Dissertations, Master's Theses and Master's Reports

2021

# Novel Methods in Computational Imaging with Applications in Remote Sensing

Adam Webb

*Michigan Technological University*, ajwebb@mtu.edu

NOVEL METHODS IN COMPUTATIONAL IMAGING WITH APPLICATIONS

IN REMOTE SENSING

By

Adam J. Webb

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Electrical Engineering

MICHIGAN TECHNOLOGICAL UNIVERSITY

2021

This dissertation has been approved in partial fulfillment of the requirements for the

Degree of DOCTOR OF PHILOSOPHY in Electrical Engineering.

**Department of** Electrical and Computer Engineering

Dissertation Co-advisor: *Dr. Timothy J. Schulz*

Dissertation Co-advisor: *Dr. Timothy C. Havens*

Committee Member: *Dr. Michael C. Roggemann*

Committee Member: *Dr. Thomas Oommen*

Department Chair: *Dr. Glen Archer*

# Dedication

To my father, who among many other things taught me the value of hard work.

# Contents

# Abstract

This dissertation is devoted to novel computational imaging methods with applications in remote sensing. Computational imaging methods are applied to three distinct applications including imaging and detection of buried explosive hazards utilizing array radar, high resolution imaging of satellites in geosynchronous orbit utilizing optical hypertelescope arrays, and characterization of atmospheric turbulence through multi-frame blind deconvolution utilizing conventional optical digital sensors.

The first application considered utilizes a radar array employed as a forward looking ground penetrating radar system with applications in explosive hazard detection. A penalized least squares technique with sparsity-inducing regularization is applied to produce imagery, which is consistent with the expectation that objects are sparsely populated but extended with respect to the pixel grid. Additionally, a series of pre-processing steps is demonstrated which result in a greatly reduced data size and computational cost. Demonstrations of the approach are provided using experimental data and results are given in terms of signal to background ratio, image resolution, and relative computation time. The second application involves a sparse-aperture telescope array configured as a hypertelescope with applications in long range imaging. The penalized least squares technique with sparsity-inducing regularization is adapted and applied to this very different imaging modality. A comprehensive study

of the algorithm tuning parameters is performed and performance is characterized using the Structure Similarity Metric (SSIM) to maximize image quality. Simulated measurements are used to show that imaging performance achieved using the proposed algorithm compares favorably in comparison to conventional Richardson-Lucy deconvolution. The third application involves a multi-frame collection from a conventional digital sensor with the primary objective of characterizing the atmospheric turbulence in the medium of propagation. In this application a joint estimate of the image is obtained along with the Zernike coefficients associated with the atmospheric PSF at each frame, and the Fried parameter $r_0$ of the atmosphere. A pair of constraints are applied to a penalized least squares objective function to enforce the theoretical statistics of the set of PSF estimates as a function of $r_0$.

Results of the approach are shown with both simulated and experimental data and demonstrate excellent agreement between the estimated $r_0$ values and the known or measured $r_0$ values respectively.

# Chapter 1

# Introduction

## 1.1   Motivation

This dissertation addresses the challenge of producing high fidelity image representations in real world sensing applications where the physical realities of the collection limit the performance achieved through conventional means. High fidelity physics based models of relevant sensing paradigms coupled with penalized and constrained least squares statistical estimation methods are utilized to overcome the challenges present in these real world sensing applications. A least squares based imaging approach is adapted and applied to three distinct applications including imaging and

detection of buried explosive hazards utilizing array radar; high resolution imaging of satellites in geosynchronous orbit utilizing optical hypertelescope arrays; and characterization of atmospheric turbulence through multi-frame blind deconvolution utilizing conventional optical digital sensors. Very similar approaches are applied to each application even though different models are required to accurately account for the physics associated with the very different collection modalities.

In the first application described in chapter 2, ground penetrating radar (GPR) is utilized to create images with the purpose of detecting buried explosive ordinances. Imaging challenges associated with this application include; limited resolution due to finite bandwidth and the finite aperture size of the sensor, speckle due to fluctuating reflectivity over space and time, and side-lobe corruption present as an artifact of conventional image formation methods. Furthermore, in the forward-looking collection regime the collection geometry is particularly unfavorable for buried targets due to the limited signal energy that penetrates the earth compared to the signal energy that is scattered from the surface.

In the second application described in chapter 3 a hypertelescope imaging system is utilized to image satellites in geosynchronous orbit from earth. A hypertelescope refers to a connected array of telescopes whose collected energy is combined coherently in a central location with the purpose of forming an image whose resolution reflects the size of the array rather than the size of the individual apertures. The primary

challenges associated with this application include the extreme sparsity of the array and the low signal levels inherent with imaging objects in geosynchronous orbit. The resulting direct images exhibit low SNR and poor PSF characteristics.

In the third application described in chapter 4, multi-frame image collections using a conventional digital camera in the visible band are utilized to characterize the levels of turbulence present in the atmosphere between the sensor and a given target object. The challenge inherent in this application is that the object as well as the atmospheric turbulence and it's effects on the PSF in each frame are unknown and must be estimated jointly.

## 1.2 Contributions

Published works which were generated as part of this dissertation are listed as follows.

† A. J. Webb, M. C. Roggemann, and M. R. Whiteley, "Atmospheric turbulence characterization through multi-frame blind deconvolution," *Appl. Opt.*, Under Review 2021

† A. J. Webb and M. C. Roggemann, "A penalized least squares technique for imaging with hypertelescopes," *Appl. Opt.*, Under Review 2021

† A. J. Webb, T. C. Havens, and T. J. Schulz, "Fast image reconstruction in forward looking gpr using dual l1 regularization," *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 470–478, 2018

† A. J. Webb, T. C. Havens, and T. J. Schulz, "An apodization approach for processing forward-looking gpr for buried target detection," in *Proc. SPIE*, vol. 9454, pp. 94540X–94540X–15, 2015

† A. J. Webb, T. C. Havens, and T. J. Schulz, "Spectral diversity for ground clutter mitigation in forward-looking gpr," in *Proc. SPIE*, vol. 9823, pp. 98231M–98231M–18, 2016

† A. J. Webb, T. C. Havens, and T. J. Schulz, "Iterative image formation for forward looking gpr," in *Military Sensing Symposium*, vol. 1, March 2016

† A. J. Webb, T. C. Havens, and T. J. Schulz, "Gpr imaging with mutual intensity," in *Proc. SPIE*, vol. 10182, pp. 101821B–101821B–8, 2017

† J. S. Rice, A. Pinar, T. C. Havens, A. J. Webb, and T. J. Schulz, "Multiple instance learning for buried hazard detection," in *Proc. SPIE*, vol. 9823, 2016

† A. J. Pinar, T. C. Havens, and A. Webb, "Multisensor fusion of flgpr and thermal and visible-spectrum cameras for standoff detection of buried objects," in *Proc. SPIE*, vol. 10182, pp. 101821A–101821A–15, 2017

† J. Burns, M. P. Masarik, I. J. Xique, B. Thelen, and A. Webb, "Comparative analysis of image formation techniques for flgpr," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII*, vol. 10182, p. 1018219, International Society for Optics and Photonics, 2017

# Chapter 2

# Fast Image Reconstruction in Forward-Looking GPR Using Dual $\ell_1$ Regularization

## 2.1 Introduction

An important goal for the U.S. Army is remediating the threats of buried targets as these devices cause uncountable deaths and injuries to both civilians and soldiers throughout the world. Modern hazard detection is accomplished using multiple modality sensors whose data are combined in an effort to detect and classify potential

threats in the presence of noise and clutter. Common modalities utilized for threat detection have included acoustics, infrared (IR) and visible-spectrum cameras, and ground-penetrating-radar (GPR) [11, 12, 13, 14, 15, 16].

Radar is a key modality whose strength, in this context, lies in it's ability to penetrate into the ground enabling the detection of buried objects. Operational constraints, expected target types, and environmental considerations dictate the manner in which GPR is used for hazard detection. Close proximity *downward-looking* GPR systems are optimal for buried targets, but in many cases *forward-looking* systems are preferential to maintain some stand-off distance from possible threats; even airborne GPR, and *foliage penetrating* (FOPEN) radar sensors have been utilized for large scale mine-clearing efforts [17].

Improving the performance of *forward-looking* GPR (FLGPR) systems has become an active research topic due to its potential to reliably detect hazards before they are encountered [18, 19, 20, 21, 22, 23, 24]. Typical standoff distances can range from few to tens of meters. FLGPR has been applied to the detection of side-attack, surface, and buried devices [25, 26, 27]. Due to system design considerations, modern FLGPR sensors typically operate in step-frequency collection mode and conventional image formation processing is accomplished using the back-projection imaging technique [28]. Imaging limitations include; limited resolution due to finite bandwidth and the finite aperture size of the sensor, speckle due to fluctuating reflectivity over

space and time, and side-lobe corruption present as an artifact of back-projection processing. Speckle has been exploited as means of differentiating targets from clutter [29] [30]. Side-lobes are controlled to a certain extent through simple integration using returns collected from multiple platform positions. Further suppression of side-lobes has been demonstrated using advanced apodization based techniques [5, 31]. Studies comparing tomographic reconstruction utilizing the Truncated Singular Value Decomposition (TSVD) method to other imaging techniques including backprojection have shown that TSVD reconstruction performs favorably in terms of both target resolution and detection performance [32, 33, 34, 35]. FLGPR imaging through compressive sensing has demonstrated reduced system complexity and improved signal to clutter performance [36]. Other sparsity inducing techniques based on $\ell_1$ regularization have shown promising performance with respect to side-lobe mitigation, clutter mitigation, and finite resolution [7, 36, 37, 38, 39, 40]. One such approach, originally introduced by Cetin et al. in the context of synthetic aperture radar (SAR), simultaneously balances the $\ell_1$ regularization of the total energy (TE) and the total variation (TV) of the magnitude of the image estimate which has the ability to induce sparsity while maintaining the extended nature of targets [41]. Another approach, introduced by Wang et al. in the context of 3-D Multistatic imaging introduces the idea of performing reconstruction based on backprojected image data while using mixed norm regularization [42].

The approach taken here is an adaptation of the methods taken by Cetin and Wang

to suit the step frequency FLGPR collection scenario. The approach is extended by introducing a Doppler signature pre-filter to control side-lobe energy resulting from bright out-of-scene objects and demonstrate that by reconstructing based on back-projected image data we enable pre-averaging as a method of further reducing computational burden. These additions result in a reduction in background energy and improved computational performance.

The remainder of this chapter is organized as follows. In Section 2.2, we develop the collected signal model relevant to the step-frequency FLGPR sensor; in Section 2.3 we describe the proposed Doppler filter applied to mitigate side-lobe energy originating from out-of-scene targets; in Section 2.4 we present the proposed fast image reconstruction approach; and in Section 2.5 we present a series of results using experimentally collected data.

## 2.2 Signal Model

Consider a vehicle mounted FLGPR multi-static array continuously collecting data at discrete frequency samples as the sensing platform approaches a particular *region of interest* (ROI). A single complex data sample is collected at each discrete frequency and for each bi-static transmitter/receiver pair. In our nomenclature, a single data frame is defined as a complete set of all discrete frequency samples collected for each

and every bi-static transmitter/receiver pair. It is assumed that many data frames are collected for each ROI. The signal history indexed by the transmitter/receiver pair $p$ and given as a function of frequency $f$ is given by

$$s_p[f] = \iiint g(\mathbf{x}) \frac{e^{-j2\pi\frac{f}{c}\left(R\left(\mathbf{x},\mathbf{x}_p^{(tx)}\right)+R\left(\mathbf{x},\mathbf{x}_p^{(rx)}\right)\right)}}{R\left(\mathbf{x},\mathbf{x}_p^{(tx)}\right)R\left(\mathbf{x},\mathbf{x}_p^{(rx)}\right)}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z, \qquad (2.1)$$

where $\mathbf{x}_p^{(tx)}$ and $\mathbf{x}_p^{(rx)}$ are the position vectors of the $p$th pair of transmitter and receiver elements, respectively. The three dimensional complex reflectivity density function is given by $g(\mathbf{x})$, where $\mathbf{x}$ is a position vector in 3-space. The functions $R\left(\mathbf{x},\mathbf{x}_p^{(tx)}\right)$ and $R\left(\mathbf{x},\mathbf{x}_p^{(rx)}\right)$ represent the ranges from the $p$th transmit and receive element pair to a point in 3-space. Given the finite resolution of the collection system it is both convenient and appropriate to approximate the continuous reflectivity density function by using a discretized collection of scatterers. In applying this approximation, the triple integral can be replaced by a single summation, and the continuous reflectivity density function $g(\mathbf{x})$ is replaced by a discrete reflectivity density function $g[\mathbf{x}_k]$. The approximate signal model is then given by

$$s_p[f] = \sum_{k=1}^{K} g[\mathbf{x}_k] \frac{e^{-j2\pi\frac{f}{c}\left(R\left(\mathbf{x},\mathbf{x}_p^{(tx)}\right)+R\left(\mathbf{x},\mathbf{x}_p^{(rx)}\right)\right)}}{R\left(\mathbf{x},\mathbf{x}_p^{(tx)}\right)R\left(\mathbf{x},\mathbf{x}_p^{(rx)}\right)}, \qquad (2.2)$$

where $K$ indicates the total number of discretized scatterers in the region of interest and can be succinctly expressed in the form of the matrix-vector product,

$$\mathbf{s}^{(p)} = \mathbf{\Phi}^{(p)}\mathbf{g}. \tag{2.3}$$

In this representation the signal for a single transmit/receive pair is encoded into the $L{\times}1$ vector $\mathbf{s}^{(p)}$, the discrete reflectivity function is encoded into the $K{\times}1$ vector $\mathbf{g}^{(p)}$, and they are related through a linear transformation represented by the $L{\times}K$ matrix $\mathbf{\Phi}^{(p)}$.

The matrix-vector formulation is extended trivially to produce an expression characterizing the collected data for an entire frame, or even for a set of frames. This is accomplished through concatenation of both the forward operator and the collected signal associated with each of the P transmit/receive pairs and then again for the entire set of transmit/receive pairs for each frame. The resulting expression is given as

$$\begin{pmatrix} \mathbf{s}^{(1,1)} \\ \mathbf{s}^{(1,2)} \\ \vdots \\ \mathbf{s}^{(1,P)} \\ \mathbf{s}^{(2,1)} \\ \vdots \\ \mathbf{s}^{(N,P)} \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}^{(1,1)} \\ \mathbf{\Phi}^{(1,2)} \\ \vdots \\ \mathbf{\Phi}^{(1,P)} \\ \mathbf{\Phi}^{(2,1)} \\ \vdots \\ \mathbf{\Phi}^{(N,P)} \end{pmatrix} \mathbf{g}, \tag{2.4}$$

10

which we express succinctly as

$$\mathbf{s} = \mathbf{\Phi g}, \tag{2.5}$$

inferring the number of frames through context. In summary, the relevant parameters characterizing the size and organization of the collected data include the following:

$L$  The number of frequency samples;

$P$  The number of bistatic transmitter/receiver pairs;

$N$  The number of data frames; and

$K$  The number of object samples.

## 2.3  Doppler Pre-Processing

A common occurrence encountered in FLGPR imaging is the existence of strong clutter returns from objects including trees, shrubs, and rocks originating from outside of the ROI. Side-lobes associated with these clutter returns can and often do corrupt the relevant portion of the formed image. This effect is amplified at the upper end of the frequency spectrum where sampling of the MIMO array is sparse relative to the shorter wavelength. These side-lobes are eliminated during reconstruction but only if the sampled grid extends out to account for the true location of these objects. Large sample grids are highly undesirable due to the increased computational burden they
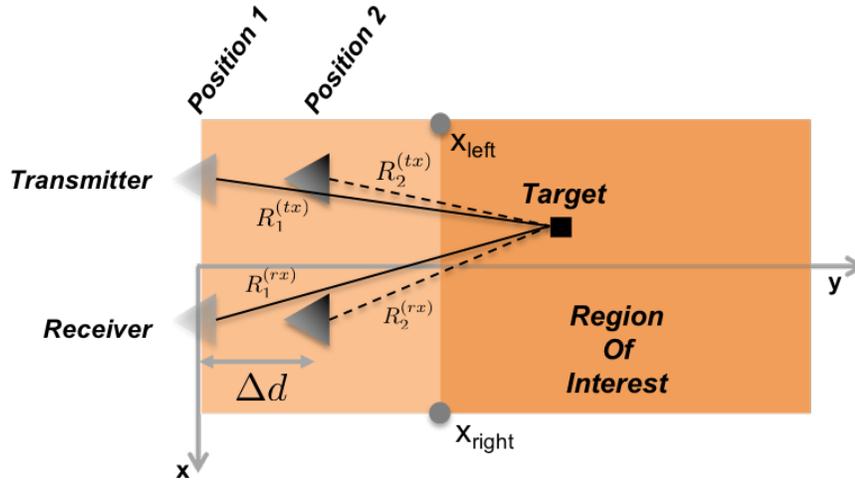
**Figure 2.1:** The change in the bistatic collection geometry as the platform approaches a ROI produces a spatially dependent Doppler shift in the collected signal.

place on the processor. Fortunately this effect can be mitigated, at least partially, by applying a filter based on the expected Doppler response of objects in the ROI. As the platform approaches the ROI, the range to relevant objects decreases producing a Doppler-shifted response. A filter can be constructed, and applied to *each* bistatic transmitter/receiver pair, to eliminate data in regions of Doppler space which are not consistent of objects in the ROI. This is similar in nature to the way stationary clutter is often filtered in pulse-Doppler radar [43].

At each platform position the phase delay associated with the bistatic propagation is given by, $\phi = kR$, where $k$ is the spatial frequency in units of m$^{-1}$ and $R = R^{(tx)} + R^{(rx)}$ is the 2-way bistatic range, as illustrated in Figure 2.1 for consecutive platform positions. At consecutive platform positions, the change in phase associated

with a particular scatterer is given by

$$\Delta\phi = k\left(R_2 - R_1\right), \tag{2.6}$$

where $R_1$ and $R_2$ represent the 2-way bistatic ranges at position 1 and position 2 of the transmitter/receiver pair. The rate of change in the observed phase over multiple platform positions is characterized by the *spatial Doppler shift* and is given, again in units of m$^{-1}$ by

$$k_D = \frac{\Delta\phi}{\Delta d} = \frac{k}{\Delta d}\left(R_2 - R_1\right), \tag{2.7}$$

where $\Delta d$ is the distance the platform travels between consecutive frames. It is noteworthy that in this context Doppler shifts are completely independent of the actual velocity of the platform.

It is our objective to construct a simple pass-band filter based on the expected spatial Doppler shift associated with in-lane targets. For the platform illustrated in Figure 2.1 traveling in the $+y$ direction, the maximum spatial Doppler shift associated with *any* in-lane target is found by considering the limiting case of a target located at

13

$x = 0$, $y = \infty$, and is given by

$$k_{D_{\max}} = \lim_{y \to \infty} \frac{k}{\Delta d} (R_2 - R_1) \tag{2.8}$$
$$= \frac{k}{\Delta d} \lim_{y \to \infty} \left( R_2^{(tx)} + R_2^{(rx)} - R_1^{(tx)} - R_1^{(rx)} \right)$$
$$= \frac{k}{\Delta d} (2\Delta d)$$
$$= 2k.$$

The minimum spatial Doppler shift associated with in-lane targets is found by considering the boundaries of the image region at the nearest range of interest. The two relevant points are shown in figure 2.1 as $\mathbf{x}_{left}$ and $\mathbf{x}_{right}$ and the bistatic range associated with these points is given by, $R_{left}$ and $R_{right}$. The minimum spatial Doppler shift observed from in-lane targets is the minimum of the Doppler shift at these two points and is given by

$$k_{D_{\min}} = \frac{k}{\Delta d} \min\{R_{left}, R_{right}\}. \tag{2.9}$$

The single transmitter/receiver pair data is collected as a function of the lane position, $y$, and the collection frequency, $f$ and is expressed as $s(y, f)$. Through Fourier transform with respect to the lane position, the collected data is expressed in terms of spatial Doppler shift and collection frequency as $S(k_D, f)$. If we let the center spatial Doppler be given as $k_{D_c} = (k_{D_{\max}} + k_{D_{\min}})/2$ and the Doppler bandwidth be

14

given as $k_{D_{bw}} = k_{D_{\max}} - k_{D_{\min}}$ then the brick-wall bandpass filter given by

$$H(k_D, f) = \text{rect}\left[\frac{k_D - k_{D_c}}{k_{D_{bw}}}, f\right], \tag{2.10}$$

passes only returns consistent with in-lane targets. The filter is implemented through

$$S_f(k_D, f) = S(k_D, f) H(k_D, f), \tag{2.11}$$

where $S_f(k_D, f)$ is the filtered result.

The support bandwidth associated with the spatial Doppler shift is dependent on the platform spacing between consecutive frames and is given by $k_{bw} = 1/\Delta d$. Unless $\Delta d$ is very small, it is likely that aliasing will play a role in construction of the spatial Doppler filter. The aliased spatial Doppler frequency can be computed from the true Doppler frequency and the support bandwidth and is given by

$$k_{D_A} = \mod\left(k_D - \frac{k_{bw}}{2}, k_{bw}\right) - \frac{k_{bw}}{2}. \tag{2.12}$$

When aliasing occurs, the filter generated in equation (2.10) is simply mapped onto the supported spatial Doppler axis.

## 2.4 Image Formation

Given the signal model described in Section 2.2, the collected data *for a single frame* is modeled as

$$\mathbf{s} = \mathbf{\Phi}\mathbf{g} + \boldsymbol{\eta},$$ (2.13)

where $\mathbf{g}$ is the true object, $\mathbf{\Phi}$ is the forward model, $\mathbf{s}$ is the collected data in the *frequency domain*, and $\boldsymbol{\eta}$ is assumed to be a zero-mean *additive white Gaussian* (AWG) noise process.

Image reconstruction is accomplished through minimization of the objective function

$$C\left(\mathbf{g}|\mathbf{s}\right) = \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{s}_n - \mathbf{\Phi}_n \mathbf{g}\|_2^2 + \beta_E \|\|\mathbf{g}\|\|_1^1 + \beta_V \|\|\nabla \|\mathbf{g}\|\|\|_1^1.$$ (2.14)

The first term of the objective function represents the data fit term measuring the *mean squared error* (MSE) between the collected data and the theoretical signal associated with the object $\mathbf{g}$. The second term is a *total energy* (TE) regularization term penalizing the $\ell_1$ norm of the magnitude of the object. The third term is a *total variation* (TV) regularization penalizing the $\ell_1$ norm of the numerical gradient of the magnitude of the object. The dual regularization approach is consistent with the prior belief that objects are sparsely populated but extended with respect to the sampled grid density. The $\ell_1$ total energy regularization promotes sparsity in the

object estimate but has no preference as to the spatial distribution of energy. When the total energy regularization is applied alone, estimates tend to be composed of a collection of disjoint impulses. The inclusion of the $\ell_1$ TV regularization encourages smooth reconstructions while maintaining a sparse collection of objects. The values associated with the regularization parameters determine the characteristics of the resulting image reconstruction and must be chosen to balance data fit against the sparsity and smoothness of the solution. Increasing the TE regularization parameter increases sparsity while increasing the TV regularization increases smoothness by creating flat regions in the solution.

No closed form solution to the optimization problem exists due to the non-linearity of the $\ell_1$ norms, thus an iterative solution via convex optimization is required. In practice, the objective function given in equation (2.14) is costly to minimize due to the fact that the matrix-vector product, $\boldsymbol{\Phi}_n\mathbf{g}$, must be computed at every iteration, and the size of the matrix, $\boldsymbol{\Phi}_n$ is large ($LP{\times}K$). We therefore aim to pre-process the data in such a way as to minimize the size of this operation.

We recognize that for our frequency domain collection each frequency sample contains information pertaining to the entire scene which includes everywhere in the beam of the sensor array. By applying the DFT operator $\mathbf{W}$ to transform the data into the time domain, "range-gating" can be applied to eliminate data relevant to ranges not associated with the ROI. This idea can be extended to the cross-range dimension by
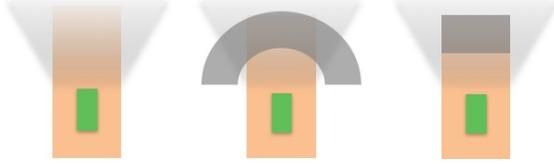
**Figure 2.2:** By pre-processing the data, the total sampled area can be reduced through range-gating (center) or imaging (right).

applying a deterministic image formation algorithm to the data. Using this approach the data can be spatially filtered, as illustrated in Figure 2.2, in an attempt to strictly retain information relevant to the ROI. The obvious imaging operation is the conventional back-projection imaging operation, which is applied though multiplication of the transpose of the forward model $\mathbf{\Phi}^T$. It should be noted that some loss of information should be expected after application of the imaging operator, but in many cases this is a trade that we will be willing to make in order to yield a reduced problem size. After application of the backprojection pre-processor to the collected data, and also including the backprojection operator as part of the forward model, the modified objective function is given as

$$C = \frac{1}{N} \sum_{n=1}^{N} \left\| \mathbf{\Phi}^T \mathbf{s} - \left( \mathbf{\Phi}^T \mathbf{\Phi} \right)_{K \times K} \mathbf{g} \right\|_2^2 + \beta_E \left\| |\mathbf{g}| \right\|_1^1 + \beta_V \left\| |\nabla |\mathbf{g}|| \right\|_1^1 . \qquad (2.15)$$

By pre-computing the net forward operator we have replaced the matrix-vector product from one including a $LP \times K$ matrix to one with a $K \times K$ matrix. More importantly, we have drastically reduced the size of $K$ itself by reducing the spatial dimensions of the necessary image grid.

Image reconstruction is typically performed on a finely sampled grid with respect to the conventional image resolution. In contrast, by the Nyquist sampling theorem, back-projection imaging can be performed on a critically sampled grid without any loss of information. As such, we now introduce a second back-projection operator $\mathbf{\Psi}^T$ of size $K_c{\times}LP$ (where $K_c < K$) which produces a near-critically sampled image. A further reduction in data size can thus be achieved by replacing the upsampled backprojection operator, with it's near-critically sampled counterpart as is given by

$$C = \frac{1}{N}\sum_{n=1}^{N}\left\|\mathbf{\Psi}^T\mathbf{s} - \left(\mathbf{\Psi}^T\mathbf{\Phi}\right)_{K_c\times K}\mathbf{g}\right\|_2^2 + \beta_E\left\|\left\|\mathbf{g}\right\|\right\|_1^1 + \beta_V\left\|\left\|\nabla\left\|\mathbf{g}\right\|\right\|\right\|_1^1. \qquad (2.16)$$

Pre-processing the data into the "image" domain enables the potential to average multiple consecutive frames prior to reconstruction. Pre-averaging of the data may be desirable to achieve further reduction in computational cost. If the number of frames is large, it makes sense to apply a pre-averaging operation combining sub-sets of consecutive images. In matrix notation the pre-averaging operation is a weighted horizontal concatenation of identity matrices, and is given as

$$\mathbf{S}_{K_c\times MK_c} = \frac{1}{M}\left(\mathbf{I}_{K_c\times K_c} \qquad \ldots \qquad \mathbf{I}_{K_c\times K_c}\right), \qquad (2.17)$$

where $M$ is the number of frames to be averaged. When the pre-averaging operation is applied we utilize the multi-frame representation of the signal model as described

19

in equation (2.4), such that $\boldsymbol{\Phi}$ is of size $MPL{\times}K$ and $\boldsymbol{\Psi}$ is of size $MPL{\times}K_c$. The aggregate forward operator remains of size $K_c{\times}K$ and the new objective function is given by

$$C = \frac{1}{N}\sum_{n=1}^{N}\left\|\mathbf{S}\boldsymbol{\Psi}^T\mathbf{s} - \left(\mathbf{S}\boldsymbol{\Psi}^T\boldsymbol{\Phi}\right)_{K_c{\times}K}\mathbf{g}\right\|_2^2 + \beta_E\left\|\left|\mathbf{g}\right|\right\|_1^1 + \beta_V\left\|\left|\nabla\left|\mathbf{g}\right|\right|\right\|_1^1, \qquad (2.18)$$

where $M$ is the number of frames to include in the pre-averaging operation and $N$ becomes the number of sets of pre-averaged frames to process.

## 2.5   Results

The proposed imaging algorithm was demonstrated using experimental data provided by the U.S. Army Night Vision and Electronic Sensors Directorate (NVESD). Since 2005, NVESD has been integrating sensing technologies to aid the soldier in detecting buried targets while maintaining Situational Awareness (SA) in route clearance and convoy escort vehicles. The experimental sensing platform used to collect the data provided for this work is shown in Figure 2.3. The relevant sensor is an L-band MIMO array consisting of 384 multi-static transmitter/receiver pairs. The step-frequency collection system records 2701 complex frequency samples over a bandwidth of approximately 2.8 GHz. Data collected between 400 MHz and 2.5 GHz was utilized for in this work. Data for this work was collected by the MIMO array as it slowly travels

20

**Figure 2.3:** Since 2005, the U.S. Army Night Vision and Electronic Sensors Directorate (NVESD) has been integrating sensing technologies to aid the soldier in detecting buried targets while maintaining Situational Awareness (SA) in route clearance and convoy escort vehicles.

down prepared experimental lanes. Lanes are composed of sand and gravel and a wide variety of targets are buried at shallow depths below the surface (on the order to a few cm's). While full polarimetric data is collected, only *HH* polarized data is utilized in this work.

In this section we first demonstrate the Doppler pre-filtering process and compare images with and without Doppler pre-filtering applied to demonstrate removal of side-lobe energy originating from objects located outside the intended ROI. We then, utilizing data with the Doppler pre-filter applied, perform a series of image reconstructions (without pre-averaging) and compare results to conventional back-projected image formation applied to two sample targets. Finally, we compare computation time and imaging performance with different levels of pre-averaging applied.

### 2.5.1 Doppler Pre-Filtering

A demonstration of Doppler filter pre-processing is illustrated using a particular ROI selected from the NVESD data set. The Doppler filter is constructed based on a lane width of 4 m, and an approximate frame-rate of 7 frames/m determining the support of the spatial Doppler shift. The resulting filter for a particular transmitter/receiver pair is shown in Figure 2.4a. The input data relevant to the Doppler filter shown in Figure 2.4a, is found in Figure 2.4b, and shows the original data displayed as a function of collection frequency and platform distance traveled. The collection frequency/spatial Doppler frequency representation of the data is produced through a single FFT in the distance dimension of the data, and is shown in Figure 2.4c. Finally, through application of equation (2.11), the filtered data is shown in Figure 2.4d.

It should be noted that some accommodations were necessary in order to achieve acceptable results for our experimental data. Platform velocity naturally varies over the course of the collection producing an inconsistent spacing, $\Delta d$, between consecutive platform positions resulting in phase noise. In order to account for this phase noise it was necessary to increase the size of the passband region. The results of a brute-force search determined that expanding the bandwidth of the filter by 40% resulted in improved post-filter signal to background performance. We further replaced the

**(a)** Passband Filter

**(b)** Collected Data

**(c)** Doppler Space, Pre-Filter

**(d)** Doppler Space, Post-Filter
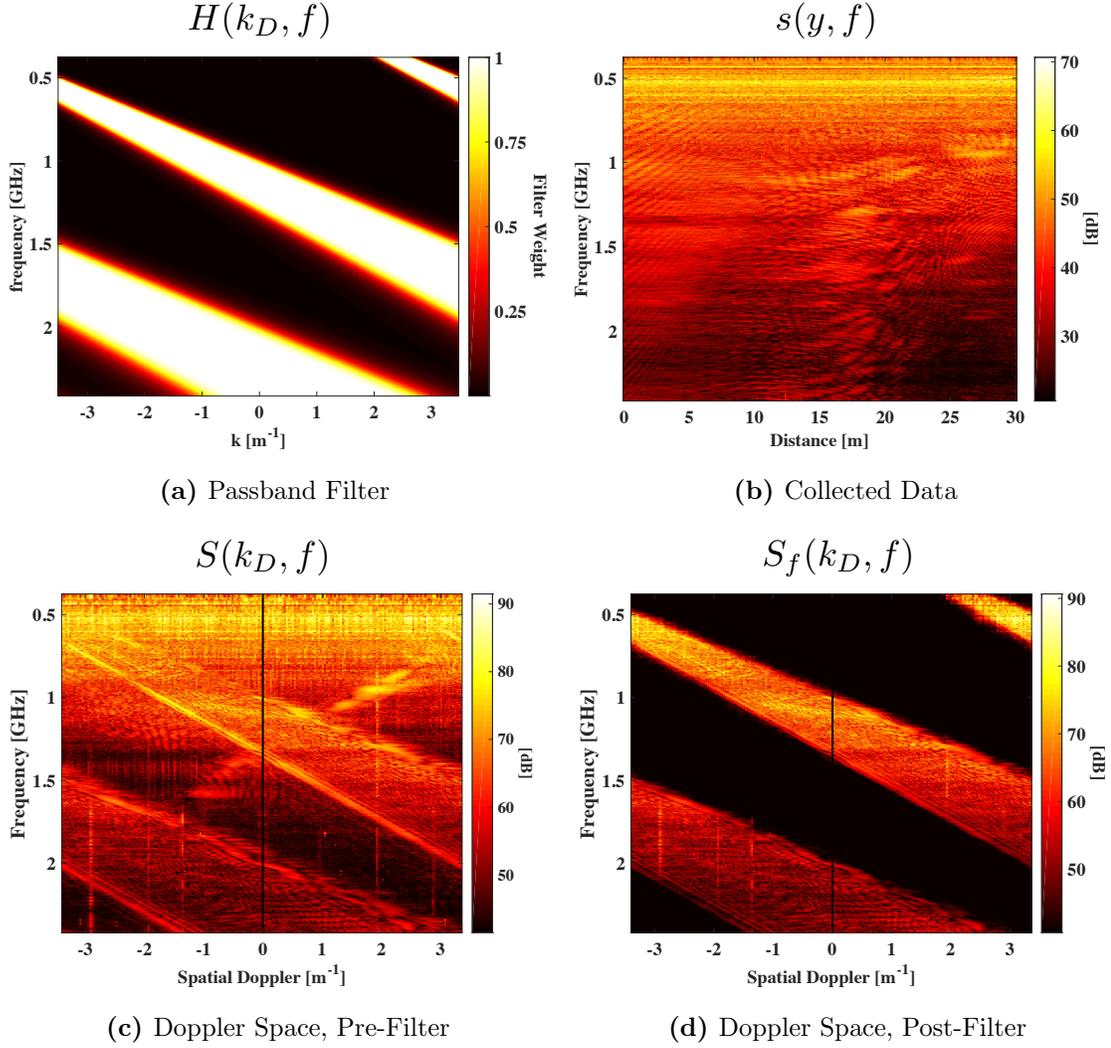
**Figure 2.4:** The passband Doppler filter (a) is applied to the collected data (b) in the transform domain (c), resulting in the filtered data (d).

brick-wall filter presented in equation (2.10) with a 6th order passband Butterworth filter in order to gracefully transition out of the passband region and minimize ringing. It should further be noted that a DC notch filter is also applied in the spatial Doppler dimension to remove systematic frequency dependent bias.

Figures 2.5a and 2.5b show the spatial Doppler contours of the surface in the region

23

**(a)** Surface Doppler at 0.5 GHz

**(b)** Surface Doppler at 3.0 GHz
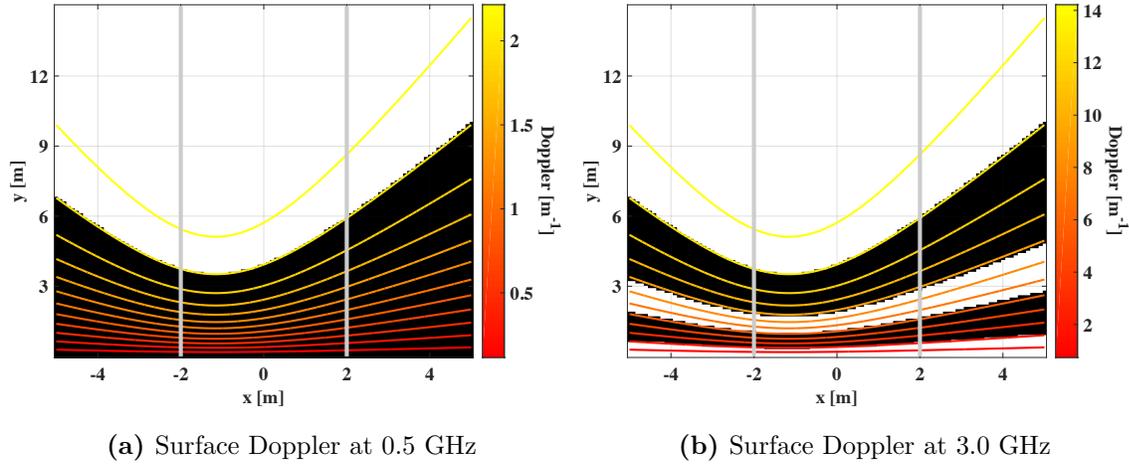
**Figure 2.5:** The surface Doppler contours at the lowest frequency (a) and highest frequency (b) determine which regions pass through or are rejected by the filter.

of the platform computed at the lowest and highest collected frequencies respectively. The passband mask mapped onto the ground plane is superimposed onto the Doppler contour plots clearly illustrating regions where the scattered surface energy is either passed or rejected by the filter. Based on the passband masks, it is clear that at the low end of the collected frequency spectrum the spatial Doppler associated with surface returns is unaliased while aliasing occurs at the higher end of the spectrum. The presence of aliasing results in out-of-lane regions whose Doppler matches in-lane returns and as a result are passed through the filter.

Images formed using conventional backprojection processing with and without spatial Doppler filtering are shown in Figures 2.6a and 2.6b respectively. Each image is normalized to the peak response and a dynamic range of 15 dB is shown. (This will continue to be the case for the all images in this section.) Two known targets are

24

**(a)** Pre-Filter           **(b)** Post-Filter
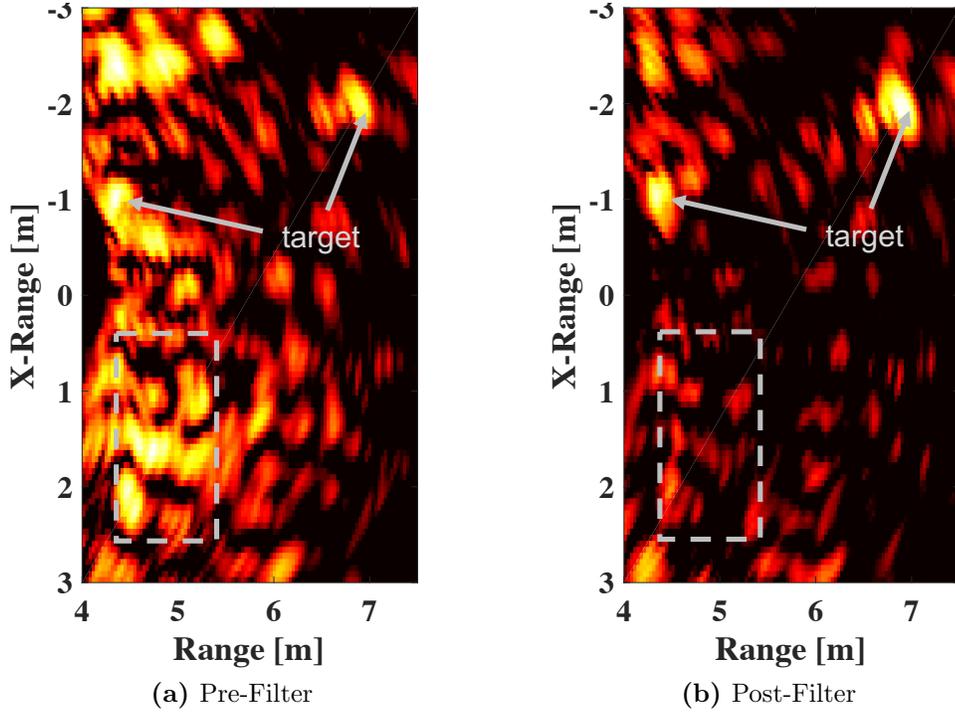
**Figure 2.6:** The sample image demonstrates significant reduction in background energy when comparing before (a) and after (b) Doppler filtering.

identified along with a selected background region with significant corruption. The peak signal to background ratio is computed for both targets with respect to the identified background region before and after Doppler filtering by

$$SBR = 10 \log \left( \frac{|\mathrm{x}_{pk}|^2}{\frac{1}{N_{bgd}} \|\mathbf{x}_{bgd}\|^2} \right), \tag{2.19}$$

where, $\mathrm{x}_{pk}$ is the peak value of the target, $\mathbf{x}_{bgd}$ is the vector of pixel values in the identified background region, and $N_{bgd}$ is the number of samples in the background region. In general the level of improvement is dependent on the level of corruption. For this particular example the observed signal to background ratio improves from

5.0 dB to 13.20 dB for the target on the left side of the image and similarly from 7.5 dB to 12.3 dB for the target on the right.

## 2.5.2 Imaging Results

Two high-contrast targets are chosen for demonstration which will simply be referred to as *Target A* and *Target B* (Note that these are not the targets illustrated in Figure 2.6). Imaging is performed using 30 frames taken at ranges of approximately 6 to 10 m from scene center. Conventional backprojection imaging is applied and compared to image reconstructions performed using the proposed technique. In all cases the non-linear unconstrained optimization problem is solved using the limited memory BFGS method. Processed images of *Target A* and *Target B* and are shown in figures 2.7 and 2.8 respectively. The intended targets are easily observed near scene-center. In both figures, the conventional backprojected images (Figures 2.7a and 2.8a) are shown next to reconstructions using dual TV and TE regularization (Figures 2.7b and 2.8b), and reconstructions using the TE regularization alone (Figures 2.7c and 2.8c).

In a fielded system representative training data should be used to optimize the choice of regularization parameters with respect to detection/classification performance. In other work, even multiple reconstructions have been considered with the intent of

26

**(a)** Backprojection     **(b)** Dual Regularization     **(c)** TE Regularization
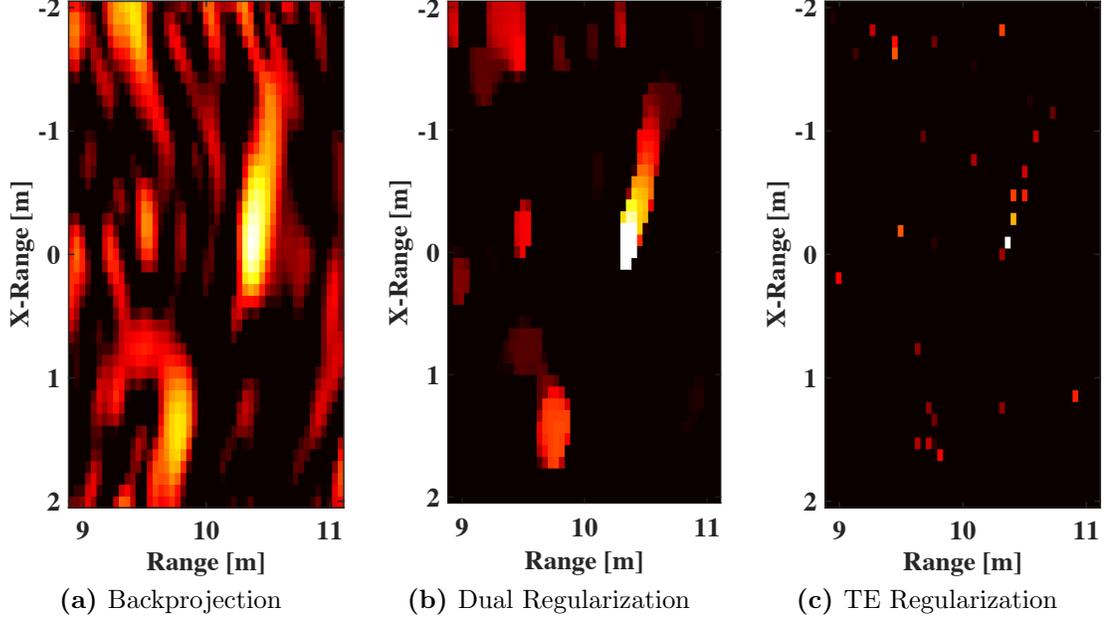
**Figure 2.7:** Target A is imaged using conventional backprojection (a), dual TV and TE regularization (b) and with TE regularization alone (c).

emphasizing certain features to be used for classification [41]. For the purposes of our demonstration, regularization parameter selection was done through trial and error. In the dual regularization case, a grid search was performed and a set of regularization parameters were hand picked such that resolution was improved but extended objects remain "intact" when compared to the backprojected result. In the TE regularization only case the regularization parameter was hand picked such that extended objects were reduced to only a few point scatterers. The dual regularized reconstruction is generated using regularization coefficients of $\beta_{TE} = 120$, and $\beta_{TV} = 20$ while the TE only reconstruction is generated using regularization coefficients of $\beta_{TE} = 240$, $\beta_{TV} = 0$.
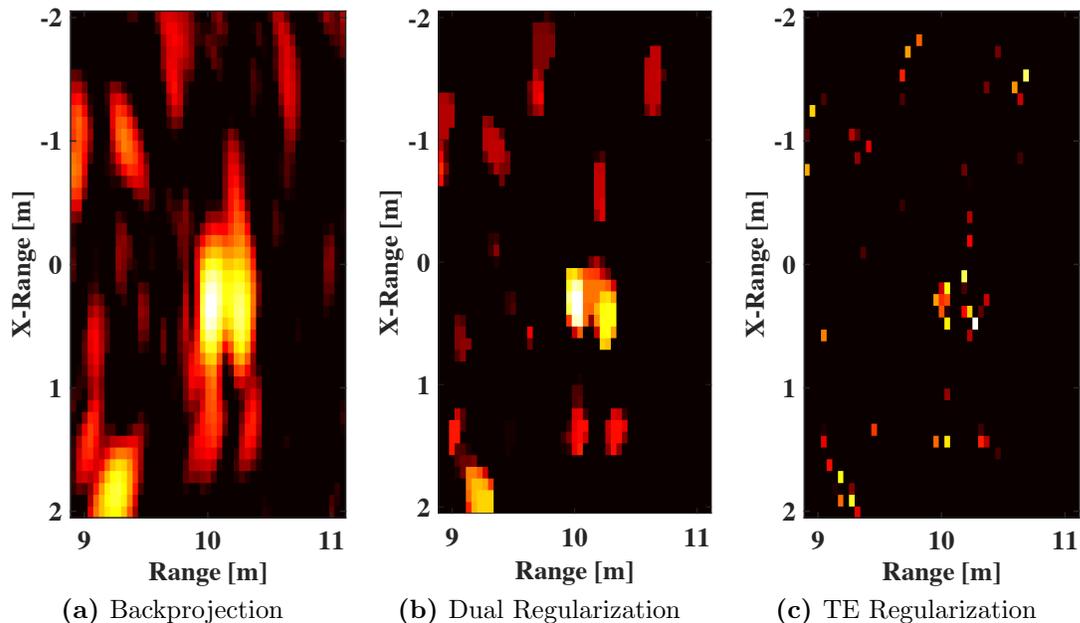
**Figure 2.8:** Target B is imaged using conventional backprojection (a), dual TV and TE regularization (b), and with TE regularization alone (c).

Slice plots through the peak target value in the range and cross-range dimensions allow for easy comparison of the target response and are shown in Figures 2.9 and 2.10 for targets $A$ and $B$ respectively. In both cases, slices are compared between backprojected images and image reconstructions using dual regularization. The dual regularization approach demonstrates better resolution when compared to the conventional backprojection. The improvement is dramatic in the cross-range dimension where the conventional image resolution is large in comparison to the actual object dimensions, but minimal in the range dimension where the conventional resolution is small due to the wide-band system.
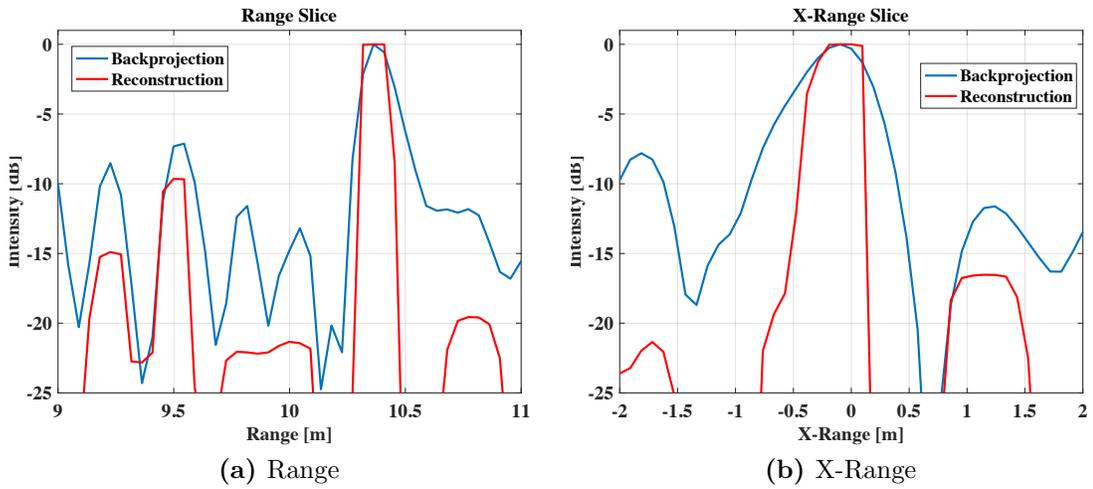
**(a)** Range  **(b)** X-Range

**Figure 2.9:** Slice plots through the peak response of target A in the range (a) and x-range (b) dimensions allow for easy comparison in the resulting target resolution.
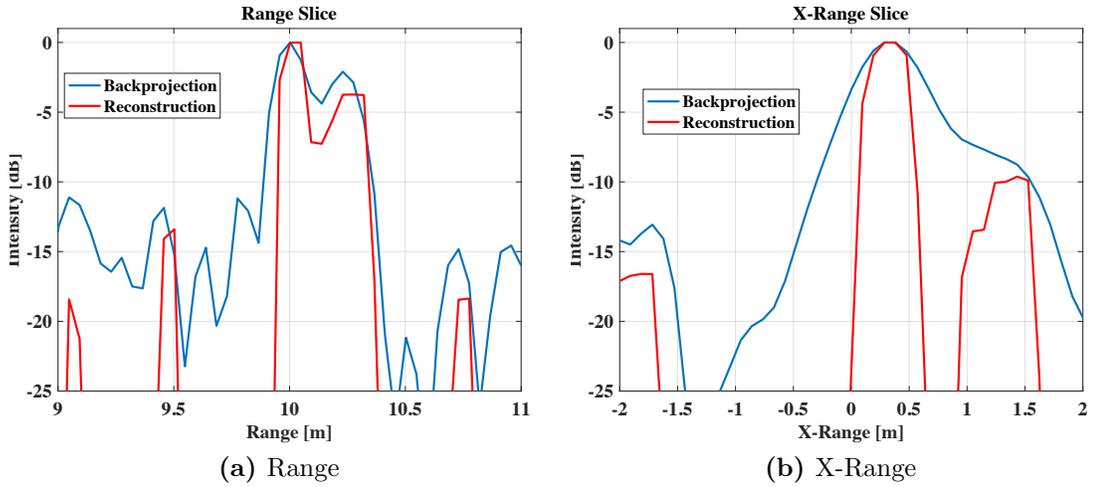


**(a)** Range  **(b)** X-Range

**Figure 2.10:** Slice plots through the peak response of target B in the range (a) and x-range (b) dimensions allow for easy comparison in the resulting target resolution.
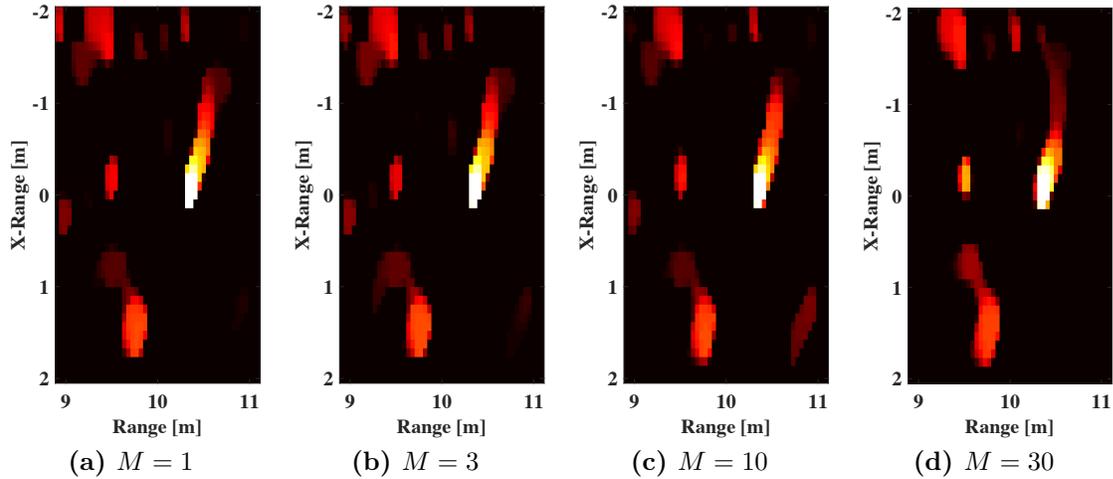
**Figure 2.11:** Image reconstructions of *Target A* are formed with pre-averaging of $M = 1, 3, 10,$ and $30$ frames.

### 2.5.3 Pre-Averaging Performance

In this section we include the pre-averaging operation and make comparisons regarding its computational benefits and effects on performance. Reconstruction of *Target A* and *Target B* is repeated with pre-averaging included and the resulting images are shown in Figures 2.11 and 2.12, respectively. Images are formed with pre-averaging of $M = 1, 3, 10,$ and $30$ frames such that when $M = 1$ the reconstruction is performed with no pre-averaging applied and when $M = 30$ reconstruction is performed after pre-averaging the entire set of 30 frames. In fact, in can be noted that Figure 2.11a is identical to Figure 2.7b and that Figure 2.12a is identical to Figure 2.8b.

If we stipulate that the best imaging performance will be achieved in the absence
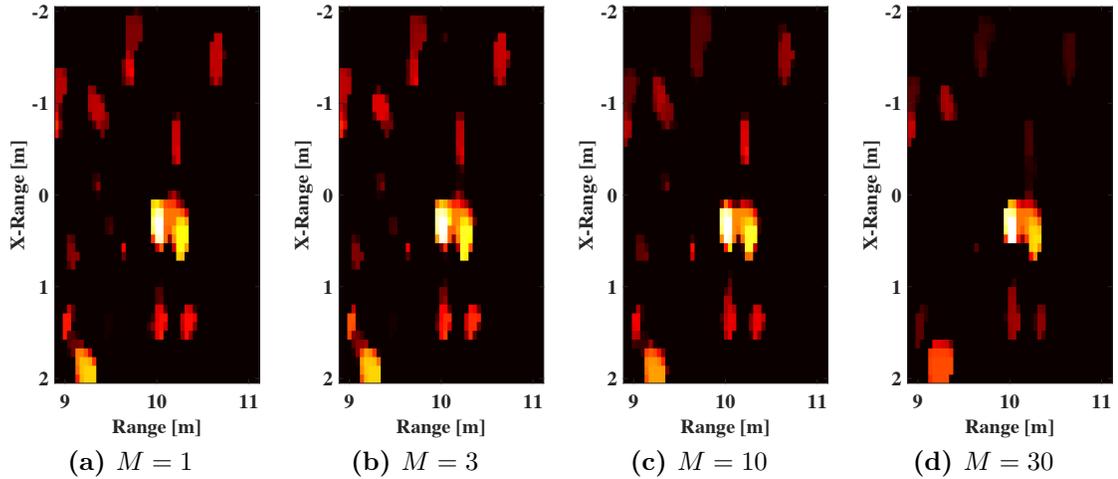
**Figure 2.12:** Image reconstructions of *Target B* are formed with pre-averaging of $M = 1, 3, 10$, and 30 frames.

| (a) $M = 1$ | (b) $M = 3$ | (c) $M = 10$ | (d) $M = 30$ |

of pre-averaging then the effects of averaging can be characterized using the *mean squared error* (MSE) relative to the image formed with no averaging applied. Image reconstructions were performed using an off-the-shelf 2013 MacBook Pro (Processor: 2.6 GHz Intel Core i5, Memory: 16 GB 1600 MHz DDR3) and the total reconstruction times were recorded. The average computation time and the effects on performance are summarized in Figure 2.13. As the number of averaged frames increases the relative error increases and the computation time decreases. The improvement in computation time diminishes for higher levels of averaging where the setup time required to pre-process the data and build the forward operator dominates the solution of the actual optimization problem.
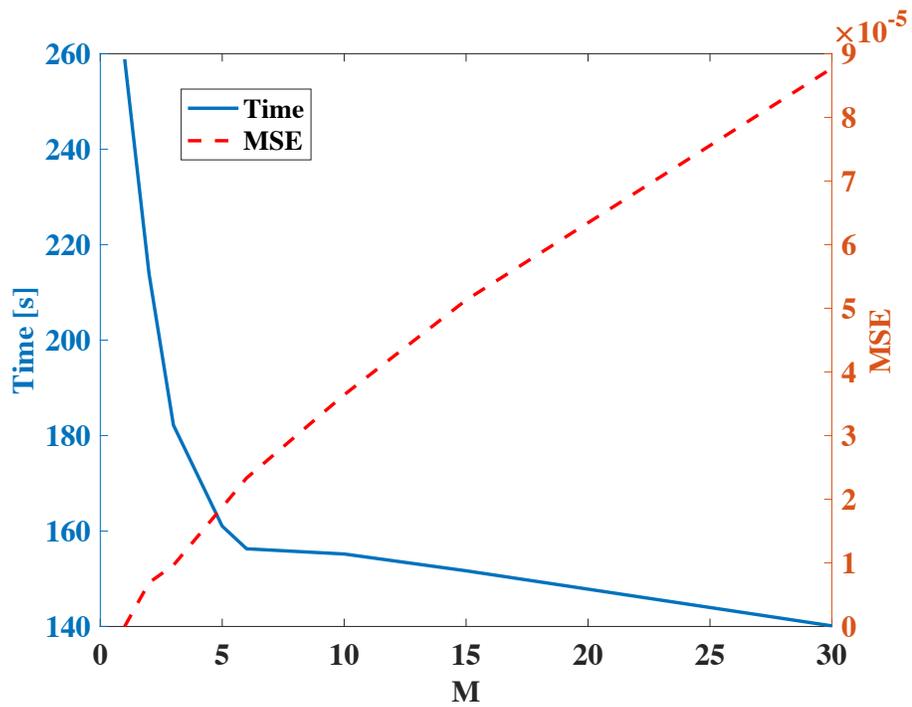
**Figure 2.13:** The performance of pre-averaging is characterized using image reconstruction time and MSE.

# Chapter 3

# A Penalized Least Squares Technique for Imaging with Hypertelescopes

## 3.1 Introduction

Imaging of space-based objects from ground sensors is a highly desirable technical capability for astronomers and has important applications within the defence community as well. Practical considerations limit the aperture size of conventional telescopes, which when combined with atmospheric turbulence effects limit the achievable

angular resolution even in the presence of adaptive optics. Sparse aperture telescopes represent a viable alternative achieving increased angular resolution by combining light collected from elements distributed over a wide spatial area. There are two ways to combine the light collected in this manner: (1) using amplitude interferometery based on the van Cittert-Zernike theorem [44]; and (2) by using a so-called densifier followed by an imaging lens to form a direct image. The densified hypertelescope concept has demonstrated excellent imaging performance with direct image formation on targets of interest to the astronomical community and is the relevant concept in this chapter [45, 46].

Central to the densified hypertelescope concept is a *pupil densification* transformation. *Pupil densification* describes a system where signals collected from array elements are transported to a central location where they are expanded and combined using a scaled replica of the actual array (i.e. the densified pupil). The output light from the densifier represents a generalized pupil plane which will be discussed in detail in section 3.2. A final image forming lens is used to create the measured direct image. This approach is distinct from amplitude interferometry, where the final Fourier transform required to form an image is computed digitally. One practical implementation of a very large hypertelescope design utilizes optical fibers to carry the signals collected at each array element to the beam combiner. [46, 47, 48] An inevitable consequence of sparse aperture sampling is that the array point-spread-function (PSF) associated with the direct image exhibits enhanced side-lobes which

are indicative of the placement of the array elements.[49] Densification has the effect of concentrating light which would be scattered into a very wide angle in image space, into the center of the image, but at the cost of imposing and envelope function which makes the measured images be non-shift invariant. However, without densification the measured images would have very low signal-to-noise ratio, and would also be corrupted with artifacts from the repetitive nature of the array PSF. As such, a post-processing image reconstruction step is highly desirable to maximize image quality.

Previous efforts relevant to hypertelescope image reconstruction have successfully demonstrated the ability to partially mitigate the adverse effects associated with the sparse aperture. Thiebaut and Young present an excellent tutorial paper covering many techniques applicable to interferometric imaging in general. [50] Aim et.al. demonstrates the Richardson-Lucy algorithm for hypertelescope imaging specifically as well as the Image Space Reconstruction Algorithm (ISRA) which finds the maximum likelihood estimate subject to a positivity constraint based on the model that the data is corrupted by a Gaussian noise process. [51] Surya proposes a speckle imaging approach when adaptive phasing of the sub-apertures is not available. [52]

In this chapter, the penalized least squares image reconstruction approach described in chapter 2 is adapted to densified hypertelescope imaging. The penalized least squares approach is demonstrated using a variety of regularization parameter configurations and compared to the conventional Richardson-Lucy deconvolution algorithm which

35

has been adapted to the densified hypertelescope paradigm. Image reconstruction quality is presented in terms of the structure similarity metric (SSIM). [53]

The key findings are as follows; penalized least squares image reconstruction is a robust approach to mitigating the adverse effects associated with the PSF in sparse aperture imaging; The SSIM metric achieved using the penalized least squares approach is significantly higher than what is achieved using conventional Richardson-Lucy deconvolution given a suitable selection of regularization parameters; Many combinations of regularization parameters can be applied which achieve good results and can be modified depending on the desired properties of the resulting image reconstruction.

The remainder of this chapter is organized as follows. In Section 3.2 we characterize the theoretical signal model appropriate for a fibered hypertelescope, in Section 3.3 we present the relevant image reconstruction approaches, and in=n Section 3.4 we postulate particular hypertelescope design parameters simulate direct images, perform reconstruction, and asses performance.

## 3.2   Signal Model

The signal model characterizing the object-image relationship for a variety of hypertelescope implementations has been presented previously in the literature [45, 46].

The version relevant to this work has been presented in [47] and is summarized here. A conceptual illustration of the densified hypertelescope is shown in figure 3.1. The densification is quantified by the scale factor relating the collection array and the densified array and is given by $\gamma_b = \frac{b'}{b}$ where $b$ is the baseline associated with the actual array and $b'$ is the baseline associated with the densified array. The hypertelescope samples the incident wavefront using a collection of sub-apertures dispersed over a large spatial region. Each sub-aperture is coupled to an optical fiber which
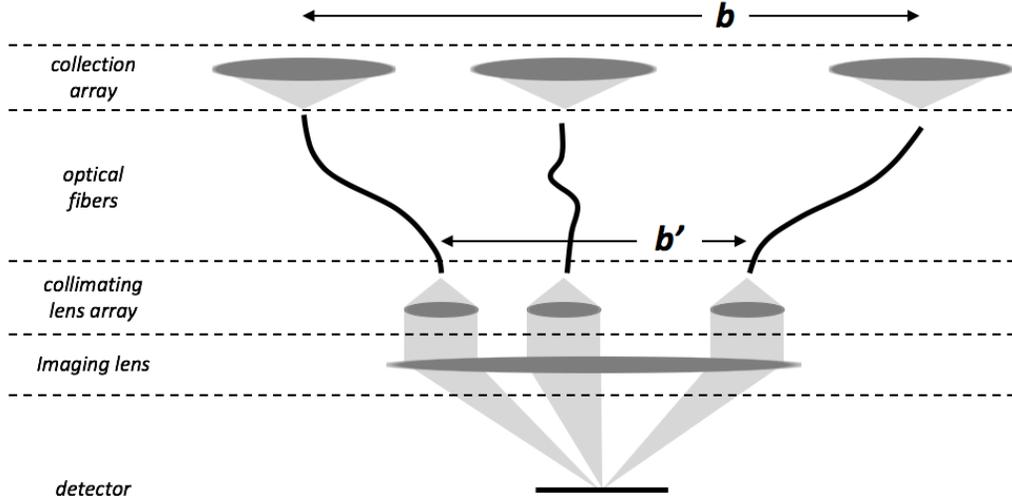


**Figure 3.1:** A conceptual model illustrates the hypertelescope imaging concept.

transports the collected signal to a densified replica of the array. It is important to note that the tip of the optical fiber in the collecting telescopes is a strong spatial filter, so that the field entering each fiber from an array telescope is reasonably modeled as the zero frequency component of the field falling on that aperture. The signal emitted from each optical fiber in the beam combiner is passed through a collimating

lens. The wavefront produced by the entire densified array is then passed through a final imaging lens before detection. The wavefront exiting the densified array can be visualized as a densified replica of the wavefront sampled by the original collection array with an an additional weighting applied to each sub-aperture incurred due to the amplitude distribution pattern produced by the fiber and collimating lens. In our model it is assumed that the amplitude distribution across the spatial extent of the collimating lens uniform. As such the amplitude distribution of the signal associated with each sub-aperture after passing through the fiber and collimating lens is given by,

$$\psi(x, y) = \text{circ}\left(\frac{\sqrt{x^2 + y^2}}{R_0}\right), \tag{3.1}$$

where $R_0$ is the radius of the collimating lens. An array function characterizing the placement of the array elements is defined as,

$$\mathcal{A}(x, y) = \sum_{n=1}^{N} \delta(x - x_n, y - y_n), \tag{3.2}$$

where the coordinates of the $N$ array elements are given by $(x_n, y_n)$. The pupil function for the entire array is then the convolution of the array function with the amplitude distribution and is given by,

$$P(x, y) = \psi(x, y) \circledast \mathcal{A}(x, y). \tag{3.3}$$

From the pupil function the object-image relationship can be computed through the array theorem and is the product of an *envelope function* which represents the weighting function imposed by the presence of the fiber/collimating lens combination and an *interference function* which is a direct consequence of the placement of the array elements. [54]. The *envelope function* is the magnitude squared of the Fourier transform of the amplitude distribution function $\psi(x, y)$, while the *interference function* is the magnitude squared of the Fourier transform of the array function $\mathcal{A}(x, y)$. A unique feature resulting from propagation through the fiber is that the tilt associated with the angle of arrival within each of the sub-apertures is lost. This results in an envelope function which is independent of the location of the source while the interference function remains a function of the source coordinates. If we let the vector $\boldsymbol{\theta} = [\theta_x, \theta_y]^T$ describe the angular coordinates in object space and the vector $\boldsymbol{\alpha} = [\alpha_x, \alpha_y]^T$ describe the angular coordinates in image space then the object space coordinates are related to the image space coordinates by $\boldsymbol{\alpha} = \frac{\boldsymbol{\theta}}{\gamma_b}$. The envelope function can then be expressed as,

$$\mathcal{E}(\alpha_x, \alpha_y) = \left| \mathcal{F}\{\psi(x, y)\} \right|^2 \Bigg|_{k_x = \frac{\alpha_x}{\lambda}, k_y = \frac{\alpha_y}{\lambda}} \tag{3.4}$$

while the interference function is given by,

$$\mathcal{I}(\theta_x, \theta_y) = \left| \mathcal{F}\{\mathcal{A}(x, y)\} \right|^2 \Bigg|_{k_x = \frac{\theta_x}{\gamma_b \lambda}, k_y = \frac{\theta_y}{\gamma_b \lambda}}. \tag{3.5}$$

39

Upon inspection, the object-image relationship can be expressed as a convolution between a scaled version of the object with the scaled interference function which is then multiplied by the envelope function. The signal model for the direct image can finally be expressed as,

$$x\left(\alpha_x, \alpha_y\right) = \mathcal{E}\left(\alpha_x, \alpha_y\right) \times \left(\mathcal{I}\left(\gamma_b \alpha_x, \gamma_b \alpha_y\right) \circledast g\left(\gamma_b \alpha_x, \gamma_b \alpha_y\right)\right), \qquad (3.6)$$

where $x\left(\alpha_x, \alpha_y\right)$ is the direct image and $g\left(\gamma_b \alpha_x, \gamma_b \alpha_y\right)$ is the true object evaluated in terms of the image coordinates scaled by the array densification factor.

For image processing applications it is typically sufficient to consider a discretized signal model which utilizes a series of linear operators. The two-dimensional discretized object is represented as a single lexicographically reordered vector $\mathbf{g}$. The convolution operation is expressed as a matrix-vector multiplication where the convolution kernel is encoded into a square matrix $\mathbf{H}$. The envelope is applied as a square diagonal matrix $\mathbf{\Psi}$. Finally, the direct image is represented by the lexicographically ordered vector $\mathbf{x}$. In it's discretized form the object image relationship is given by,

$$\mathbf{x} = \mathbf{\Psi H g}. \qquad (3.7)$$

Extensions beyond this linear model may be necessary to account for differences in the sizes of the object and data vectors due to differences in sample size or differences

in the spatial extents of the object and data as it is often necessary to account for objects which extend beyond the field of view of the detector. As such we introduce a resampling operator $\mathbf{B}$ and a truncation operator $\mathbf{M}$. The resampling operator is applied through a non-square matrix which reduces the sample size through local area integration (i.e. binning). The truncation operator is a second non-square matrix which extracts only the samples included within the finite extent of the detector. After accounting for resampling and truncation the linear model for the direct image is given by,

$$\mathbf{x} = \mathbf{MB\Psi Hg}. \tag{3.8}$$

The collected data will inevitably be corrupted by noise resulting from the photodetection process. Contributions to the collected data include the intensity of the light incident on the detector as well as a read noise contribution intrinsic to the detector itself. The light intensity is modeled as a Poisson random variable with an expected value of $\mathbf{x}$, while the read noise contribution is modeled as a zero mean Normal random variable variance of $\sigma_R^2$. In the sections that follow we use the variable $\mathbf{d}$ to represent the direct image after accounting for noise.

## 3.3   Image Reconstruction

In this section we describe the traditional Richardson-Lucy algorithm as it is applied to fibered hypertelescope imaging. We then present our novel penalized least squares (PLS) approach which has been applied previously for image reconstruction in radar applications and is adapted here for hypertelescope imaging. In this section we express the forward model by the generic linear operator $\mathbf{F}$. The details describing how the forward model is implemented will be described in detail in section 2.5.

### 3.3.1   Richardson-Lucy

Likely the most prevalent approach to deconvolution of images is the Richardson-Lucy algorithm [55, 56]. The classical Richardson-Lucy algorithm computes the maximum likelihood estimate of an object which has been blurred by a linear space invariant convolution and corrupted by a Poisson noise process subject to a positivity constraint. The log-likelihood associated with a vector of independent Poisson random variables is given with constant terms omitted by,

$$\mathcal{L}\left(\mathbf{g}\right) = \mathbf{d}^{T}\log\left(\mathbf{Fg}\right) - \mathbf{1}^{T}\mathbf{Fg}. \tag{3.9}$$

The maximum likelihood estimate of the object is solved through the constrained optimization problem defined by,

$$\hat{\mathbf{g}} = \arg\max\left(\mathcal{L}\left(\mathbf{g}\right)\right) \qquad \text{s.t. } \mathbf{g} \geq 0. \tag{3.10}$$

The Richardson-Lucy approach solves the optimization problem iteratively using the expectation maximization (EM) algorithm which produces the update equation given by,

$$\hat{\mathbf{g}}_{k+1} = \frac{\hat{\mathbf{g}}_k}{\mathbf{F}^T\mathbf{1}} \odot \left(\mathbf{F}^T \frac{\mathbf{d}}{\mathbf{F}\mathbf{g}_k}\right) \tag{3.11}$$

[57]. An inherent feature of the Richardson-Lucy approach is that as long as the initial estimate contains values which are all greater than zero, then the estimates at each iteration will also remain greater than zero, and the positivity constraint is inherently satisfied.

## 3.3.2 Penalized Least Squares

Penalized least squares provides a flexible framework which utilizes penalty functions to mitigate overfitting by encouraging solutions that are consistent with prior expectations. The penalized least squares objective function considered in this chapter is given by,

$$\mathcal{O}\left(\mathbf{g}\right) = ||\mathbf{d} - \mathbf{F}\mathbf{g}||_2^2 + \beta_E||\mathbf{g}||_p^p + \beta_V||\nabla\mathbf{g}||_q^q. \tag{3.12}$$

Object estimation is accomplished through the constrained optimization problem defined by,

$$\hat{\mathbf{g}} = \arg\min\left(\mathcal{O}\left(\mathbf{g}\right)\right) \qquad \text{s.t. } \mathbf{g} \geq 0. \tag{3.13}$$

In this case, the estimated object, $\hat{\mathbf{g}}$, is found through convex optimization. The first term of the objective function represents the data fit term measuring the *mean squared error* (MSE) between the collected data and the predicted data based on the object $\mathbf{g}$. The proposed penalty terms include a *total energy* (*TE*) regularization term penalizing the $\ell_p$ norm of the magnitude of the object and a *total variation* (*TV*) regularization term penalizing the $\ell_q$ norm of the numerical gradient of the magnitude of the object. In the absence of regularization, object estimates based on MSE alone tend to include rapidly fluctuating high energy values which are unexpected, if not unphysical, yet result in low error in measurement space. Four tuning parameters exist in equation 3.12 to guide object estimates. The coefficient $\beta_E$ controls the weight associated with the *TE* penalty while $p$ controls the level of sparsity in the object estimate. Similarly, the coefficient $\beta_V$ controls the weight associated with the *TV* penalty while $q$ controls the level of sparsity in the numerical gradient of the object estimate.

## 3.4 Results

In this section we postulate a notional hypertelescope design. Direct images are simulated based on the discussion in section 3.2 and reconstructions are performed as described in section 3.3. Performance is assessed in terms of the structure similarity metric (SSIM) which has been shown to align well with subjective measures of quality associated with the human visual experience. [53].

### 3.4.1 Simulation

The postulated array consists of 76 apertures arranged in a spoke pattern with a maximum diameter of $115\,m$. Imaging is performed at a wavelength of $1150\,nm$ resulting in a theoretical angular resolution of approximately $10\,nRad$. Based on this design the envelope and interference patterns are computed as described in equations 3.1-3.5. The postulated array locations, the resulting interference function, and the envelope are shown in figures 3.2a, 3.2b, and 3.2c respectively.

The postulated detector is a $20 \times 20$ array with $12.5\,nRad$ pixels and the collected data is represented by a vector $\mathbf{d}$ consisting of 400 samples. Note that we choose a detector sampling such that the direct image is slightly aliased with respect to the resolution
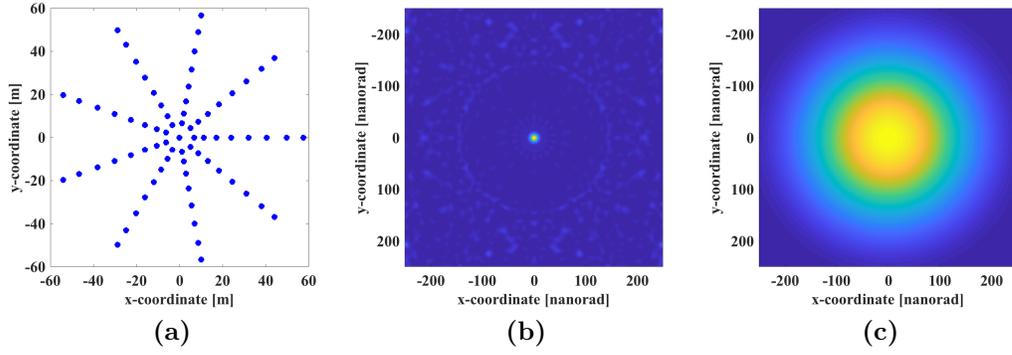
**Figure 3.2:** The postulated array (a) is shown along with the applicable interference function (b) and envelope function (c).

dictated by the postulated array. We make this trade in order to increase the per pixel signal level in order to improve imaging performance of low intensity objects. For simulation purposes high resolution object representations consisting of $(400 \times 400)$ pixels are utilized resulting in a total of $K = 160000$ object samples. The object samples are sampled at rate of $1.25\,nRad$, $10\times$ upsampled with respect to the detector sampling, and represent an angular area with dimensions that are twice that of the field of view observed by the detector. As such, in our simulation the convolution and envelope are expressed as linear operators characterized by $\mathbf{H} \in \mathbb{R}^{K \times K}$ and $\mathbf{\Psi} \in \mathbb{R}^{K \times K}$ respectively. The downsampling operator reduces the sample spacing from the simulated object spacing of $1.25\,nRad$ to the detector spacing of $12.5\,nRad$ and is characterized by $\mathbf{B} \in \mathbb{R}^{M \times K}$ where $M = 1600$. Finally, the truncation operator extracts only the samples included in the $20 \times 20$ detector array and is characterized by $\mathbf{M} \in \mathbb{R}^{N \times M}$ where $N = 400$.

A model of the Galaxy 15 telecommunications satellite is chosen for demonstration

46

of the proposed image reconstruction algorithms. Representative data is produced by first simulating the noise-free direct image as defined in equation 3.8 and then accounting for noise as described at the end of section 3.2. In our model we assume a relative reflectivity map and set the total photon count based on a visual magnitude of 8 relative to the sun as described in [58]. Given a bandwidth of $(300)\,nm$, a total aperture area of $30\,m^2$, a transmission coefficient of $0.035$, and an integration time of $50\,ms$ the total collection results in approximately $225000\,photons$. The read noise associated with the detector is modeled as a zero mean Gaussian random variable with a standard deviation given by $\sigma = 8$. The relative reflectivity map associated with the high resolution test object is shown in 3.3a with the relevant field of view indicated by a dashed yellow line. The resulting simulation of the direct image is shown in figure 3.3b. Lastly, a resampled version of the pristine reflectivity map is shown in figure 3.3c where the resampling is done such that the sample rate matches the rate at which reconstructions will be performed.
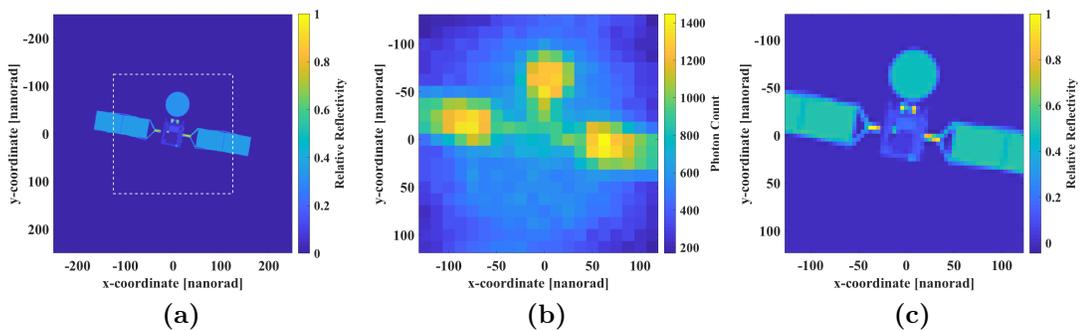


**Figure 3.3:** The high resolution model of the galaxy 15 telecommunications satellite (a) is shown along with the simulated test data (b) and an ideal representation of the object sampled at the reconstruction sampling rate (c).

### 3.4.2 Reconstruction

The classical Richardson-Lucy formulation is derived such that the sample rate of the object estimate matches that of the collected data. The reconstructed object is represented by the vector $\mathbf{g}$ consisting of $K = 1600$ samples representing a $40 \times 40$ grid sampled at $12.5 \, nRad$. Note that the object grid is defined with sufficient guard-band to account for the potential of bright objects outside the field of view of the detector. The forward model utilized for our Richardson-Lucy reconstructions is then given by,

$$\mathbf{F} = \mathbf{M\Psi Hg}. \tag{3.14}$$

Note the absence of any downsampling operation. The Richardson-Lucy estimate $\hat{\mathbf{g}}$ of the test object is shown in Figure 3.4a at detector resolution and an upsampled version is shown in Figure 3.4b for comparison with the penalized least squares reconstructions below. Upsampling is performed using standard bi-cubic interpolation. The structure similarity metric computed based on the upsampled Richardson-Lucy estimate is 0.60.

The presence of regularization in the penalized least squares formulation provides the opportunity to perform object estimates at an increased sample rate with respect to the collected data. We choose an object-estimate grid which is sampled at a rate which is 3× the sample rate of the detector and just over 2× the sample rate associated
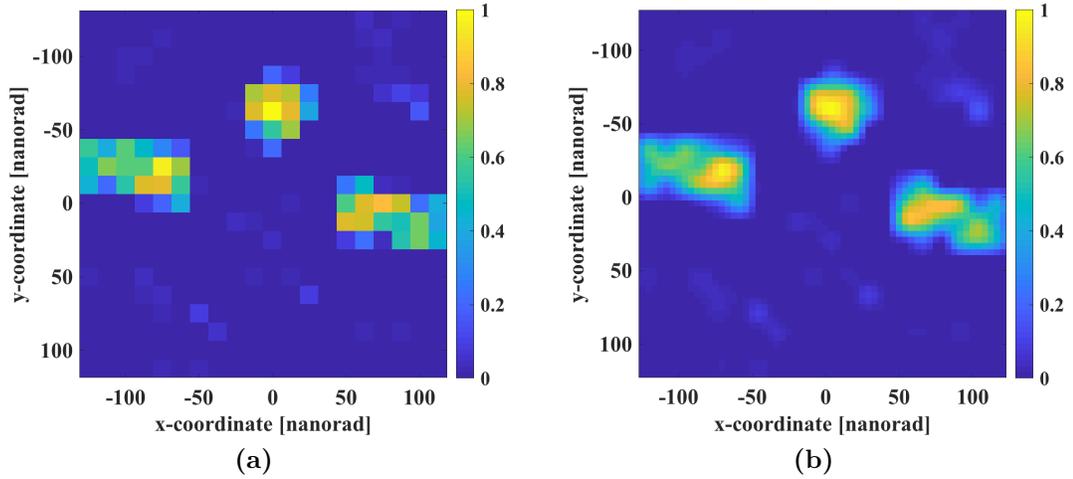
**Figure 3.4:** The Richardson-Lucy reconstruction is shown at detector resolution (a) and upsampled by 3× for comparison with the penalized least squares reconstruction (b).

with the theoretical resolution of the array. For penalized least squares estimates, the object is then represented by a vector $\mathbf{g}$ consisting of $K = 14400$ samples representing a $120 \times 120$ grid sampled at a rate of $12.5/3\,nRad$. The downsampling operator $\mathbf{B}$ reduces the sample rate through binning from $12.5/3\,nRad$ to $12.5\,nRad$ to match the sample rate of the detector and the truncation operator $\mathbf{M}$ once again extracts only the samples representing the $20 \times 20$ detector array. The forward model utilized for our penalized least squares reconstructions has the same terms as the model used for simulation but differs due to the size of the reconstructed object. The model is repeated here for convenience,

$$\mathbf{F} = \mathbf{MB\Psi Hg}. \tag{3.15}$$

49

As discussed in section 3.3 the penalized least squares objective function includes four parameters that can be adjusted in order to tune the properties of the object estimate. Here we present a collection of results which utilize different combinations of regularization parameters in order to demonstrate the efficacy of the approach and how the choice of regularization effects the object estimate.

Our approach is to first choose $p$ and $q$ values and then to perform a brute force grid search for $\beta_{TE}$ and $\beta_{TV}$ values which optimize the SSIM value. We consider a complete set of permutations of $p$ and $q$ values given that $p, q \in \{1.0, 1.5, 2.0\}$.

The comprehensive results of the grid search for the regularization parameters are illustrated in Figure 3.5 with the best SSIM values identified. The object estimates associated with the combination of parameters producing the best SSIM values are then shown in figure 3.6. The best overall result according to the SSIM value is achieved when the $\ell_1$ version of both penalty terms is applied ($p = 1, q = 1$). The sparsity inducing properties of the $\ell_1$ norm applied to both the total energy as well as the total variation lead to reconstructions which consist of contiguous energy containing regions of finite extent which is often the expectation when imaging objects in space.

As a general statement as $p$ increases the best reconstructions include more observed background energy which was suppressed when $p$ was chosen to induce sparsity. Similarly, as $q$ increases the best reconstructions contain edges which are less sharp and

more variability can be seen within distinct regions including the background. For higher values of $p$ and $q$ the best reconstructions are achieved as $\beta_E$ approaches 0, or in other words using *total variation* alone. This is particularly apparent from the plots in figures 3.5d - 3.5i where the weight associated with the *total energy* regularization term effectively goes to zero.
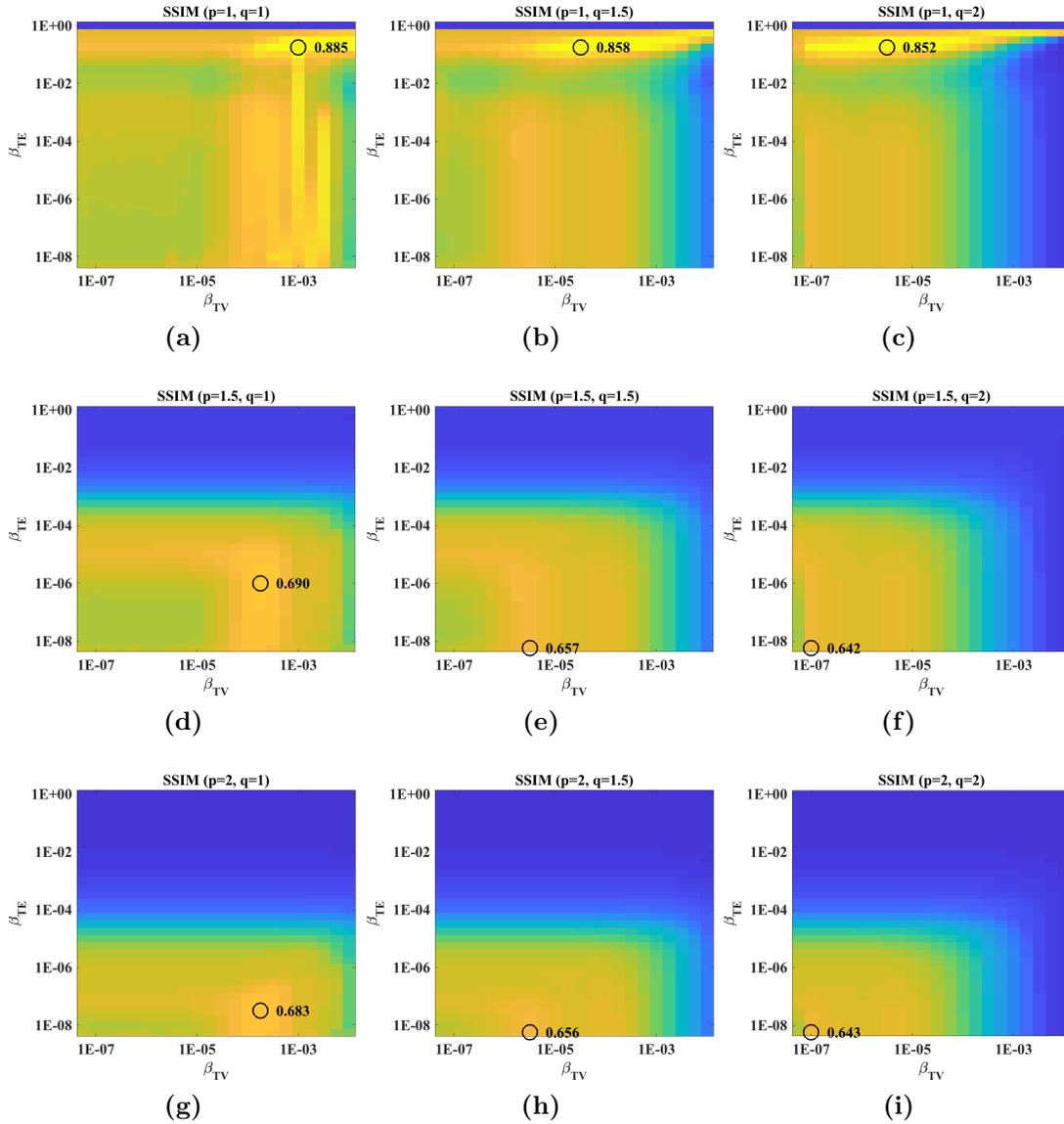
**Figure 3.5:** The SSIM values are computed for chosen combinations of $p$ and $q$ over an exhaustive grid of weight parameters including $p = 1, q = 1$ (a), $p = 1, q = 1.5$ (b), $p = 1, q = 2$ (c), $p = 1.5, q = 1$ (d), $p = 1.5, q = 1.5$ (e), $p = 1.5, q = 2$. (f), $p = 2, q = 1$ (g), $p = 2, q = 1.5$ (h), and $p = 2, q = 2$ (i) (Note all figures are shown on a [0 0.8] color axis) .
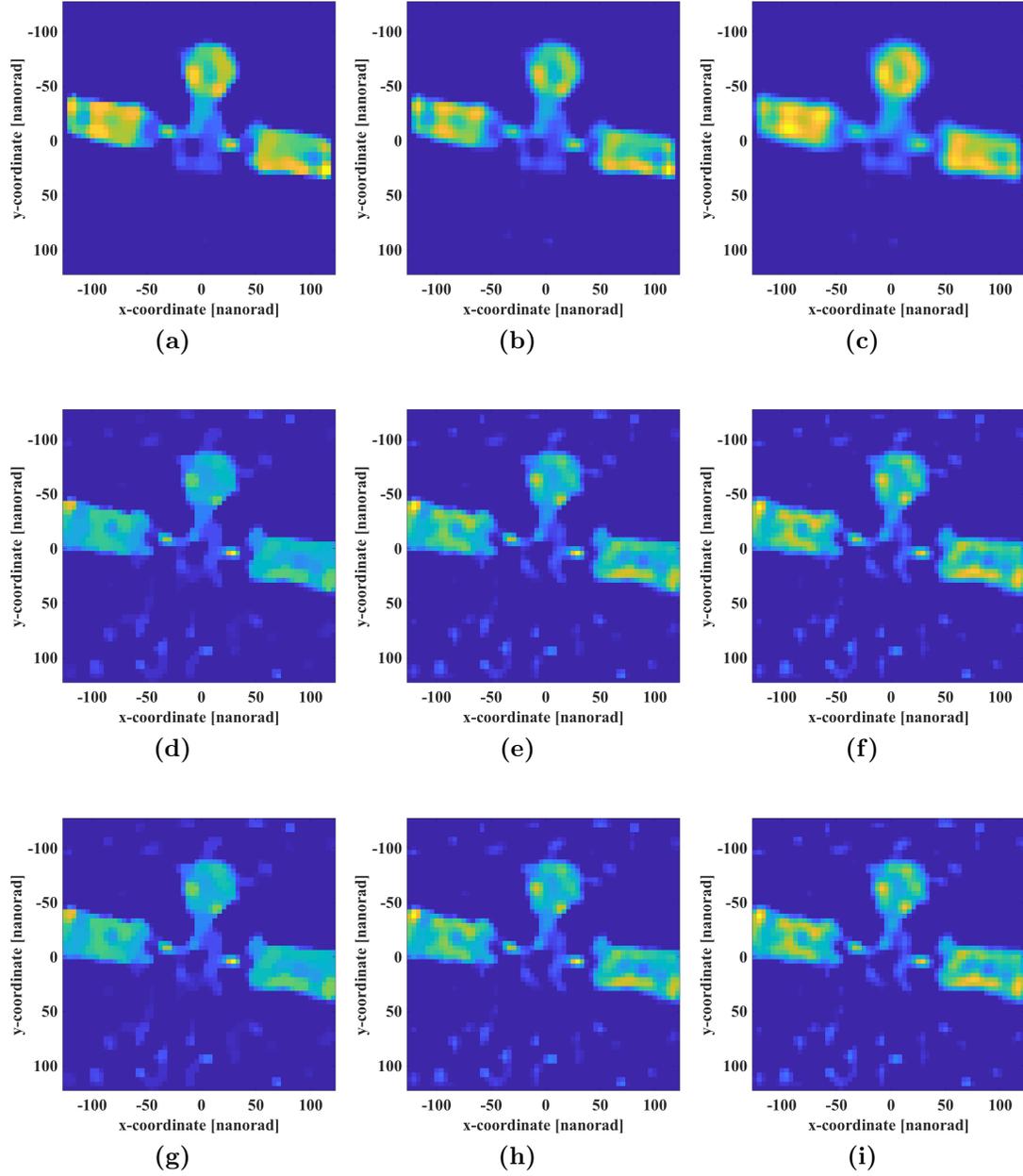
**Figure 3.6:** The reconstructions associated with the best SSIM values resulting from the grid search are shown for chosen combinations of $p$ and $q$ including $p = 1, q = 1$ (a), $p = 1, q = 1.5$ (b), $p = 1, q = 2$ (c), $p = 1.5, q = 1$ (d), $p = 1.5, q = 1.5$ (e), $p = 1.5, q = 2$. (f), $p = 2, q = 1$ (g), $p = 2, q = 1.5$ (h), and $p = 2, q = 2$ (i)

, and $p = 2, q = 2$ (i) (Note all figures are shown on a $[0\ 1.0]$ color axis)

# Chapter 4

# Atmospheric Turbulence Characterization through Multi-Frame Blind Deconvolution

## 4.1  Introduction

Propagation of light through turbulent media is commonly the dominant mechanism attributed to limiting performance of long range imaging, communication, and directed energy systems [59, 60]. In optical systems turbulence results in random fluctuations in space and time of the refractive index which perturbs the phase of the

wave front arriving at the sensor from an object of interest. Optical systems experience turbulence as a time varying aberration which degrades the ability to focus light. Complicated adaptive optics (AO) systems are sometimes used to compensate for the phase perturbations in real time, improving the instantaneous point spread function of the atmosphere-telescope system. AO system performance is highly dependent on the knowledge of the characteristics of the current turbulence environment. In this chapter we describe a method for simultaneously estimating both an improved image of the object, and the Fried parameter $r_0$ for the path by processing a set of short exposure images of the target in the presence of both turbulence and measurement noise.

Methods used to measure atmospheric turbulence using direct optical measurements include the differential image motion monitor (DIMM) [61], scintillation detection and ranging (SCIDAR) [62], path resolved optical profiling (PROPS) [63], and processing differential tilt measurements combined with analytic path weighting called the delayed tilt anisoplanatism (DELTA) method [64][65]. In astronomical applications it is also possible, at least in principle, to look at a bright star to measure one or more long exposure images, and use this data to estimate $r_0$. Note that while DIMM, SCIDAR, PROPS, and DELTA are powerful methods they require additional hardware, and the long exposure image method requires a point source target. In this chapter we present a method for estimating $r_0$ that does not require additional hardware, or a point source. This method is based upon the multi-frame blind deconvolution

(MFBD) idea [66], but with unique constraints placed on the point spread function (PSF) estimates as a function of $r_0$.

Deconvolution in image processing is a classical technique used to remove the blurring effect of the known point-spread-function (PSF). Blind deconvolution is the process of removing blur which is incurred from a PSF which in unknown or has unknown components. Multi-frame blind deconvolution (MFBD) is an approach which exploits multiple short exposure images of a common object in order to jointly estimate the object as well as the unique PSF observed at each frame. Previous applications of MFBD have been developed with the objective of improving image quality in a variety of applications where the PSF is partially unknown. [66, 67, 68, 69, 70, 71, 72, 73]

In this chapter we use MFBD as a means to estimate the magnitude of turbulence present over the path which light travels between an object and the sensor. The unique PSFs which are incurred at each frame fluctuate in a manner which is dictated by the level of turbulence. Stronger turbulence causes more dramatic fluctuations in the PSF while mild turbulence conditions cause less change in PSF. Here we exploit the theoretical relationship between the statistics of the PSF observed over a series of frames with the strength of the turbulence present in the medium of propagation. The natural metric to use to characterize the strength of the turbulence is the Fried parameter, also known as the *atmospheric coherence length*. As such, a joint estimation strategy is presented where the atmospheric coherence length value is estimated

along with the common object and the PSF at each frame. Key results are as follows; a novel technique for estimating $r_0$ from extended targets observed over multiple short-exposure frames is presented and using the presented technique excellent $r_0$ estimation performance is achieved; additionally highly accurate PSF estimates are obtained for each frame of the multi-frame processing interval and excellent image reconstruction performance of the common object is demonstrated.

The remainder of the chapter is organized as follows; in section 4.2 the relevant theory for imagining through turbulence including the so-called "short-exposure" PSF statistics is reviewed, in section 4.3 we describe our approach to MFBD with atmospheric coherence length estimation, and in section 4.4 we demonstrate the efficacy of the approach on both simulated and experimentally collected data.

## 4.2 Imaging Through Turbulence

The image formation process is characterized by the convolution of the system PSF and the object intensity distribution. This can be expressed mathematically as,

$$\boldsymbol{\mu} = \mathbf{h} \circledast \mathbf{x}, \tag{4.1}$$

where $\mathbf{x}$ is the object, $\mathbf{h}$ the non-coherent point spread function (PSF), $\boldsymbol{\mu}$ is the noise-free image and $\circledast$ represents the convolution operation.

The effects of imaging in the presence of atmospheric turbulence are well studied. The *Generalized Pupil Function* (GPF) characterizes the perturbations on the wavefront caused by propagation through the turbulent atmosphere in addition to the pupil and is given by,

$$\mathbf{q} = \mathbf{p} \odot e^{j\phi}, \tag{4.2}$$

where $\mathbf{p}$ is the pupil function, $\phi$ represents the phase induced by the wavefront perturbation, and $\odot$ is the Hadamard product operator. The coherent PSF is given as a Fourier transform of the GPF and the non-coherent PSF is given as the modulus squared of the coherent PSF and can be expressed as,

$$\mathbf{h} = |\mathcal{F}\{\mathbf{q}\}|^2. \tag{4.3}$$

The perturbed wavefront phase at the entrance pupil is well approximated in terms of it's Zernike polynomial decomposition such that,

$$\phi = \mathbf{Za}, \tag{4.4}$$

where the Zernike components are stored in the columns of the matrix $\mathbf{Z}$ and the

coefficients associated with each of the Zernike terms are stored in the vector, $\mathbf{a}$. Statistical models for the perturbed wavefront have been studied extensively and are given explicitly in terms of the Zernike polynomial coefficients as a function of the atmospheric coherence length, $r_0$. [59, 74] The distribution associated with the Zernike polynomial coefficients is zero-mean Gaussian with covariance, $\mathbf{\Sigma_a}$ where the elements of $\mathbf{\Sigma_a}$ are given by,

$$\Sigma_{i,j} = \frac{1}{r_0^{5/3}} 0.0072 D^{5/3} \pi^{8/3} (-1)^{n_i + n_j - 2m_i} \left( (n_i + 1)(n_j + 1) \right)^{1/2} \delta_{m_i, m_j} \tag{4.5}$$

$$\times \frac{\Gamma(14/3)\,\Gamma\left((n_i + n_j - 5/3)/2\right)}{\Gamma\left((n_i - n_j + 17/3)/2\right)\Gamma\left((n_j - n_i + 17/3)/2\right)\Gamma\left((n_i + n_j + 23/3)/2\right)}$$

for $i - j = even$, and zero for $i - j = odd$. In 4.5, $m_i$ and $n_i$ refer to the azimuthal and radial orders associated with the $i$th polynomial, respectively, $m_j$ and $n_j$ refer to the azimuthal and radial orders associated with the $j$th Zernike polynomial, and $\delta_{mi, mj}$ is the Kronecker delta function. It will prove convenient to express $\mathbf{\Sigma_a}$ with the $r_0$ dependence explicitly factored out. We thus define a matrix, $\mathbf{C}$ which is related to $\mathbf{\Sigma_a}$ as follows,

$$\mathbf{\Sigma_a} = \frac{1}{r_0^{5/3}} \mathbf{C}, \qquad \mathbf{\Sigma_a}^{-1} = r_0^{5/3} \mathbf{C}^{-1}, \qquad |\mathbf{\Sigma_a}| = \frac{1}{r_0^{N5/3}} |\mathbf{C}|. \tag{4.6}$$

## 4.2.1 Short Exposure PSF Statistics

The so-called short-exposure OTF characterizes the average blur induced by turbulence over a series of short-exposure frames discounting the translational effects associated with the first two terms in the Zernike decomposition (tip/tilt). [59] The *short-exposure* OTF is given by,

$$\mathbf{H} = \mathbf{H}_{ATM} * \mathbf{H}_{DL} \tag{4.7}$$

where $\mathbf{H}_{DL}$ represents the diffraction limited OTF resulting from the finite size of the aperture and $\mathbf{H}_{ATM}$ represents the effects of the atmosphere and is given by,

$$\mathbf{H}_{ATM} = e^{-3.44\left(\frac{\lambda f_l \rho}{D}\right)^{5/3}\left[1-\left(\frac{\lambda f_l \rho}{D}\right)^{1/3}\right]} \tag{4.8}$$

where $D$ is the pupil diameter, $f_l$ is the focal length, and $\rho$ is the spatial frequency in units of cycles per unit distance. The short-exposure OTF can naturally be transformed to provide an expression for the short-exposure PSF through an inverse Fourier transform,

$$\mathbf{h_{SE}} = \mathcal{F}^{-1}\left\{\mathbf{H}\right\}. \tag{4.9}$$

In this work we extend the concept of the short-exposure PSF to consider the short-exposure PSF covariance, $\mathbf{\Sigma}_{SE}$, which we define as the covariance of the PSF in turbulence discluding the translational effects of the first two Zernike components (tip/tilt). Because no known closed form expression exists for the short-exposure PSF covariance Monte-Carlo methods will be used to obtain it. PSF realizations can be generated based on the theoretical distribution of Zernike coefficients for a given value of $r_0$ and a complete relationship between $r_0$ and the short-exposure PSF covariance can be characterized.

## 4.3  MFBD with Turbulence Characterization

A classical deconvolution approach given a known PSF is to minimize the mean squared error between the collected image data, $\mathbf{d}$ and the data predicted by the convolutional forward model applied to an object estimate $\mathbf{x}$. A regularizing function is typically applied to prevent overfitting to noisy data. A popular choice for the regularizing function is the $\ell_1$ norm of the total variation (TV) of the image estimate which is particularly attractive in this application due to it's known edge preserving capabilities.[75] The resulting objective function is given as,

$$\mathcal{O}\left(\mathbf{x}\right) = \left\|\mathbf{d}_n - \boldsymbol{\mu}\left(\mathbf{x}\right)\right\|^2 + \beta_{TV}\left\|\nabla_{2D}\mathbf{x}\right\|_1. \tag{4.10}$$

where $\boldsymbol{\mu}$ is computed from equation 4.1, $\nabla_{2D}\mathbf{x}$ is a two-dimensional numerical gradient of the object $\mathbf{x}$ and $\beta_{TV}$ is a regularization parameter.

In this application it is our objective to not only estimate the object, but the PSF and the atmospheric coherence length as well. The PSF under turbulence is expressed efficiently in equation 4.3 in terms of the Zernike coefficients. In order to account for the additional unknown quantities, additional constraints must also be applied to avoid convergence to undesirable if not unphysical solutions. It is prudent to constrain the PSF estimates such that the statistics of the collection of PSFs is consistent with expectations for some fixed level of turbulence. We choose to apply constraints to the PSF directly rather than indirectly through the wavefront at the entrance pupil. This choice is made in order to avoid complications caused by the non-unique relationship between the wavefront and the PSF. As such we apply a pair of constraints based on the theoretical PSF statistics associated with the $r_0$ value of the atmosphere. The first constraint requires that the average PSF taken over the collection of frames match the theoretical short exposure PSF under a given $r_0$ condition. The second constraint requires the covariance of the PSF taken over the collection of frames to match the theoretical covariance under a given $r_0$ condition. Letting, $\boldsymbol{\mu}_{psf}$ and $\boldsymbol{\Sigma}_{psf}$ represent the sample mean and sample covariance of the estimated PSFs, the resulting objective

function is given as,

$$\min_{\mathbf{x},\mathbf{a},r_0} \quad \sum_{n=1}^{N} \|\mathbf{d}_n - \boldsymbol{\mu}_n(\mathbf{x},\mathbf{a})\|^2 + \beta_{TV} \|\nabla_{2D}\mathbf{x}\|_1$$

$$\text{s.t.} \quad \boldsymbol{\mu}_{psf}(\mathbf{a}) = \mathbf{h}_{SE}(r_0) \tag{4.11}$$

$$\boldsymbol{\Sigma}_{psf}(\mathbf{a}) = \boldsymbol{\Sigma}_{SE}(r_0)$$

The quantities to be estimated include the object $\mathbf{x}$, the Zernike coefficient vector, $\mathbf{a}$, and the atmospheric coherence length, $r_0$. The constraints are enforced through a pair of quadratic penalty functions,

$$P\left(\boldsymbol{\mu}_{psf}, r_0\right) = \frac{1}{\beta_{\boldsymbol{\mu}}} \left\|\boldsymbol{\mu}_{psf}(\mathbf{a}) - \mathbf{h}_{SE}(r_0)\right\|^2 \tag{4.12}$$

and

$$P\left(\boldsymbol{\Sigma}_{psf}, r_0\right) = \frac{1}{\beta_{\boldsymbol{\Sigma}}} \left\|\boldsymbol{\Sigma}_{psf}(\mathbf{a}) - \boldsymbol{\Sigma}_{SE}(r_0)\right\|^2. \tag{4.13}$$

To enforce the constraints, the parameters $\beta_{\boldsymbol{\mu}}$ and $\beta_{\boldsymbol{\Sigma}}$ are evaluated in the limit as they go towards zero.

## 4.4 Results

The approach is demonstrated both with simulated and experimentally collected data. Simulated results enable us to make direct comparisons between estimated $r_0$ values

and known ground truth while experimental results validate the efficacy of the approach in a real-world setting. The system configuration used both for simulation and experiment is based on an $8\,inch$ aperture, an angular resolution of $5\mu rad$, and a nominal wavelength of $637\,nm$. Sets of 50 consecutively collected frames are used for processing. Solution of the optimization problem is achieved using conventional convex optimization routines based on the $LBFGS$ technique. A brute force search based on the L-curve criterion was used for selection of the appropriate TV regularization parameters. [76]

The simulated collection scenario was devised using a spoke calibration target. Atmospheric conditions characterized by $r_0$ values of $3, 5$, and $7\,cm$ were considered. Higher levels of turbulence indicated by lower $r_0$ values result in more blur in the short-exposure PSF and more uncertainty as indicated in the short-exposure PSF covariance. The PSF statistics specific for the system configuration considered were characterized through Monte-Carlo simulation for the three considered values for $r_0$. The results of the Monte-Carlo simulation are illustrated in Figure 4.1. The plots in Figure 4.1a show slices through the short-exposure PSF. The plots in Figure 4.1b show slices through the standard deviation associated with the diagonal terms of the short-exposure PSF covariance. Note that the PSFs are normalized such that the peak of the diffraction limited PSF is unity. The diffraction limited PSF is included in Figure 4.1a for reference.

**(a)** Short-Exposure PSF      **(b)** Short-Exposure PSF Standard Deviation
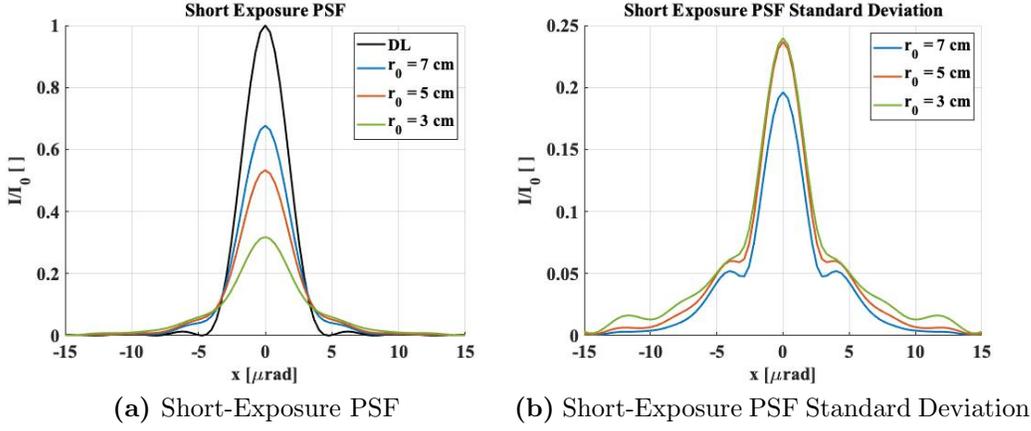
**Figure 4.1:** One-dimensional plots illustrate the Monte-Carlo generated short-exposure PSF ((a)) and short-exposure PSF standard deviation ((b)) under the simulated $r_0$ conditions.

A series of images are shown in figures 4.2, 4.3 and 4.4, illustrating the results applicable to a single frame selected from the multi-frame collections for each of the three turbulence levels considered. The set of images include ((a)) a representative image frame $\mathbf{d}_n$ selected from the multi-frame collections, ((b)) the fit term $\boldsymbol{\mu}_n$ for that image frame resulting from MFBD processing, ((c)) the true PSF used for the simulation of that data frame, and ((d)) the estimated PSF resulting from MFBD processing. Finally, The true object is shown in figure 4.5a along with the reconstructions achieved under the three representative levels of turbulence. The $r_0$ estimates associated with the three reconstructions are $2.98\,cm$, $5.04\,cm$, and $7.12\,cm$ respectively. It should be noted that some deviation from the true $r_0$ values is expected due to the finite number of frames used for processing.

A second demonstration illustrates the efficacy of the approach using experimentally collected data and validated using the PROPS atmospheric measurement system. [63]

66

**(a)** data frame

**(b)** data fit

**(c)** true PSF
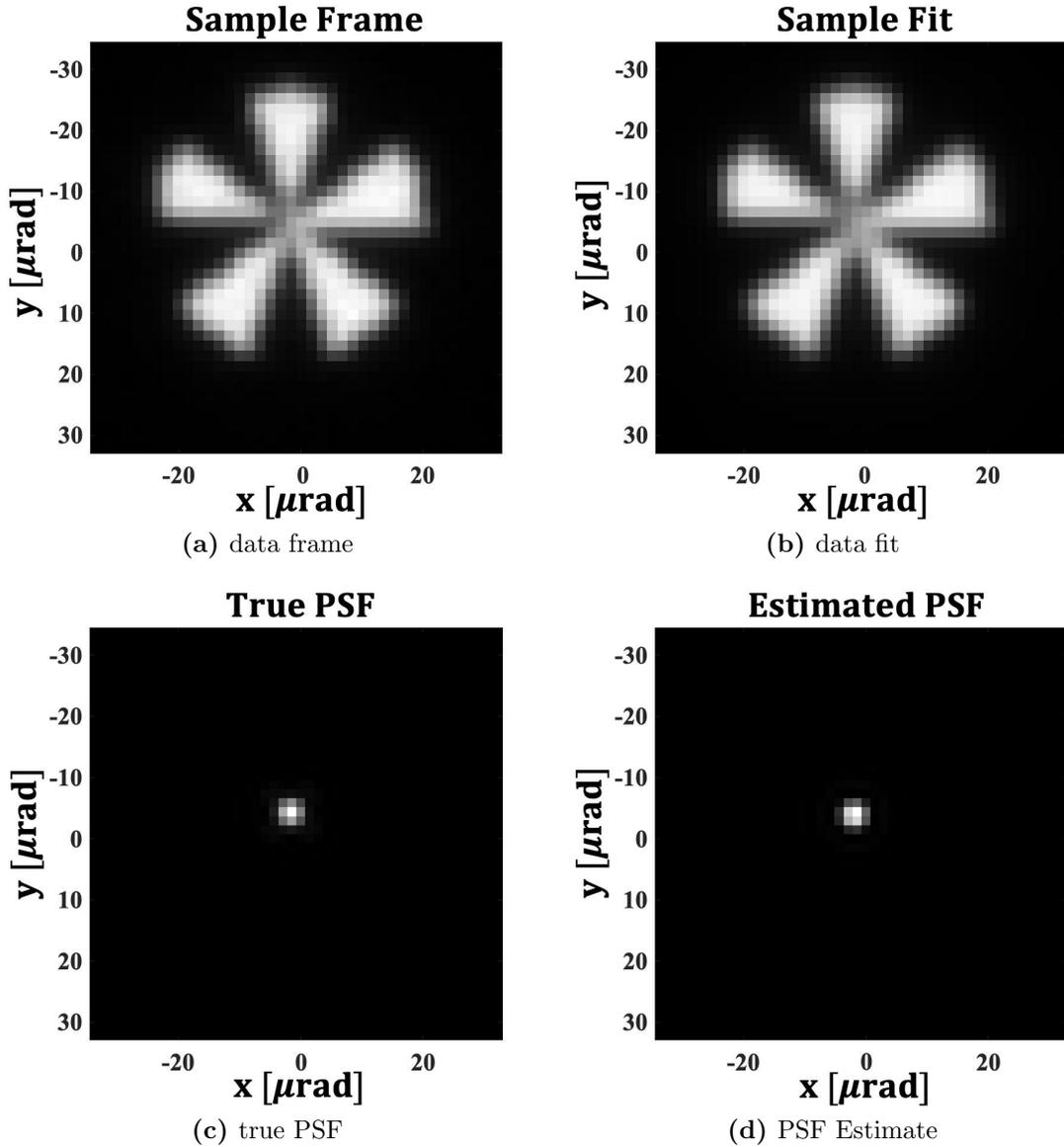
**(d)** PSF Estimate

**Figure 4.2:** A (a) representative image frame $\mathbf{d}_n$, (b) the fit term $\boldsymbol{\mu}_n$ for that image frame resulting from MFBD processing, (c) the true PSF used for the simulation of that data frame, and (d) the estimated PSF resulting from MFBD processing for an $r_0$ value of $7\,cm$.

The PROPS system measurements occurred at $1\,min$ intervals. The relevant PROPS measurements to our 50 frame image set recorded $r_0$ values of $3.09\,cm$ and $2.70\,cm$ with the entirety of the image set collected between the two PROPS measurements.

**(a)** data frame



**(b)** data fit



**(c)** true PSF
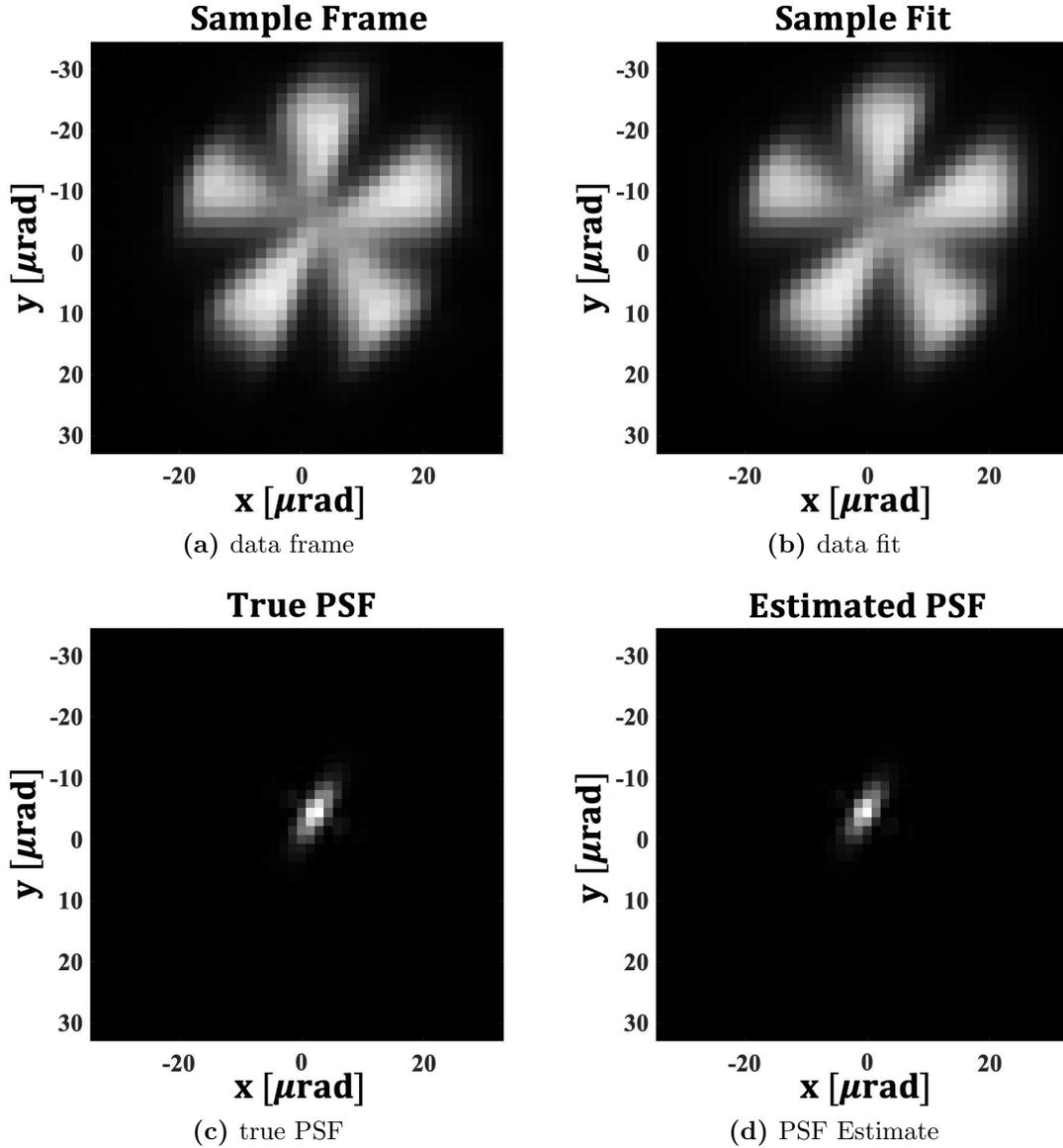


**(d)** PSF Estimate

**Figure 4.3:** A ((a)) representative image frame $\mathbf{d}_n$, ((b)) the fit term $\boldsymbol{\mu}_n$ for that image frame resulting from MFBD processing, ((c)) the true PSF used for the simulation of that data frame, and ((d)) the estimated PSF resulting from MFBD processing for an $r_0$ value of $5\,cm$.

In this experimental collection a target-board was imaged at a range of $1.364\,km$. A full sample frame is shown in 4.6a with a chip identified for MFBD processing. A small chip is chosen both to manage computational burden and to maximize isoplanatism.

68

**(a)** data frame

**(b)** data fit

**(c)** true PSF
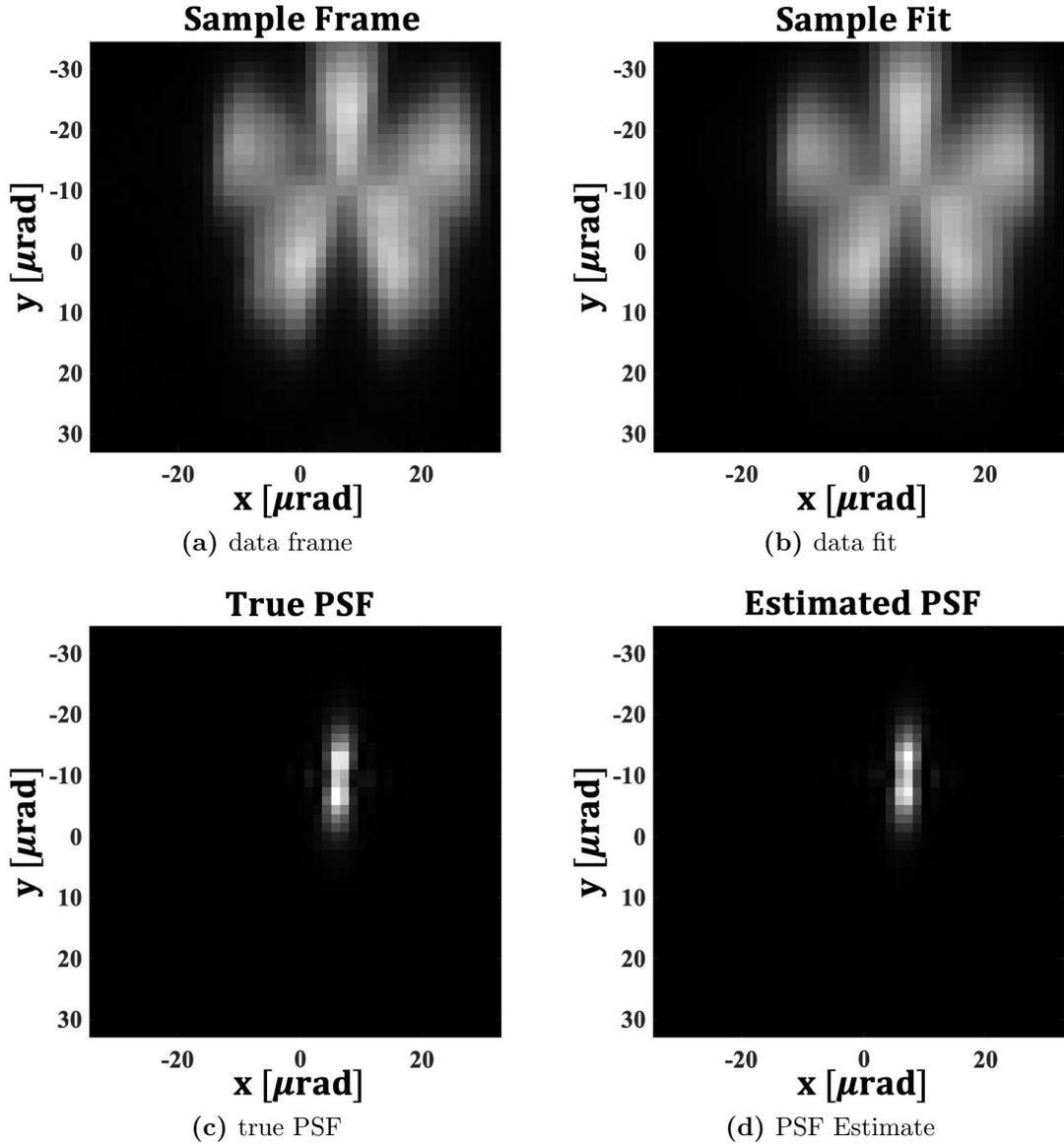
**(d)** PSF Estimate

**Figure 4.4:** A ((a)) representative image frame $\mathbf{d}_n$, ((b)) the fit term $\boldsymbol{\mu}_n$ for that image frame resulting from MFBD processing, ((c)) the true PSF used for the simulation of that data frame, and ((d)) the estimated PSF resulting from MFBD processing for an $r_0$ value of $3\,cm$.

The sample image chip is shown in 4.6b, and finally the MFBD result is shown in 4.6c. The estimated $r_0$ value associated with MFBD processing was found to be $2.67\,cm$ demonstrating excellent agreement with the PROPS measurements.

**Ground Truth**

**(a)** Truth

**Object Estimate**

**(b)** $r_0 = 7\,cm$

**Object Estimate**

**(c)** $r_0 = 5\,cm$

**Object Estimate**
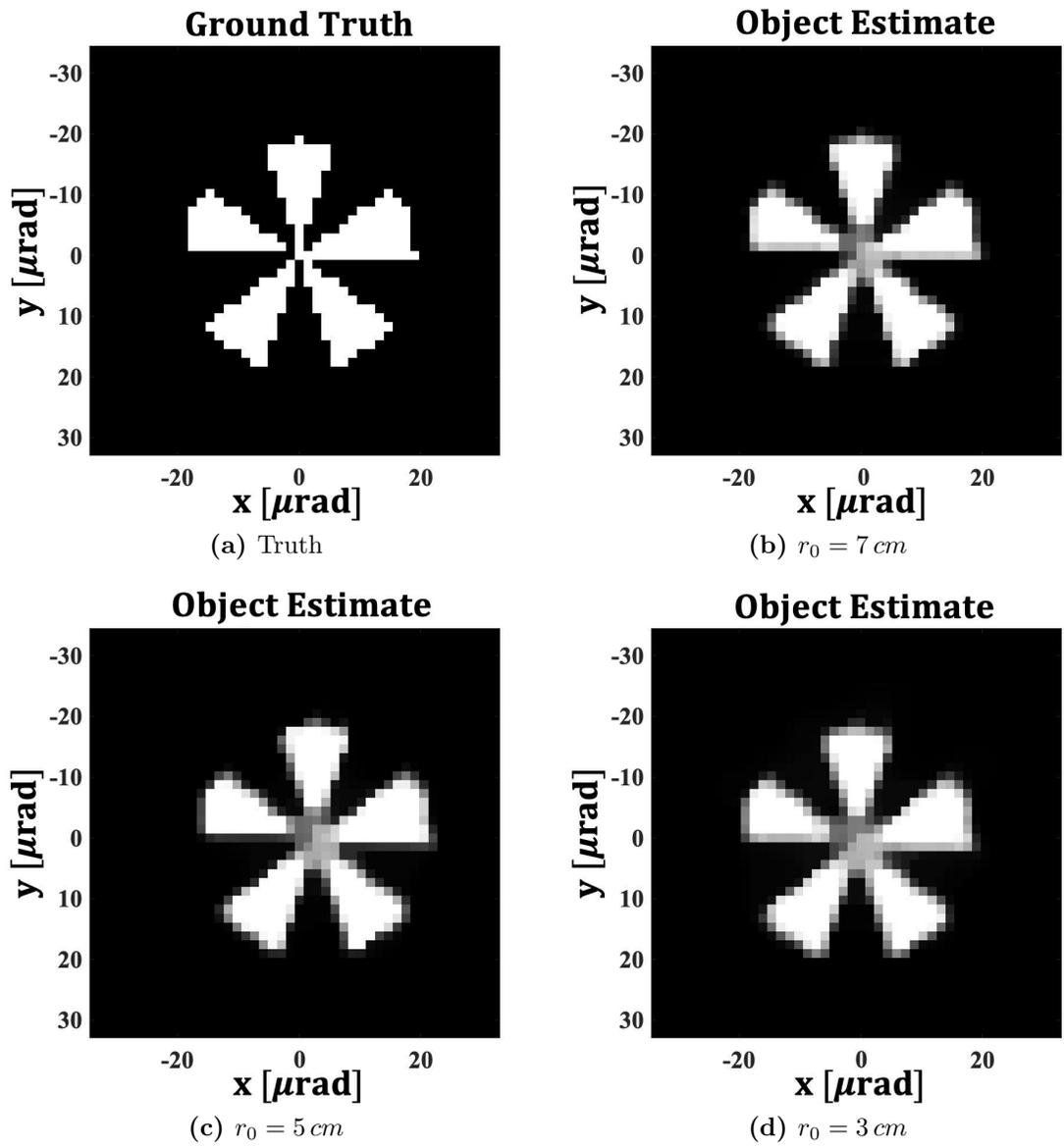
**(d)** $r_0 = 3\,cm$

**Figure 4.5:** The true spoke target ((a)) and the reconstructions under $r_0 = 7\,cm$ ((b)), $r_0 = 5\,cm$ ((c)) , and $r_0 = 3\,cm$ ((d)) respectively.

70

**(a)** Full Frame  **(b)** Image Chip  **(c)** Chip Reconstruction
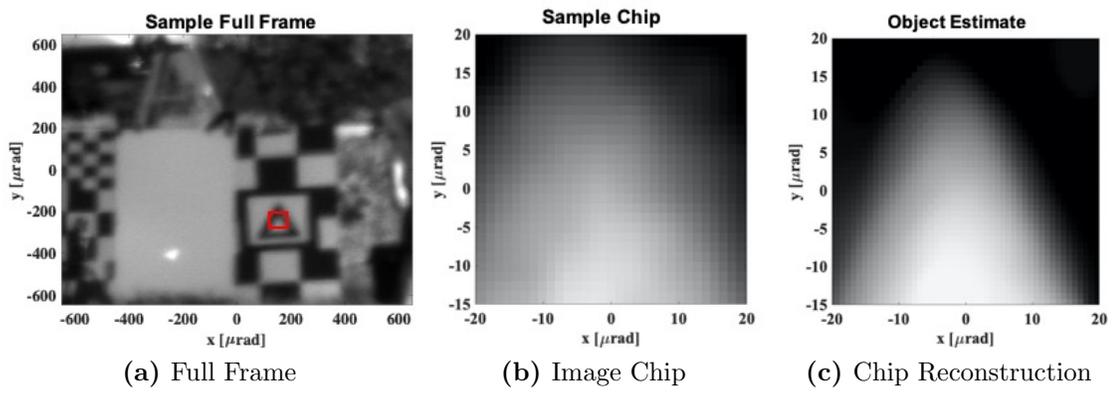
**Figure 4.6:** The chip taken from the full collected frame ((a)) produces a sample chip ((b)), corresponding to the MFBD reconstruction ((c)).

71

# Chapter 5

# Conclusions

In this dissertation novel regularized and constrained least squares methods were adapted and applied to three distinct applications including imaging and detection of buried explosive hazards utilizing array radar; high resolution imaging of satellites in geosynchronous orbit utilizing optical hypertelescope arrays; and characterization of atmospheric turbulence through multi-frame blind deconvolution utilizing conventional optical digital sensors.

In Chapter 2 a least squares image reconstruction technique applicable to FLGPR imaging was developed with two regularization terms designed to balance the expectation that targets should be sparsely populated but extended with respect to the sampled image grid. It was shown that by spatially filtering the collected data through

a combination of Doppler pre-filtering and deterministic image formation processing, significant computational savings can be achieved due to the reduced size of the collected data and also the decreased size of the footprint required for a given ROI. Furthermore, it was shown that image domain reconstruction enables pre-averaging of the collected data which further reduces computational requirements by reducing the number of effective frames used during reconstruction. Results were presented, using experimental data, demonstrating the efficacy of the proposed algorithm.

In Chapter 3 the penalized least squares image reconstruction technique was adapted to sparse aperture imaging based on the fibered hypertelescope concept. Demonstration of the proposed technique was performed using simulated imagery produced based on a postulated hypertelescope design and a notional model of the Galaxy 15 telecommunications satellite. Imaging performance was compared to a conventional implementation of the Richard-Lucy deconvolution algorithm adapted to the fibered hypertelescope paradigm. Reconstructions using the penalized least squares algorithm compared favorably to Richard-Lucy reconstructions based on evaluation using the SSIM score. A thorough analysis of the tuning parameters associated with the penalized least squares technique was conducted. The best results in terms of the SSIM score were obtained by finding a careful balance between the sparsity inducing $\ell_1$ norms associated with the *total energy* and *total variation* penalty terms.

In Chapter 4 a least squares imaging approach was developed to address the important

need to measure and mitigate atmospheric turbulence in optical imaging systems. The method presented in this work avoids complex sensor and illuminator schemes which are typically employed for turbulence measurement and instead characterizes the atmospheric turbulence condition passively through image processing performed on a sequence of collected image frames. By exploiting the theoretical effects of turbulence on the PSF of an optical sensor over time we have constructed a pair of constraints requiring the statistics of the PSF estimates to match expectations for a given level of turbulence. Results demonstrating the efficacy of the approach were generated using both a simulated spoke calibration target and a patch extracted from an experimental data collection. Accurate $r_0$ estimates were shown for levels of turbulence ranging from $r_0 = 7\,cm$ (low) to $r_0 = 3\,cm$ (high). Furthermore, by accounting explicitly for turbulence effects in a physics based manner we have demonstrated excellent image deconvolution performance.

# References

[1] A. J. Webb, M. C. Roggemann, and M. R. Whiteley, "Atmospheric turbulence characterization through multi-frame blind deconvolution," *Appl. Opt.*, Under Review 2021.

[2] A. J. Webb and M. C. Roggemann, "A penalized least squares technique for imaging with hypertelescopes," *Appl. Opt.*, Under Review 2021.

[3] A. J. Webb, T. C. Havens, and T. J. Schulz, "Fast image reconstruction in forward looking gpr using dual l1 regularization," *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 470–478, 2018.

[4] A. J. Webb, T. C. Havens, and T. J. Schulz, "An apodization approach for processing forward-looking gpr for buried target detection," in *Proc. SPIE*, vol. 9454, pp. 94540X–94540X–15, 2015.

[5] A. J. Webb, T. C. Havens, and T. J. Schulz, "Spectral diversity for ground

clutter mitigation in forward-looking gpr," in *Proc. SPIE*, vol. 9823, pp. 98231M–98231M–18, 2016.

[6] A. J. Webb, T. C. Havens, and T. J. Schulz, "Iterative image formation for forward looking gpr," in *Military Sensing Symposium*, vol. 1, March 2016.

[7] A. J. Webb, T. C. Havens, and T. J. Schulz, "Gpr imaging with mutual intensity," in *Proc. SPIE*, vol. 10182, pp. 101821B–101821B–8, 2017.

[8] J. S. Rice, A. Pinar, T. C. Havens, A. J. Webb, and T. J. Schulz, "Multiple instance learning for buried hazard detection," in *Proc. SPIE*, vol. 9823, 2016.

[9] A. J. Pinar, T. C. Havens, and A. Webb, "Multisensor fusion of flgpr and thermal and visible-spectrum cameras for standoff detection of buried objects," in *Proc. SPIE*, vol. 10182, pp. 101821A–101821A–15, 2017.

[10] J. Burns, M. P. Masarik, I. J. Xique, B. Thelen, and A. Webb, "Comparative analysis of image formation techniques for flgpr," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII*, vol. 10182, p. 1018219, International Society for Optics and Photonics, 2017.

[11] R. D. Costley, J. M. Sabatier, and N. Xiang, "Forward-looking acoustic mine detection system," in *Proc. SPIE*, vol. 4394, pp. 617–626, 2001.

[12] F. Cremer, J. G. Chavemaker, W. deJong, and K. Schutte, "Comparison of

vehicle-mounted forward-looking polarimetric infrared and downward-looking infrared sensors for landmine detection," in *Proc. SPIE*, vol. 5089, pp. 517–526, 2003.

[13] N. Playle, D. M. Port, R. Rutherford, I. A. Burch, and R. Almond, "Infrared polarization sensor for forward-looking mine detection," in *Proc. SPIE*, vol. 4742, pp. 11–18, 2002.

[14] T. C. Havens, K. Stone, J. M. Keller, and K. C. Ho, "Sensor-fused detection of explosive hazards," in *Proc. SPIE*, vol. 7303, p. 73032A, 2009.

[15] K. Stone, J. M. Keller, M. Popescu, T. C. Havens, and K. C. Ho, "Forward-looking anomaly detection via fusion of infrared and color imagery," in *Proc. SPIE*, vol. 7664, p. 766425, 2010.

[16] T. C. Havens, C. J. Spain, K. C. Ho, J. M. Keller, T. T. Ton, D. C. Wong, and M. Soumekh, "Improved detection and false alarm rejection using ground-penetrating radar and color imagery in a forward-looking system," in *Proc. SPIE*, vol. 7664, p. 76441U, 2010.

[17] G. Moussally, K. Breiter, and J. Rolig, "Wide-area landmine survey and detection system," in *Proceedings of the Tenth International Conference on Grounds Penetrating Radar, 2004. GPR 2004.*, pp. 693–696, June 2004.

[18] T. C. Havens, K. C. Ho, J. Farrell, J. M. Keller, M. Popescu, T. T. Ton, D. C.

Wong, and M. Soumekh, "Locally adaptive detection algorithm for forward-looking ground-penetrating radar," in *Proc. SPIE*, vol. 7664, p. 76442E, 2010.

[19] J. Farrell, T. C. Havens, K. C. Ho, J. M. Keller, T. T. Ton, D. C. Wong, and M. Soumekh, "Detection of explosive hazards using spectrum features from forward-looking ground penetrating radar imagery," in *Proc. SPIE*, vol. 8017, p. 80171F, 2011.

[20] J. Farrell, T. C. Havens, K. C. Ho, J. M. Keller, T. T. Ton, D. C. Wong, and M. Soumekh, "Evaluation and improvement of spectral features for the detection of buried explosive hazards using forward-looking ground-penetrating radar," in *Proc. SPIE*, vol. 8357, p. 8357C, 2012.

[21] T. C. Havens, J. M. Keller, K. C. Ho, T. T. Ton, D. C. Wong, and M. Soumekh, "Narrow band processing and fusion approach for explosive hazard detection in FLGPR," in *Proc. SPIE*, vol. 8017, p. 80171F, 2011.

[22] T. C. Havens, J. M. Keller, K. Stone, K. C. Ho, T. T. Ton, D. C. Wong, and M. Soumekh, "Multiple kernel learning for explosive hazards detection in FLGPR," in *Proc. SPIE*, vol. 8357, p. 83571D, 2012.

[23] T. C. Havens, J. T. Becker, A. J. Pinar, and T. J. Schulz, "Multi-band sensor-fused explosive hazards detection in forward-looking ground-penetrating radar," in *Proc. SPIE*, vol. 9072, p. 90720T, 2014.

[24] Y. Fuse, B. Gonzalez-Valdes, J. A. Martinez-Lorenzo, and C. M. Rappaport, "Advanced sar imaging methods for forward-looking ground penetrating radar," in *2016 10th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–4, April 2016.

[25] M. R. Bradley, T. R. Witten, M. Duncan, and R. McCummins, "Anti-tank and side-attach mine detection with forward-looking GPR," in *Proc. SPIE*, vol. 5415, pp. 421–432, 2004.

[26] R. Cosgrove, P. Milanfar, and J. Kositsky, "Trained detection of buried mines in sar images via the deflection-optimal criterion," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, pp. 2569 – 2575, Nov. 2004.

[27] Y. Sun and J. Li, "Plastic landmine detection using time-frequency analysis for forward-looking ground-penetrating radar," in *Proc. SPIE*, vol. 5089, pp. 851–862, 2003.

[28] L. Nyguyen, "Signal processing techniques for stepped frequency ultrawideband radar," in *SPIE Defense+ Security*, pp. 90771H–90771H, International Society for Optics and Photonics, 2014.

[29] D. Comite, F. Ahmad, D. Liao, T. Dogaru, and M. G. Amin, "Multiview imaging for low-signature target detection in rough-surface clutter environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 5220–5229, Sept 2017.

[30] D. Comite, F. Ahmad, T. Dogaru, and M. G. Amin, "Coherence factor for rough surface clutter mitigation in forward-looking gpr," in *2017 IEEE Radar Conference (RadarConf)*, pp. 1803–1806, May 2017.

[31] L. Nguyen and R. Innocenti, "Suppression of sidelobes and noise in airborne SAR imagery using recursive sidelobe minimization technique," in *Proc. IEEE Radar Conference*, pp. 522–525, May 2010.

[32] M. A. González-Huici, I. Catapano, and F. Soldovieri, "A comparative study of gpr reconstruction approaches for landmine detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 4869–4878, Dec 2014.

[33] F. Soldovieri, G. Gennarelli, I. Catapano, D. Liao, and T. Dogaru, "Forward-looking radar imaging: A comparison of two data processing strategies," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, pp. 562–571, Feb 2017.

[34] D. Comite, A. Galli, I. Catapano, and F. Soldovieri, "Advanced imaging for down-looking contactless gpr systems," in *2017 International Applied Computational Electromagnetics Society Symposium - Italy (ACES)*, pp. 1–2, March 2017.

[35] I. Catapano, A. Affinito, A. D. Moro, G. Alli, and F. Soldovieri, "Forward-looking ground-penetrating radar via a linear inverse scattering approach," *IEEE*

*Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 5624–5633, Oct 2015.

[36] J. Yang, T. Jin, X. Huang, J. Thompson, and Z. Zhou, "Sparse mimo array forward-looking gpr imaging based on compressed sensing in clutter environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 4480–4494, July 2014.

[37] B. Burns, "Forward looking GPR sidelobe reduction using l1-norm minimization," in *Proc. SPIE*, vol. 8357, p. 835713, 2012.

[38] M. Ndoye, J. M. M. Anderson, and D. J. Greene, "An mm-based algorithm for $\ell_1$-regularized least-squares estimation with an application to ground penetrating radar image reconstruction," *IEEE Transactions on Image Processing*, vol. 25, pp. 2206–2221, May 2016.

[39] H. C. Ogworonjo, J. M. M. Anderson, M. Ndoye, and L. Nguyen, "An l1-regularized least squares algorithm for reconstructing step-frequency ground penetrating radar images," in *2016 IEEE Radar Conference (RadarConf)*, pp. 1–5, May 2016.

[40] H. C. Ogworonjo, J. M. M. Anderson, and L. H. Nguyen, "An iterative parameter-free map algorithm with an application to forward looking gpr imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 1573–1586, March 2017.

[41] M. Cetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Transactions on Image Processing*, vol. 10, pp. 623–631, April 2001.

[42] L. G. Wang, L. Li, J. Ding, and T. J. Cui, "A fast patches-based imaging algorithm for 3-d multistatic imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, pp. 941–945, June 2017.

[43] M. Richards, *Fundamentals Of Radar Signal Processing*. McGraw-Hill Education (India) Pvt Limited, 2005.

[44] J. W. Goodman, *Statistical optics*. John Wiley & Sons, 2015.

[45] A. Labeyrie, "Resolved imaging of extra-solar planets with future 10-100 km optical interferometric arrays," *Astronomy and Astrophysics Supplement Series*, vol. 118, no. 3, pp. 517–524, 1996.

[46] O. Lardière, F. Martinache, and F. Patru, "Direct imaging with highly diluted apertures–i. field-of-view limitations," *Monthly Notices of the Royal Astronomical Society*, vol. 375, no. 3, pp. 977–988, 2007.

[47] F. Patru, D. Mourard, O. Lardière, and S. Lagarde, "Optimization of the direct imaging properties of an optical-fibred long baseline interferometer," *Monthly Notices of the Royal Astronomical Society*, vol. 376, pp. 1047–1053, January 2007.

[48] A. Labeyrie, D. Mourard, F. Allouche, R. Chakraborthy, J. Dejonghe, A. Surya, Y. Bresson, C. Aime, D. Mary, and A. Carlotti, "Concept study of an extremely large hyper telescope (elhyt) with 1200m sparse aperture for direct imaging at 100 micro-arcsecond resolution," in *Optical and Infrared Interferometry III*, vol. 8445, p. 844512, International Society for Optics and Photonics, 2012.

[49] F. Patru, N. Tarmoul, D. Mourard, and O. Lardière, "Direct imaging with highly diluted apertures–ii. properties of the point spread function of a hypertelescope," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 4, pp. 2363–2372, 2009.

[50] Éric Thiébaut and J. Young, "Principles of image reconstruction in optical interferometry: tutorial," *J. Opt. Soc. Am. A*, vol. 34, pp. 904–923, Jun 2017.

[51] C. Aime, H. Lantéri, M. Diet, and A. Carlotti, "Strategies for the deconvolution of hypertelescope images," *Astronomy & Astrophysics*, vol. 543, p. A42, 2012.

[52] A. Surya, S. K. Saha, and A. Labeyrie, "Speckle imaging with hypertelescopes," *Monthly Notices of the Royal Astronomical Society*, vol. 443, pp. 852–859, January 2014.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[54] E. Hecht, *Optics*. Pearson education, Addison-Wesley, 2002.

[55] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Optical Society of America*, vol. 62, no. 1, pp. 55–59, 1972.

[56] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical J.*, vol. 79, no. 6, pp. 745–754, 1974.

[57] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. CRC Press, 1998.

[58] P. North and N. Kazmi, "Shining a light on visual magnitude," tech. rep., AGI, 2019.

[59] M. C. Roggemann and B. M. Welsh, *Imaging through turbulence*. CRC press, 1996.

[60] L. C. Andrews and R. L. Phillips, *Laser Beam Propagation Through the Atmosphere, Second Edition*. SPIE Press, 2005.

[61] M. Sarazin and F. Roddier, "The ESO differential image motion monitor," *aap*, vol. 227, pp. 294–300, Jan. 1990.

[62] D. Coburn, D. Garnier, and J. Dainty, "A single star SCIDAR system for profiling atmospheric turbulence," in *Optics in Atmospheric Propagation and Adaptive Systems VIII* (K. Stein and A. Kohnle, eds.), vol. 5981, pp. 105 – 114, International Society for Optics and Photonics, SPIE, 2005.

[63] *Path-Resolved Optical Profiler System (PROPS), PR-05-600.* https://www.mza.com/doc/misc/MZA_PROPS_PR-05-600_Specifications.pdf.

[64] M. R. Whiteley, D. C. Washburn, and L. A. Wright, "Differential-tilt technique for saturation-resistant profiling of atmospheric turbulence," in *Adaptive Optics Systems and Technology II* (R. K. Tyson, D. Bonaccini, and M. C. Roggemann, eds.), vol. 4494, pp. 221 – 232, International Society for Optics and Photonics, SPIE, 2002.

[65] *DELTA Imaging Path Turbulence Monitor, PM-02-600.* https://www.mza.com/doc/misc/MZA_DELTA_PM-02-600_Specifications.pdf.

[66] G. E. Archer, J. P. Bos, and M. C. Roggemann, "Comparison of bispectrum, multiframe blind deconvolution and hybrid bispectrum-multiframe blind deconvolution image reconstruction techniques for anisoplanatic, long horizontal-path imaging," *Optical Engineering*, vol. 53, no. 4, pp. 1 – 16, 2014.

[67] T. J. Schulz, "Multiframe blind deconvolution of astronomical images," *JOSA A*, vol. 10, no. 5, pp. 1064–1073, 1993.

[68] A. J. Lambert, D. Fraser, M. R. S. Jahromi, and B. R. Hunt, "Superresolution in image restoration of wide-area images viewed through atmospheric turbulence," in *Image Reconstruction from Incomplete Data II* (P. J. Bones, M. A. Fiddy, and R. P. Millane, eds.), vol. 4792, pp. 35 – 43, International Society for Optics and Photonics, SPIE, 2002.

[69] R. C. Hardie, M. A. Rucci, B. K. Karch, A. J. Dapore, and D. R. Droege, "Super-resolution in the presence of atmospheric optical turbulence," in *Long-Range Imaging III* (E. J. Kelmelis, ed.), vol. 10650, pp. 123 – 136, International Society for Optics and Photonics, SPIE, 2018.

[70] X. Zhu and P. Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 157–170, 2013.

[71] M. Furhad, M. Tahtali, and A. Lambert, "Restoring atmospheric-turbulence-degraded images," *Applied Optics*, vol. 55, p. 5082, July 2016.

[72] M. A. Rucci, R. C. Hardie, and A. J. Dapore, "Comparing multiple turbulence restoration algorithms performance on noisy anisoplanatic imagery," in *Long-Range Imaging II* (E. J. Kelmelis, ed.), vol. 10204, pp. 34 – 39, International Society for Optics and Photonics, SPIE, 2017.

[73] S. R. Bose-Pillai, J. E. M. Jr., R. A. Wood, C. E. . Murphy, C. A. Rice, and S. T. Fiorino, "Characterizing atmospheric turbulence over long paths using time-lapse imagery," in *Long-Range Imaging III* (E. J. Kelmelis, ed.), vol. 10650, pp. 58 – 67, International Society for Optics and Photonics, SPIE, 2018.

[74] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *JOsA*, vol. 66, no. 3, pp. 207–211, 1976.

[75] 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.

[76] P. Hansen, *The L-Curve and Its Use in the Numerical Treatment of Inverse Problems*, vol. 4, pp. 119–142. 01 2001.

# Appendix A

# Mathematical Background

This dissertation presents penalized and constrained least squares computational imaging methods applied to three unique remote sensing modalities. In this appendix the Bayesian interpretation for least squares and penalized least squares is described in the context of image reconstruction.

In section A.1 least squares is derived as the maximum likelihood estimator (MLE) for a Normally distributed random vector with constant diagonal covariance. In section A.2 penalized least squares (PLS) is derived as a maximum a posteriori (MAP) estimator and where penalty functions are based on $\ell_1$ norms. In each section the necessary gradients are derived for use in numerical optimization.

## A.1   Probabilistic Interpretation of Least Squares

It is common for measured data to be modeled as a normally distributed random vector. The familiar *pdf* associated with a Normally distributed random vector, $\mathbf{d}$ consisting of $N_d$ elements, is given by,

$$f\left(\mathbf{d}|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \frac{1}{\sqrt{\left(2\pi\right)^{N_d}|\boldsymbol{\Sigma}|}}e^{-1/2(\mathbf{d}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{d}-\boldsymbol{\mu})} \tag{A.1}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance respectively. If multiple independent observations are available, such as a multi-frame collection of a common scene, then the joint distribution associated with the collection of observations is given by

$$\begin{aligned} f\left(\mathbf{d_1}, \mathbf{d_2}, ..., \mathbf{d_{N_f}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}_n\right) &= \prod_{n=1}^{N_f} \frac{1}{\sqrt{\left(2\pi\right)^{N_d}|\boldsymbol{\Sigma}_n|}}e^{-1/2(\mathbf{d_n}-\boldsymbol{\mu}_n)^T\boldsymbol{\Sigma}_n^{-1}(\mathbf{d_n}-\boldsymbol{\mu}_n)} \\ &= \left(\frac{1}{\sqrt{\left(2\pi\right)^{N_d}|\boldsymbol{\Sigma}_n|}}\right)^{N_f} e^{-1/2\sum_{n=1}^{N_f}(\mathbf{d_n}-\boldsymbol{\mu}_n)^T\boldsymbol{\Sigma}_n^{-1}(\mathbf{d_n}-\boldsymbol{\mu}_n)}, \end{aligned} \tag{A.2}$$

where $N_f$ is the number of independent observations.

The log-likelihood equation associated with such a distribution is then given by,

$$\mathcal{L}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{d_1}, \mathbf{d_2}, ..., \mathbf{d_{N_f}}\right) = -\frac{N_f N_d}{2} \log(2\pi) - \frac{1}{2} \sum_{n=1}^{N_f} \log\left(|\boldsymbol{\Sigma_n}|\right) \qquad \text{(A.3)}$$

$$-\frac{1}{2} \sum_{n=1}^{N_f} (\mathbf{d_n} - \boldsymbol{\mu}_n)^T \boldsymbol{\Sigma_n}^{-1} (\mathbf{d_n} - \boldsymbol{\mu}_n).$$

## A.1.1   Constant Diagonal Covariance Assumption

A common assumption simplifying the likelihood expression is that the elements in the random vector, $\mathbf{d}$ are uncorrelated and have a common variance which remains fixed over the set of observations. In this case, the covariance reduces to a constant diagonal matrix, $\boldsymbol{\Sigma}_n = \sigma^2 \mathbf{I} \quad \forall n$, and the determinant and inverse can be computed trivially by recalling that the determinant of a diagonal matrix is the product of the diagonal terms and the inverse of a diagonal matrix is a diagonal matrix with the elements inverted. As such,

$$|\boldsymbol{\Sigma}_n| = \sigma^{2N_d}, \qquad \text{and} \qquad \boldsymbol{\Sigma}_n^{-1} = \frac{1}{\sigma^2} \mathbf{I}. \qquad \text{(A.4)}$$

The log-likelihood expression associated with such a set of independent random vectors can then be reduced as follows

$$\mathcal{L}\left(\boldsymbol{\mu}, \sigma | \mathbf{x}\right) = -\frac{N_d N_f}{2} \log\left(2\pi\right) - N_f N_d \log\left(\sigma\right) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (\mathbf{d_n} - \boldsymbol{\mu}_n)^T (\mathbf{d_n} - \boldsymbol{\mu}_n) \quad \text{(A.5)}$$

$$= -\frac{N_d N_f}{2} \log\left(2\pi\right) - N_f N_d \log\left(\sigma\right) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} \|\mathbf{d_n} - \boldsymbol{\mu}_n\|_2^2 .$$

Often the expected value of the data is directly dependent on the object of interest, $\mathbf{g}$, and as such, $\boldsymbol{\mu}_n \Rightarrow \boldsymbol{\mu}_n(\mathbf{g})$. If only estimation of the object is required then the terms that are independent of $\mathbf{g}$ are unimportant and the least squares estimator is equivalent to the MLE estimator,

$$\mathcal{O}\left(\mathbf{g}\right) = \sum_{n=1}^{N} \|\mathbf{d_n} - \boldsymbol{\mu}_n\left(\mathbf{g}\right)\|_2^2 . \quad \text{(A.6)}$$

If $\boldsymbol{\mu}$ is linearly dependent on $\mathbf{g}$ as is often the case, then the relationship can be expressed as a matrix vector product such that $\boldsymbol{\mu} = \mathbf{A}\mathbf{g}$

## A.1.2 Least Squares Gradient

The gradient of the objective function is required as part of the optimization process. The expression for the gradient is given here for the least squares data fit term in

terms of $\boldsymbol{\mu}_n(\mathbf{g})$ as

$$\frac{\partial \mathcal{O}}{\partial \mathbf{g}} = \sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{g}} \|\mathbf{d_n} - \boldsymbol{\mu}_n(\mathbf{g})\|_2^2 \tag{A.7}$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{g}} (\mathbf{d_n} - \boldsymbol{\mu}_n(\mathbf{g}))^T (\mathbf{d_n} - \boldsymbol{\mu}_n(\mathbf{g}))$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{g}} \left( \mathbf{d_n}^T \mathbf{d_n} - \boldsymbol{\mu}_n(\mathbf{g})^T \mathbf{d_n} - \mathbf{d_n}^T \boldsymbol{\mu}_n(\mathbf{g}) + \boldsymbol{\mu}_n(\mathbf{g})^T \boldsymbol{\mu}_n(\mathbf{g}) \right)$$

$$= 2 \sum_{n=1}^{N} \frac{\partial \boldsymbol{\mu}_n(\mathbf{g})}{\partial \mathbf{g}} (\boldsymbol{\mu}_n(\mathbf{g}) - \mathbf{d_n}).$$

The gradient with respect to $\boldsymbol{\mu}(\mathbf{g})$ then is necessary and must be computed based on the signal model associated with the measured data. For the linear case, $\frac{\partial \boldsymbol{\mu}_n(\mathbf{g})}{\partial \mathbf{g}} = \mathbf{A}^T$ and as such,

$$\frac{\partial \mathcal{O}}{\partial \mathbf{g}} = 2 \sum_{n=1}^{N} \mathbf{A}^T (\mathbf{Ag} - \mathbf{d_n}). \tag{A.8}$$

## A.2 Probabilistic Interpretation of Penalized Least Squares

A probabilistic interpretation of the penalized least squares objective function is described as follows; Let $\mathbf{g}$ represent the object and let the two-dimensional numerical gradient of the object be given as $\mathbf{v} = \nabla_{2D}\mathbf{g}$. From Bayes Rule, the joint probability

95

of $\mathbf{g}$ and $\mathbf{v}$ given data $\mathbf{d}$ can be expanded as

$$p\left(\mathbf{g}, \mathbf{v} | \mathbf{d}\right) = \frac{p\left(\mathbf{d} | \mathbf{g}, \mathbf{v}\right) p\left(\mathbf{g}, \mathbf{v}\right)}{p\left(\mathbf{d}\right)}. \tag{A.9}$$

The relationship between $\mathbf{g}$ and its numerical gradient, $\mathbf{v}$, is completely deterministic and as such $p\left(\mathbf{d} | \mathbf{g}, \mathbf{v}\right) = p\left(\mathbf{d} | \mathbf{g}\right)$ and through the same reasoning $p\left(\mathbf{g}, \mathbf{v} | \mathbf{d}\right) = p\left(\mathbf{g} | \mathbf{d}\right) p\left(\mathbf{v} | \mathbf{g}, \mathbf{d}\right) = p\left(\mathbf{g} | \mathbf{d}\right)$. In contrast, in order to simplify the joint probability distribution, $p\left(\mathbf{g}, \mathbf{v}\right)$ the approximation is made that

$$p\left(\mathbf{g}, \mathbf{v}\right) = p\left(\mathbf{v}\right) p\left(\mathbf{g} | \mathbf{v}\right) \tag{A.10}$$

$$= p\left(\mathbf{v}\right) p\left(\mathbf{g}\right),$$

such that equation A.9 can now be expressed as

$$p\left(\mathbf{g} | \mathbf{d}\right) = \frac{p\left(\mathbf{d} | \mathbf{g}\right) p\left(\mathbf{v}\right) p\left(\mathbf{g}\right)}{p\left(\mathbf{d}\right)}. \tag{A.11}$$

In order to obtain the $\ell_1$ norm penalty function it is assumed that the distributions associated with the object $\mathbf{g}$ and it's numerical gradient $\mathbf{v}$ are modeled as *iid* Laplacians such that,

$$p\left(\mathbf{v}\right) = \prod_{i=1}^{N} C_v e^{-\beta_{TV} |v_i|}, \tag{A.12}$$

and

$$p\left(\mathbf{g}\right) = \prod_{i=1}^{N} C_e e^{-\beta_{TE} |g_i|}. \tag{A.13}$$

The denominator is unrelated to $\mathbf{g}$ and as such plays no role in the optimization. If the data term $p(\mathbf{d}|\mathbf{g})$ is reduced to the squared error as was demonstrated in section A.1 then the Log-Likelihood associated with equation A.11 reduces to

$$\mathcal{L}(\mathbf{g}|\mathbf{d}) = \|\mathbf{d} - \boldsymbol{\mu}(\mathbf{g})\|^2 + \beta_{TE}\|\mathbf{g}\|_1 + \beta_{TV}\|\mathbf{v}\|_1, \tag{A.14}$$

and if multiple independent observations are made then the expression can be extended as in section A.1 such that

$$\mathcal{L}(\mathbf{g}|\mathbf{d}) = \sum_{n=1}^{N_f} \|\mathbf{d}_n - \boldsymbol{\mu}_n(\mathbf{g})\|^2 + \beta_{TE}\|\mathbf{g}\|_1 + \beta_{TV}\|\mathbf{v}\|_1. \tag{A.15}$$

### A.2.1  Total Energy Penalty Gradient

In this section the gradient associated with the total energy penalty term is derived. The penalty function associated with total energy regularization is given by

$$P = \beta_{TE}\|\mathbf{g}\|_p^p, \tag{A.16}$$

where the $L_p$ norm taken over the entire set of pixels is defined as

$$\|\mathbf{g}\|_p^p = \sum_{k=1}^{K} |g_k|^p \tag{A.17}$$

From this the gradient for any $p$ can be given trivially,

$$\frac{\partial \|\mathbf{g}\|_p^p}{\partial \mathbf{g}} = p \, |\mathbf{g}|^{p-1}, \tag{A.18}$$

and for the special case of $p = 1$, the gradient reduces to the ones vector such that $\frac{\partial \|\mathbf{g}\|_1}{\partial \mathbf{g}} = \mathbf{1}$.

## A.2.2 Total Variation Penalty Gradient

In this section the gradient associated with the total variation penalty term is derived. If we consider the finite difference in the $x$ and $y$ directions computed using the appropriate finite difference matrices, $\mathbf{D}_x$ and $\mathbf{D}_y$, then the finite difference approximation to the gradient vector is given as

$$\nabla \mathbf{g} = \begin{pmatrix} \mathbf{D}_x \mathbf{g} \\ \mathbf{D}_y \mathbf{g} \end{pmatrix}, \tag{A.19}$$

and the magnitude (2-norm) of the gradient vector for each image sample is then approximated by

$$\boldsymbol{\gamma} = |\nabla \mathbf{g}| = \sqrt{(\mathbf{D}_x \mathbf{g})^2 + (\mathbf{D}_y \mathbf{g})^2}. \tag{A.20}$$

It will prove convenient to further define

$$\boldsymbol{\xi}\left(\mathbf{g}\right) = \left(\mathbf{D}_x\mathbf{g}\right)^2 + \left(\mathbf{D}_y\mathbf{g}\right)^2,\tag{A.21}$$

such that

$$\boldsymbol{\gamma} = \sqrt{\boldsymbol{\xi}\left(\mathbf{g}\right)}.\tag{A.22}$$

Finally, the $L_p$ norm taken over the entire set of pixels is

$$\|\boldsymbol{\gamma}\|_p^p = \sum_{k=1}^{K} \gamma_k^p.\tag{A.23}$$

Through the chain rule we compute the gradient vector w.r.t. $\mathbf{g}$,

$$\frac{\partial}{\partial \mathbf{g}}\left(\||\nabla\mathbf{g}|\|_p^p\right) = \left(\frac{\partial \boldsymbol{\xi}}{\partial \mathbf{g}}\right)\left(\frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\xi}}\right)\left(\frac{\|\boldsymbol{\gamma}\|_p^p}{\partial \boldsymbol{\gamma}}\right).\tag{A.24}$$

Evaluation of each of the terms results in,

$$\frac{\partial \boldsymbol{\xi}}{\partial \mathbf{g}} = 2\left(\operatorname{diag}[\mathbf{D}_x\mathbf{g}]\,\mathbf{D}_x + \operatorname{diag}[\mathbf{D}_y\mathbf{g}]\,\mathbf{D}_y\right)^T,\tag{A.25}$$

$$\frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\xi}} = \frac{1}{2}\operatorname{diag}\left[\frac{1}{\sqrt{\boldsymbol{\xi}}}\right],\tag{A.26}$$

$$\frac{\|\boldsymbol{\gamma}\|_p^p}{\partial \boldsymbol{\gamma}} = p\,|\boldsymbol{\gamma}| \odot \left(|\boldsymbol{\gamma}|^2 + \epsilon\right)^{p/2-1}.\tag{A.27}$$

such that the final gradient of the T.V. penalty can be reduced as follows

$$\frac{\partial}{\partial \mathbf{g}}\left(\|\boldsymbol{\gamma}\|_p^p\right) = 2\left(\mathrm{diag}[\mathbf{D}_x\mathbf{g}]\,\mathbf{D}_x + \mathrm{diag}[\mathbf{D}_y\mathbf{g}]\,\mathbf{D}_y\right)^T \tag{A.28}$$

$$\frac{1}{2}\mathrm{diag}\left[\frac{1}{\sqrt{\boldsymbol{\xi}}}\right] p\,|\boldsymbol{\gamma}| \odot \left(|\boldsymbol{\gamma}|^2 + \epsilon\right)^{p/2-1}$$

$$= p\left(\mathrm{diag}[\mathbf{D}_x\mathbf{g}]\,\mathbf{D}_x + \mathrm{diag}[\mathbf{D}_y\mathbf{g}]\,\mathbf{D}_y\right)^T$$

$$\mathrm{diag}\left[\frac{1}{\boldsymbol{\gamma}}\right]|\boldsymbol{\gamma}| \odot \left(|\boldsymbol{\gamma}|^2 + \epsilon\right)^{p/2-1}$$

$$= p\left(\mathrm{diag}[\mathbf{D}_x\mathbf{g}]\,\mathbf{D}_x + \mathrm{diag}[\mathbf{D}_y\mathbf{g}]\,\mathbf{D}_y\right)^T \left(|\boldsymbol{\gamma}|^2 + \epsilon\right)^{p/2-1}$$

$$= p\left(\mathbf{D}_x^T\left(\mathbf{D}_x\mathbf{g} \odot \left(|\boldsymbol{\gamma}|^2 + \epsilon\right)^{p/2-1}\right) + \mathbf{D}_y^T\left(\mathbf{D}_y\mathbf{g} \odot \left(|\boldsymbol{\gamma}|^2 + \epsilon\right)^{p/2-1}\right)\right).$$