ground truth response intensity images. As, both the metrics mentioned above have a linear relationship, the results here will be shown only in terms of the SNR value. The SNR metric can be defined as

$$SNR = 10 \log \left( \frac{\|\hat{x}\|_2^2}{\|\hat{x} - x\|_2^2} \right) \quad (4.1)$$

where,  $\hat{x}$  is the ground truth image matrix ( $\hat{x} \in \mathbb{R}_+^{N \times L}$ ) and  $x$  is the estimated intensity image matrix of same dimensions, with  $N$  as the number of pixels in each image and  $L$  as the



Figure 4.1: The setup that was repeatedly imaged over different peak wavelengths using the mentioned SPL

total number of wavelengths, and,  $\|.\|_2^2$  represents the squared  $l_2$  norm.

The acquisition time for the data can be scaled to alter the expected photon counts per pixel. Table 4.1 shows the correspondences between different average photon counts over the data and the corresponding sub-sampling ratios. The relation is that when the average photon counts per pixel over each image is doubled, the number of considered pixels in the image is halved. Such an approach highlights the benefit of compressed sensing under low-light condition. The table mentions the average photon counts per pixel over each band, so, for the data matrix  $X$  consisting the images over each wavelength vertically stacked, the average photons per pixel will be 32 times the mentioned value.

The joint sparsity model was tested using Discrete Cosine Transform (DCT) and Daubechies Wavelet Transform (DWT) as two separate sparsity bases. Both of these bases  $\Psi$  have shown to promote sparsity for the tested data. Daubechies wavelets have been widely used image processing for basic applications such as filtering, compression etc., to complex applications such as, image in-painting etc.. Daubechies wavelets have a finite number of non-zero scaling coefficients meaning that they have a compact support. Another property of *order – K* Daubechies wavelets is that the first  $k$ -moments are zero that makes them useful as basis functions. They

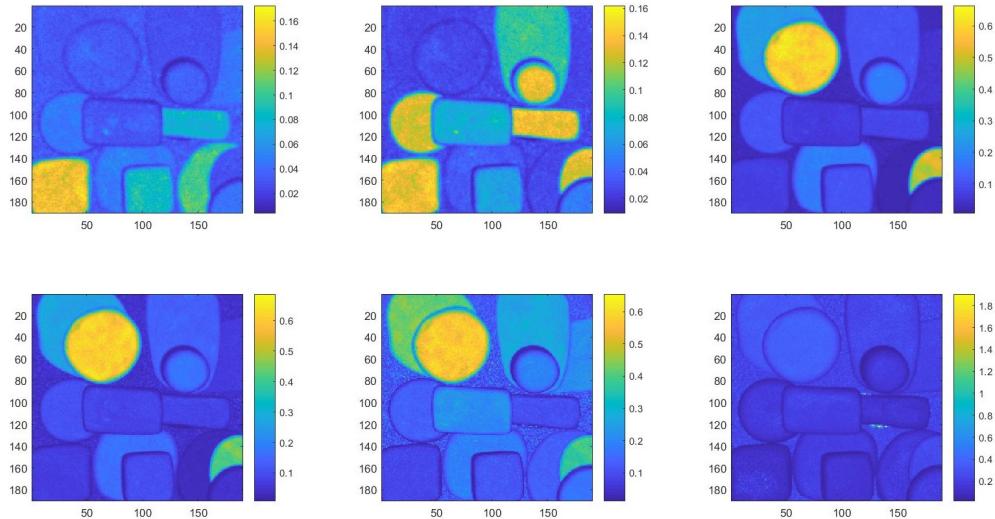


Figure 4.2: Reconstructions from data at high average photon counts that was used as ground truth for quantitative analysis (showing only 6 out of 32 images)

Table 4.1: Table showing the correspondence of sub-sampling ratio ( $\alpha$ ) with the mean photons per pixel (ppp) per each band of wavelength

mean ppp (before sampling)	sub-sampling ratio	mean ppp (after sub-sampling)
0.5, 1, 2, 4, 8	1, 1/2, 1/4, 1/8, 1/16	0.5
1, 2, 4, 8, 16	1, 1/2, 1/4, 1/8, 1/16	1
10, 20, 40, 80, 160	1, 1/2, 1/4, 1/8, 1/16	10
50, 100, 200	1, 1/2, 1/4	50
100, 200	1, 1/2	100

Table 4.2: Table showing the SNR (dB) values for the reconstructions obtained using the proposed methods for different average values of photons per pixel (rows) against different sub-sampling ratios (columns).

$\alpha$	1	1/2	1/4	1/8	1/16
100	TVNN	<b>53.08</b>	50.48		
	JS-DCT	<b>53.59</b>	49.89		
	JS-DWT	<b>54.59</b>	51.16		
50	TVNN	46.74	<b>48.69</b>	46.33	
	JS-DCT	<b>47.95</b>	47.27	42.68	
	JS-DWT	<b>48.99</b>	47.97	43.79	
10	TVNN	34.38	41.33	<b>42.79</b>	39.75
	JS-DCT	37.58	<b>42.12</b>	40.22	36.03
	JS-DWT	39.29	<b>43.14</b>	41.43	37.34
1	TVNN	22.65	27.77	34.42	<b>35.89</b>
	JS-DCT	28.25	30.64	<b>31.06</b>	30.34
	JS-DWT	27.22	27.31	<b>27.41</b>	27.15
0.5	TVNN	20.91	21.32	22.33	<b>22.53</b>
	JS-DCT	26.68	<b>26.95</b>	26.89	<b>26.78</b>
	JS-DWT	26.70	28.75	<b>29.67</b>	28.99

are also orthogonal wavelets that enable the use of compressed sensing in the inverse problem. They also promote smoothness over the image which is a general property for most real-world images. The first eight orthogonal Daubechies bases were used in these tests.

The Discrete Cosine Transform is closely related to the Discrete Fourier Transform, real and implicitly performs symmetric extension. One of the major advantages of DCT is that only 1% of coefficients are sufficient to approximate the image, thus, promoting sparsity. DCT is also an orthogonal basis and promotes smoothness in the image. DCT has also been widely used in image processing applications, primarily in compression. The method using Nuclear norm and TV regularization will be called TVNN and the joint sparsity based method with the DCT and DWT wavelets bases will be called JS-DCT and JS-DWT, respectively, for simplicity.

Table 4.2 shows the SNR values for the reconstructions from the data with different average photon counts per pixel at different levels of sub-sampling. The table is divided into cells of three rows each corresponding to the three proposed methods for some fixed average number of photons per pixel (first column). The best reconstruction for each cell is highlighted as the bold SNR value. It may be observed from the highlighted values that JS-DWT method has the best reconstructions out-performing the TVNN method by a margin, specially, in the case low measurements ( $\leq 10$  ppp). The three methods have equally good reconstructions from data with high number of detected photons. The improvement with compressed sensing can be seen

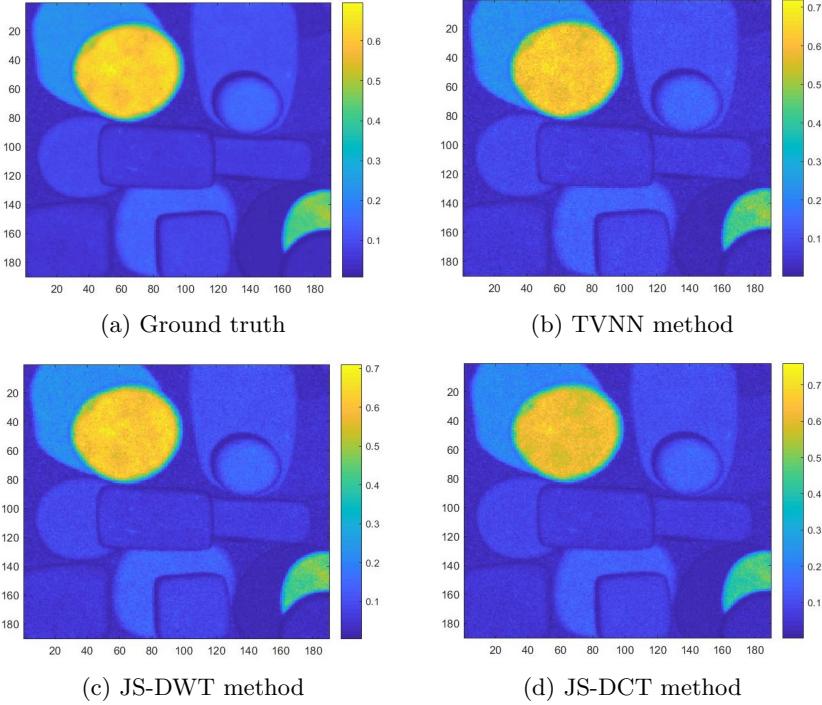


Figure 4.3: Response Intensity image reconstruction results using the three methods at high measurements ( $\sim 100$  photons per pixel per band) at the 15<sup>th</sup> band

at low observations (shown with an underline highlighting the best value of that row). It can be observed that in some cases, it is better to consider fewer samples (upto 1/8<sup>th</sup> of the total number of pixels) than to measure data at all the pixel locations. This improvement applies only to the case of low number of detected photons as the data is not sparse enough when the number of photons is high, even after considering fewer samples. It can also be noted that the TVNN method shows big improvement with compressed sensing both quantitatively and visually.

The results for the three methods are also visually appealing. It is an achievement by the two models to obtain visually good results, given that, sometimes data with only  $\sim 1\%$  of observations is considered. Figure 4.3 shows the reconstruction of the response intensity images for one wavelength from the data with an average of 100 photons per pixel per band. The varied effect of smoothness from the three different images can be seen from those images. It may also be noted that the three methods have equally appealing results, visually. The difference between the three methods can be seen clearly for the case of reconstruction at low observations.

Figure 4.4 shows the reconstructed response intensity images for the three methods at dif-

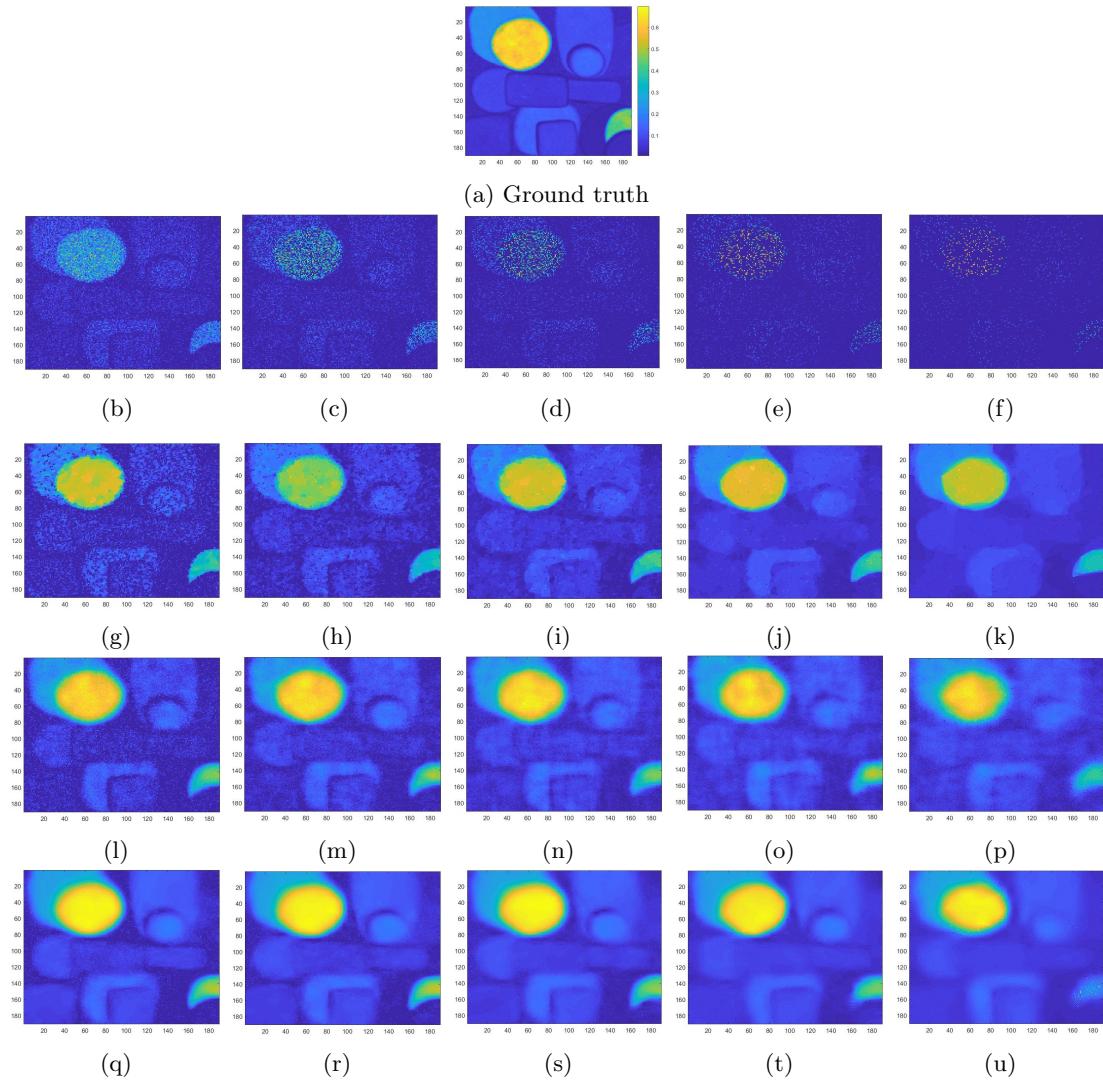


Figure 4.4: Response Intensity image reconstruction results using the three methods at low measurements ( $\sim 1$  photon per pixel per band) at the 15<sup>th</sup> band. Row 2 ((b)-(f)) show the measurements at different sub-sampling ratios from 1 – 1/16 with an overall average of 1 photon per pixel, respectively. Rows 3,4 and 5 show the respective reconstructed images using TVNN, JS-DCT and JS-DWT methods from the measurements

ferent number of considered samples. The measurements are also shown for a comparative evaluation of the results. These images also show the improvement with compressed sensing. It may be observed that smoothness over the image increases while the noise decreases with sampling fewer pixels to provide more detail in the image. This improvement is more evident in the results from TVNN method. However, it might be interesting to note that there is a limit to how few samples can be considered. The results, in general, degrade when  $1/16^{th}$  of the pixels are considered. Hence, it may be said that we still need to sample at least  $1/8^{th}$  of the pixels to obtain visually satisfactory results.

From these images, it can also be said that JS-DWT method is better than the rest in preserving detail or edges to provide better contrast in the reconstructions. However the joint sparsity model does not show enough improvement as the TVNN method when compressed sensing is considered. Lastly, it has been observed that data with  $\sim 0.5$  photon per pixel per band is the limit to obtain visually satisfactory results. The methods start to collapse when the measurements go lower than that. However, again a comparison can be made between the TVNN and the joint sparsity models, while looking at reconstructions at different wavelengths for both the methods. Before looking at the images, it may be said that the joint sparsity method outperforms the TVNN method with good results over all the wavelengths even at  $\sim 0.5$  photons per pixel. Figure 4.5 shows that the joint sparsity method performs very well over all the wavelengths even from data with extremely low measurements. As the distribution of intensities is different over each wavelength, taking measurements lower than  $\sim 0.5$  ppp would smooth the images with uniformly distributed intensities (e.g, band 32), and on the other hand, resulting in discontinuities in images where the intensities are concentrated only on some part of the image (e.g, band 1).

## 4.1 Discussion

We have seen a detailed comparison between the different methods that were proposed. It was discussed earlier that the three methods have equally good results with the JS-DWT method slightly outperforming the rest. It can be argued that the TVNN method does not perform very well because the data consists of only 32 images (number of wavelengths considered) that is not enough to promote low-rankness. TV regularization in TVNN method does its job by preserving the edges while promoting smoothness over the image. On the other hand, the nuclear norm is supposed to improve the reconstruction at all the wavelengths by adding information about spectral correlation. The nuclear norm fails to promote spectral correlation in this case because of the low number of wavelengths that are available. Moreover, the nuclear norm adds to the computational complexity by a great deal because of the SVD computation at each iteration.

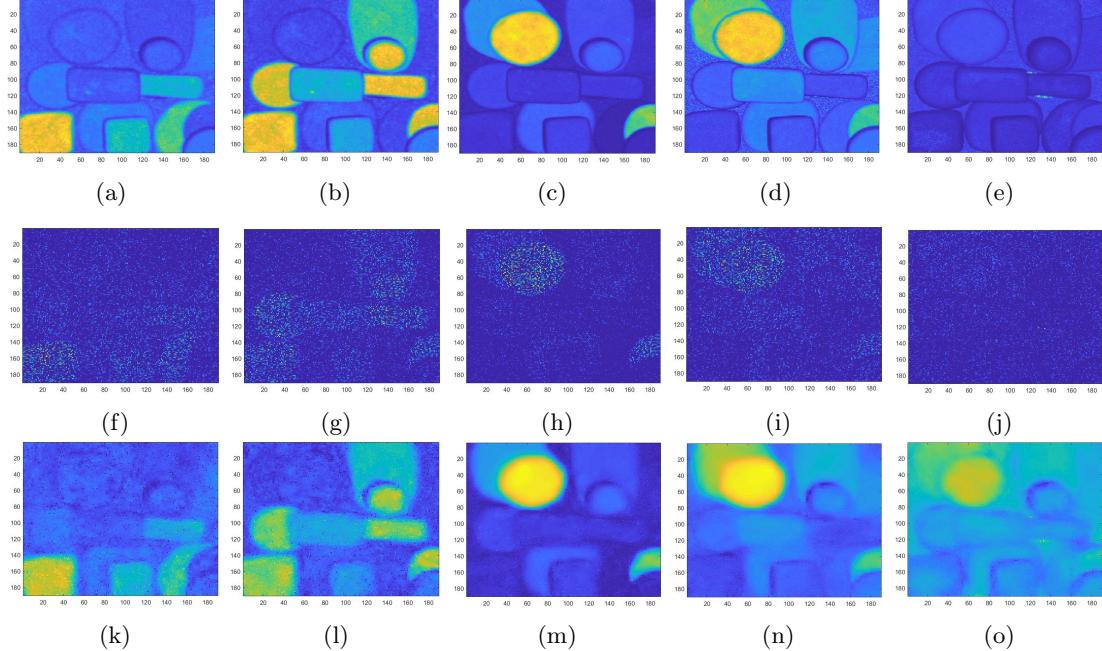


Figure 4.5: Response Intensity image reconstruction results using the JS-DWT method. The columns correspond to the 1<sup>st</sup>, 6<sup>th</sup>, 15<sup>th</sup>, 24<sup>th</sup> and 32<sup>nd</sup> wavelength bands respectively. Row 1 ((a)-(e)) shows the ground truth images, Row 2 ((f)-(j)) shows the measurements corresponding to an average of  $\sim 0.5$  photons per pixel per band collected only on 1/4<sup>th</sup> of the total number of pixel locations, and, Row 3 ((k)-(o)) shows the reconstructions using the JS-DWT method.

This can be a costly trade-off, given that the joint sparsity method performs better in much less time.

This drawback can be avoided by considering a group-TV norm instead of taking a TV norm of individual images at each wavelength. This group-TV can suppress spurious detail present in the image, and at the same time, it can suppress unwanted detail at different wavelengths by using the information present in the complete data cube. The joint sparsity method achieves the same function as that of the group-TV norm by taking an  $l_{2,1}$  norm on the data cube. This improvement can be clearly seen from Figure 4.5, i.e, reconstruction is equally good at all the considered wavelengths irrespective of the distribution of the intensities over the images. It can be noticed from that figure that the JS-DWT method is able to preserve edges in the images at all wavelengths through the use of the  $l_{2,1}$  norm. Although, JS-DCT method is very similar to the JS-DWT method, it was noticed that DCT as the sparsity basis is not as appropriate as DWT for the considered scene. DCT is not very efficient at preserving edges and it also adds extra smoothness to the images, thus, reducing contrast and detail in the images.

# Chapter 5

## Conclusion

### 5.1 Conclusion

The contributions of this work can be listed starting with the observation model. Firstly, an in-painting method has been successfully implemented for a Poisson noise model by minimizing the negative log-likelihood of the observed data. We also saw a detailed understanding of the relevance of Poisson statistics to our hyper-spectral image in-painting problem. The intensity image estimation problem was efficiently modelled by separating the estimation process into two sequential steps of estimating the baseline intensity and response intensity. Secondly, intensity images at multiple wavelengths were simultaneously reconstructed through optimization with guaranteed convergence using a variant of ADMM algorithm that can handle minimization problems consisting of multiple terms. We also saw the implementation of an approach to iteratively vary the step size of the optimization algorithm to achieve convergence faster.

One of the most important contributions of this work is in its applications to low-light imaging. The proposed minimization problems have shown to be able to model the data appropriately, specially, when the data has observations consisting of an average of  $<10$  photons per pixel. We have also seen good reconstructions (both quantitatively and visually) at extremely low observations (data with an average of  $\sim 0.5$  photons per pixel). Additionally, the methods have shown to perform extremely well while reconstructing intensity images at all the considered wavelengths, simultaneously.

Lastly, it was shown that the proposed observation model and the minimization problems support the compressed sensing framework, in that sense that, we have seen better reconstructions when photons were collected randomly at selected pixel locations than when photons were collected uniformly over all the image pixels. We also saw that intensity image reconstruction

is better when data was collected at only  $\sim 6\%$  of pixel locations, than when data was collected uniformly on all the pixel locations. Having listed the contributions, it can be claimed that this work answers the research questions posed at the beginning while justifying the title.

## 5.2 Future Work

A few directions for future work can be listed as follows. First of all, this work did not make any contributions towards the estimation of the depth profile of the scene. So, further research can be conducted in the direction of estimating the depth images of the hyper-spectral SPL data, while keeping in mind that estimating the depth at each pixel location poses a non-convex problem. Although, depth estimation is non-convex, the reconstructed response intensity images should facilitate depth reconstruction. As this work obtains good reconstructions of the response intensity images, future research can focus on material classification using the reconstructed response intensity images. By looking at the quality of the reconstructed response intensity images and the contrast between reflectivity of different objects present in the scene, material classification can be easily achieved even at low-light observations.

## Appendix A

# Convex Analysis

The convexity of a function can be proved as follows. See [54, 55] for more details. Consider a function  $f : \mathbb{X} \rightarrow \bar{\mathbb{R}}$ , where  $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$  is called the extended real line and  $\mathbb{X}$  is the Hilbert space. The domain of the function  $f$  is the set  $dom(f) = \{x : f(x) < +\infty\}$ . Then, the function is said to be convex if  $f(\alpha u + (1 - \alpha)v) \leq \alpha f(u) + (1 - \alpha)f(v)$ , for any  $u, v \in \mathbb{X}$  and any  $\alpha \in [0, 1]$ . Convexity is said to be strict if the inequality holds strictly ( $<$ ) for any  $u, v \in dom(f)$  and  $\alpha \in ]0, 1[$ .

## Appendix B

# Moreau Proximity Operator

Consider a function  $\Phi : \mathbb{X} \rightarrow \bar{\mathbb{R}}$ , where  $\mathbb{X}$  is the Hilbert space, assumed to be convex, lower semi-continuous and proper. Consider also the function  $d_y : \mathbb{X} \rightarrow \mathbb{R}$ , defined as,

$$d_y = \|x - y\|^2 \quad (\text{B.1})$$

where,  $\|.\|$  defines the norm. Clearly, for any  $y$ , the function  $d_y$  is convex, and so is the function  $f : d_y + \Phi$ . Since,  $d_y$  is strictly convex,  $f$  is strictly convex, thus, the minimizer of  $f$  is unique. From these facts, the so-called Moreau proximity operator (see [54, 56–58] for details)  $\Psi_\Phi : \mathbb{X} \rightarrow \mathbb{X}$  can be defined as,

$$\Psi_\Phi(y) = \arg \min_x \left\{ \frac{d_y^2(x)}{2} + \Phi(x) \right\}. \quad (\text{B.2})$$

# Bibliography

- [1] M. A. Figueiredo and J. M. Bioucas-Dias, “Restoration of poissonian images using alternating direction optimization,” *IEEE transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, 2010.
- [2] C. Weitkamp, *Lidar: range-resolved optical remote sensing of the atmosphere*. Springer Science & Business, 2006, vol. 102.
- [3] J. Goldsmith, F. H. Blair, S. E. Bisson, and D. D. Turner, “Turn-key raman lidar for profiling atmospheric water vapor, clouds, and aerosols,” *Applied Optics*, vol. 37, no. 21, pp. 4979–4990, 1998.
- [4] A. Ansmann, M. Riebesell, and C. Weitkamp, “Measurement of atmospheric aerosol extinction profiles with a raman lidar,” *Optics letters*, vol. 15, no. 13, pp. 746–748, 1990.
- [5] J. E. Means, S. A. Acker, B. J. Fitt, M. Renslow, L. Emerson, C. J. Hendrix *et al.*, “Predicting forest stand characteristics with airborne scanning lidar,” *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 11, pp. 1367–1372, 2000.
- [6] T. Inomata, D. Triadan, F. Pinzón, M. Burham, J. L. Ranchos, K. Aoyama, and T. Haraguchi, “Archaeological application of airborne lidar to examine social changes in the ceibal region of the maya lowlands,” *PloS one*, vol. 13, no. 2, p. e0191619, 2018.
- [7] G. Zhou, C. Song, J. Simmers, and P. Cheng, “Urban 3d gis from lidar and digital aerial images,” *Computers & Geosciences*, vol. 30, no. 4, pp. 345–353, 2004.
- [8] K. Zhang, J. Yan, and S.-C. Chen, “Automatic construction of building footprints from airborne lidar data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2523–2533, 2006.
- [9] A. Maccarone, A. McCarthy, X. Ren, R. E. Warburton, A. M. Wallace, J. Moffat, Y. Petillot, and G. S. Buller, “Underwater depth imaging using time-correlated single-photon counting,” *Optics express*, vol. 23, no. 26, pp. 33 911–33 926, 2015.

- [10] N. Cadalli, P. J. Shargo, D. C. Munson, and A. C. Singer, “Three-dimensional tomographic imaging of ocean mines from real and simulated lidar returns,” in *Ocean Optics: Remote Sensing and Underwater Imaging*, vol. 4488. International Society for Optics and Photonics, 2002, pp. 155–167.
- [11] M. A. Lefsky, W. B. Cohen, G. G. Parker, and D. J. Harding, “Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists,” *AIBS Bulletin*, vol. 52, no. 1, pp. 19–30, 2002.
- [12] M. A. Wulder, C. W. Bater, N. C. Coops, T. Hilker, and J. C. White, “The role of lidar in sustainable forest management,” *The Forestry Chronicle*, vol. 84, no. 6, pp. 807–826, 2008.
- [13] D. Martinez-Ramirez, G. Buller, A. McCarthy, X. Ren, A. W. S. Morak, C. Nichol, and I. Woodhouse, “Developing hyperspectral lidar for structural and biochemical analysis of forest data,” in *Proc. EARSEL Conf. Adv. Geosci.*, 2012, pp. 1–11.
- [14] J. De Bruijne, R. Kohley, and T. Prusti, “Gaia: 1,000 million stars with 100 ccd detectors,” in *Space Telescopes and Instrumentation 2010: Optical, Infrared, and Millimeter Wave*, vol. 7731. International Society for Optics and Photonics, 2010, p. 77311C.
- [15] G. Finger, R. J. Dorn, S. Eschbaumer, D. Ives, L. Mehrgan, M. Meyer, and J. Stegmeier, “Infrared detector systems at eso,” in *Workshop Detectors for Astronomy*, 2009.
- [16] K. J. Moore, S. Turconi, S. Ashman, M. Ruediger, U. Haupt, V. Emerick, and A. J. Pope, “Single molecule detection technologies in miniaturized high throughput screening: fluorescence correlation spectroscopy,” *Journal of biomolecular screening*, vol. 4, no. 6, pp. 335–353, 1999.
- [17] A. Esposito, T. Oggier, H. Gerritsen, F. Lustenberger, and F. Wouters, “All-solid-state lock-in imaging for wide-field fluorescence lifetime sensing,” *Optics express*, vol. 13, no. 24, pp. 9812–9821, 2005.
- [18] F. Müller and C. Fattinger, “Exploiting molecular biology by time-resolved fluorescence imaging,” in *Single-Photon Imaging*. Springer, 2011, pp. 329–344.
- [19] P. Seitz and A. J. Theuwissen, *Single-photon imaging*. Springer Science & Business Media, 2011, vol. 160.

- [20] G. Buller and A. Wallace, “Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition,” *IEEE Journal of selected topics in quantum electronics*, vol. 13, no. 4, pp. 1006–1015, 2007.
- [21] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, “Single-pixel imaging via compressive sampling,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [22] G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell, “Photon counting compressive depth mapping,” *Optics express*, vol. 21, no. 20, pp. 23 822–23 837, 2013.
- [23] G. S. Buller, R. D. Harkins, A. McCarthy, P. A. Hiskett, G. R. MacKinnon, G. R. Smith, R. Sung, A. M. Wallace, R. A. Lamb, K. D. Ridley *et al.*, “Multiple wavelength time-of-flight sensor based on time-correlated single-photon counting,” *Review of Scientific Instruments*, vol. 76, no. 8, p. 083112, 2005.
- [24] M. Umasuthan, A. Wallace, J. Massa, G. Buller, and A. Walker, “Processing time-correlated single photon counting data to acquire range images,” *IEE Proceedings-Vision, Image and Signal Processing*, vol. 145, no. 4, pp. 237–243, 1998.
- [25] W. Becker, *Advanced time-correlated single photon counting techniques*. Springer Science & Business Media, 2005, vol. 81.
- [26] T. Yamane, “Statistics: An introductory analysis,” 1973.
- [27] D. Shin, J. H. Shapiro, and V. K. Goyal, “Single-photon depth imaging using a union-of-subspaces model,” *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2254–2258, 2015.
- [28] J. H. Shin, Dongeek & Shapiro and V. K. Goyal, “Computational single-photon depth imaging without transverse regularization,” in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 973–977.
- [29] Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. Buller, S. McLaughlin, and A. Hero, “Robust spectral unmixing of sparse multispectral lidar waveforms using gamma markov random fields,” *IEEE Transactions on Computational Imaging*, 2017.
- [30] Y. Altmann, A. Maccarone, A. Halimi, A. McCarthy, G. Buller, and S. McLaughlin, “Efficient range estimation and material quantification from multispectral lidar waveforms,” in *Sensor Signal Processing for Defence (SSPD), 2016*. IEEE, 2016, pp. 1–5.
- [31] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, “Computational 3d and reflectivity imaging with high photon efficiency,” in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 46–50.

- [32] A. Shin, Dongeek & Kirmani, V. K. Goyal, and J. H. Shapiro, “Photon-efficient computational 3-d and reflectivity imaging with single-photon detectors,” *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 112–125, 2015.
- [33] Z. T. Harmany, R. F. Marcia, and R. M. Willett, “This is spiral-tap: Sparse poisson intensity reconstruction algorithmstheory and practice,” *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.
- [34] P. Sarder and A. Nehorai, “Deconvolution methods for 3-d fluorescence microscopy images,” *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 32–45, 2006.
- [35] J.-L. Starck and F. Murtagh, *Astronomical image and data analysis*. Springer Science & Business Media, 2007.
- [36] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, “Richardson–lucy algorithm with total variation regularization for 3d confocal microscope deconvolution,” *Microscopy research and technique*, vol. 69, no. 4, pp. 260–266, 2006.
- [37] J.-L. Starck, F. Murtagii, and A. Bijaoui, “Multiresolution support applied to image filtering and restoration,” *Graphical models and image processing*, vol. 57, no. 5, pp. 420–431, 1995.
- [38] R. Nowak and E. D. Kolaczyk, “A bayesian multiscale framework for poisson inverse problems,” in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 3. IEEE, 1999, pp. 1741–1744.
- [39] K. E. Timmermann and R. D. Nowak, “Multiscale modeling and estimation of poisson processes with application to photon-limited imaging,” *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, 1999.
- [40] R. M. Willett and R. D. Nowak, “Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 332–350, 2003.
- [41] R. D. Willett, Rebecca M & Nowak, “Fast multiresolution photon-limited image reconstruction,” in *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*. IEEE, 2004, pp. 1192–1195.
- [42] S. Lefkimiatis, P. Maragos, and G. Papandreou, “Bayesian inference on multiscale models for poisson intensity estimation: Applications to photon-limited image denoising,” *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1724–1741, 2009.

- [43] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing mri,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [44] G. Puy, J. P. Marques, R. Gruetter, J.-P. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux, “Spread spectrum magnetic resonance imaging,” *IEEE transactions on medical imaging*, vol. 31, no. 3, pp. 586–598, 2012.
- [45] A. Auria, A. Daducci, J.-P. Thiran, and Y. Wiaux, “Structured sparsity for spatially coherent fibre orientation estimation in diffusion mri,” *NeuroImage*, vol. 115, pp. 245–255, 2015.
- [46] A. Abdulaziz, A. Dabbech, A. Onose, and Y. Wiaux, “A low-rank and joint-sparsity model for hyper-spectral radio-interferometric imaging,” in *Signal Processing Conference (EUSIPCO), 2016 24th European*. IEEE, 2016, pp. 388–392.
- [47] R. E. Carrillo, J. McEwen, and Y. Wiaux, “Sparsity averaging reweighted analysis (sara): a novel algorithm for radio-interferometric imaging,” *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.
- [48] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [49] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.
- [50] A. Chambolle, “An algorithm for total variation minimization and applications,” *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.
- [51] A. Beck and M. Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.
- [52] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, “An introduction to total variation for image analysis,” *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, no. 263-340, p. 227, 2010.
- [53] M. Golbabae and P. Vandergheynst, “Compressed sensing of simultaneous low-rank and joint-sparse matrices,” *arXiv preprint arXiv:1211.5058*, 2012.
- [54] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

- [55] C. Zalinescu, *Convex analysis in general vector spaces.* World scientific, 2002.
- [56] P. L. Combettes and J.-C. Pesquet, “Proximal thresholding algorithm for minimization over orthonormal bases,” *SIAM Journal on Optimization*, vol. 18, no. 4, pp. 1351–1376, 2007.
- [57] J.-C. Combettes, Patrick L & Pesquet, “A douglas–rachford splitting approach to non-smooth convex variational signal recovery,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, 2007.
- [58] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, “A variational formulation for frame-based inverse problems,” *Inverse Problems*, vol. 23, no. 4, p. 1495, 2007.