ABSTRACT

Title of dissertation: Sparse and Redundant Representations
                           for Inverse Problems and Recognition

                           Vishal M. Patel, Doctor of Philosophy, 2010

Dissertation directed by: Professor Rama Chellappa
                           Department of Electrical and Computer Engineering

Sparse and redundant representation of data enables the description of signals as linear combinations of a few atoms from a dictionary. In this dissertation, we study applications of sparse and redundant representations in inverse problems and object recognition. Furthermore, we propose two novel imaging modalities based on the recently introduced theory of Compressed Sensing (CS).

This dissertation consists of four major parts. In the first part of the dissertation, we study a new type of deconvolution algorithm that is based on estimating the image from a shearlet decomposition. Shearlets provide a multi-directional and multi-scale decomposition that has been mathematically shown to represent distributed discontinuities such as edges better than traditional wavelets. We develop a deconvolution algorithm that allows for the approximation inversion operator to be controlled on a multi-scale and multi-directional basis. Furthermore, we develop a method for the automatic determination of the threshold values for the noise shrinkage for each scale and direction without explicit knowledge of the noise variance using a generalized cross validation method.

In the second part of the dissertation, we study a reconstruction method that recovers highly undersampled images assumed to have a sparse representation in a gradient domain by using partial measurement samples that are collected in the Fourier domain. Our method makes use of a robust generalized Poisson solver that greatly aids in achieving a significantly improved performance over similar proposed

methods. We will demonstrate by experiments that this new technique is more flexible to work with either random or restricted sampling scenarios better than its competitors.

In the third part of the dissertation, we introduce a novel Synthetic Aperture Radar (SAR) imaging modality which can provide a high resolution map of the spatial distribution of targets and terrain using a significantly reduced number of needed transmitted and/or received electromagnetic waveforms. We demonstrate that this new imaging scheme, requires no new hardware components and allows the aperture to be compressed. Also, it presents many new applications and advantages which include strong resistance to countermesasures and interception, imaging much wider swaths and reduced on-board storage requirements.

The last part of the dissertation deals with object recognition based on learning dictionaries for simultaneous sparse signal approximations and feature extraction. A dictionary is learned for each object class based on given training examples which minimize the representation error with a sparseness constraint. A novel test image is then projected onto the span of the atoms in each learned dictionary. The residual vectors along with the coefficients are then used for recognition. Applications to illumination robust face recognition and automatic target recognition are presented.

Sparse and Redundant Representations for Inverse Problems
and Recognition

by

Vishal M. Patel

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor Rama Chellappa, Chair/Advisor
Professor Ramani Duraiswami
Professor Min Wu
Professor K. J. Ray Liu
Professor John J. Benedetto
Dr. Glenn R. Easley

DEDICATION

To my parents

and

in memory of Professor Dennis M. Healy, Jr.

Acknowledgments

I would like to express my sincere gratitude to the late Prof. Dennis M. Healy Jr. under whose supervision some of the research in this dissertation was carried out. I am honored to have worked with such a great man.

I would like to express my deepest gratitude to my advisor, Prof. Rama Chellappa for his guidance, mentoring and availability over the last few years. You are truly an inspiration and thanks for your generosity and encouragement.

I am very grateful to my committee members Prof. Ramani Duraiswami, Prof. K. J. Ray Liu, Prof. Min Wu, Prof. John J. Benedetto and Dr. Glenn R. Easley for their support. I am especially indebted to Dr. Glenn Easley for spending an enormous amount of time discussing various aspects of the dissertation. I am also grateful to Dr. Nasser Nasrabadi, Prof. Demetrio Labate, Prof. Robert Hartwig, Prof. Prakash Narayan and Prof. Adrian Papamarcou for their enlightening discussions during my graduate studies.

I am also particularly thankful to my high school teacher William Phillips for his enthusiasm and passion that initiated my interest in electrical engineering.

I would like to thank the administrative and technical staff at UMIACS, ECE and CfAR. In particular, many thanks to Janice Perrone, Dr. Tracy Chung, Melanie Prange, and Maria Hoo.

I would like to thank Dr. Pavan Turaga, Dr. Ashok Veeraraghavan, Dr. Soma Biswas, and Dr. Aswin Sankaranarayanan for their stimulating discussions on various topics. I also like to thank my friends and fellow graduate students for their support and friendship. I, especially, thank Chirag Lakhani, Akhilesh Khopkar, Sachin Vachhani, Amol Borkar, Don Punnoose, Anil Jacob, Navik Agrawal, Sanjay Varma, Tao Wu, Jai Pillai, Kaushik Mitra, Mahesh Ramachandran, Ruonan Li, Dikpal Reddy, Nitesh Shroff, Raghuraman Gopalan, Sima Taheri, Mohammed Abdelkader, Ming Du, Ming Liu, Ashish Shrivastava, Rick Miller, Garrett Warnell, Hien Van Nguyen, Tho Huy, Sumit Shekhar, Jie Ni, and Nazr e Batool.

My parents and family have been a guide in all my stages of life. This dissertation would not have been possible without the constant support, encouragement and love of my entire family - my parents Madhu and Manek, my sisters Julie, Dharti and Gunjali, my late grandfather Prahlad, my grandmother Kuvar, uncle Jimmy, aunt Kamala, my cousin Apoo, aunt Sharda, uncle Natu and more than 15 other aunts, uncles and cousins.

Finally, I would like to take this opportunity to thank *Ram*.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1   Research Motivation

In many applications in the sciences and engineering, one needs to approximate an unknown object given a collection of its linearly transformed, noisy and possibly incomplete observations. Such ill-conditioned problems often arise in a variety of fields such as remote sensing, medical imaging, astronomy, seismic analysis, etc. Of particular interest in this dissertation are the image deconvolution and image reconstruction problems.

In image deconvolution problem, the objective is to estimate an image given its blurred and noisy observation. The most popular methods for solving this problem are based on wavelets [2]. The main reason for wavelets' success can be explained by their ability to sparsely represent one-dimensional signals which are smooth away from point discontinuities. However, in two-dimension, wavelets are actually not optimal for all types of images. Multidimensional representations such as shearlets can provide better approximation rates for certain types of images [3]. Hence, in the first part of the dissertation, we study a new type of deconvolution algorithm based on estimating the image from a shearlet decomposition [4].

Reconstruction of images from an incomplete set of samples from a Fourier representation is an important goal to improve scanning technologies such as Magnetic Resonance Imaging (MRI) and Synthetic Aperture Radar (SAR). Solutions to such a problem would allow a significant reduction in collection time and improve the capacity to collect very time sensitive events. For example, an MRI scanner records data in a Fourier representation so methods, capable of reconstructing from such partial samples, would greatly improve medical science and reduce a patient's exposure time. To this end, in the second part of this dissertation, we study a reconstruction method that recovers images assumed to have a sparse representation

in a gradient domain by using partial measurement samples that are collected in the Fourier domain [5].

The next part of this dissertation concerns an interesting application of Compressed Sensing (CS) in SAR imaging. Compressed sensing is a new concept in signal processing and information theory where one measures a small number of non-adaptive linear combinations of the signal. These measurements are usually much smaller than the number of samples that define the signal. From these small number of measurements, the signal is then reconstructed by a non-linear procedure [6], [7]. Motivated by the theory of CS, we study a new imaging modality that can provide a high resolution map of the spatial distribution of targets and terrain based on a significant reduction in the number of transmitted and received electromagnetic waveforms [8], [9].

One principle for designing recognition algorithms for visual imagery is to encode objects as sparse representations from an over-complete dictionary. Areas of promise for sparse representation are face recognition and automatic target recognition. To this end, in the last part of this dissertation we present recognition algorithms based on learning dictionaries, both discriminative and reconstructive, for simultaneous sparse signal approximations and feature extraction. Applications in illumination robust face recognition and automatic target recognition on forward-looking infrared imagery consisting of military vehicles are presented.

1.2   Contributions

This dissertation makes the following contributions:

- **Shearlet-based deconvolution [4],[10]:** We study a deconvolution algorithm based on estimating the image from a shearlet decomposition. The threshold values for the noise shrinkage for each scale and direction are determined using a generalized cross validation method without the explicit knowledge of the noise variance.

- **Reconstruction of gradient sparse images from incomplete Fourier**

measurements [5]: We study a method that reconstructs images that have sparse or compressible gradients from an incomplete set of Fourier samples. We use a robust general Possion solver to obtain the reconstructed image from the estimated gradients. We show that a Poisson solver-based solution is much more robust to imperfections in the estimate of gradients and does not require any constraint on the type of data that are collected.

- **SAR imaging based on CS [9],[8]:** We study a new SAR imaging modality that can provide a high resolution map of the spatial distribution of targets and terrain based on a significant reduction in the number of transmitted and received electromagnetic waveforms. We study two different types of k-space undersampling schemes for SAR and analyze the severity of artifacts introduced by these undersampling schemes using a point spread function. We provide some of the applications of this new SAR imaging modality.

- **Sparsity-induced object recognition:** We study two different algorithms for object recognition based on simultaneous sparse approximations and feature extraction.

    - **Illumination robust face recognition:** The K-SVD algorithm is used to learn a reconstructive dictionary for each class. Furthermore, a re-lighting method based on estimating albedo of a face image is adapted to deal with illumination variations.

    - **Automatic Target Recognition (ATR) [11], [12]:** We study an ATR algorithm based on sparse representations. The dictionary learning algorithm is based on class supervised simultaneous orthogonal matching pursuit while a matching pursuit-based similarity measure is used for recognition.

1.3   Overview

1.3.1   Fourier-Shearlet Regularized Deconvolution

Inverse problems often involve estimating data from noisy and incomplete observations. Inverse problems are typically ill-posed, that is, the data cannot always be recovered. One such inverse problem is deconvolution, where the observation is comprised of an input image first degraded by linear time invariant convolution with a known impulse response and then corrupted by additive noise.

In chapter 3, we propose a new type of deconvolution algorithm that is based on estimating the image from a shearlet decomposition [4]. Shearlets provide a multi-directional and multi-scale decomposition that has been mathematically shown to represent distributed discontinuities such as edges better than traditional wavelets. We develop an algorithm that allows for the approximation inversion operator to be controlled on a multi-scale and multi-directional basis. Furthermore, we develop a method for the automatic determination of threshold values for the noise shrinkage for each scale and direction without explicit knowledge of the noise variance using a generalized cross validation. Various experiments show that this method can perform significantly better than many competitive deconvolution algorithms.

1.3.2   Enhancing sparsity using gradients for CS

Several scanning technologies such as MRI will greatly benefit from being able to reconstruct an image from a reduced sampling set. The theory of CS points to ways of reconstructing such images from undersampled data yet the underlying algorithms still need improvement to obtain satisfactory and efficient results. In chapter 4, we study a new reconstruction method that recovers images that have a sparse representation in a gradient domain yet have partial samples that are collected in the Fourier domain. This technique makes use of a generalized Poisson solver and experiments show that this method performs significantly better than similar reconstruction methods.

1.3.3 CS for Synthetic Aperture Radar (SAR) Imaging

SAR is a radar imaging technology that is capable of producing high resolution images of the stationary surface targets and terrain. In chapter 5, we present a new SAR imaging scheme based on compressing the number of transmitted waveforms. We will show that if the target reflectivity function is assumed to be sparse in some domain, one can reconstruct a good estimate of the reflectivity profile using a new image formation algorithm that relies on using far fewer number of waveforms than conventional systems do. Also, some applications of this compressed aperture radar will be presented.

1.3.4 Dictionary-based object recognition

In chapters 6 and 7, we study recognition algorithms based on learning dictionaries for simultaneous sparse signal approximations and feature extraction. A dictionary is learned for each object class based on given training examples which minimize the representation error with a sparseness constraint. A novel test image is then projected onto the span of the atoms in each learned dictionary. The residual vectors along with the coefficients are then used for recognition.

1.4 Organization

The dissertation is organized as follows. Chapter 2 presents some of the background materials that are central to this dissertation. In chapter 3, we present our deconvolution method based on shearlet transform. Chapter 4, presents an application of CS using image gradients. In chapter 5, we present an application of CS for SAR imaging. Dictionary-based recognition algorithms are presented in chapter 6 for face recognition and in chapter 7 for automatic target recognition. Finally, in chapter 8 we discuss some future research directions.

Chapter 2

Background

In this chapter, we cover some of the background materials that are central to this dissertation. The chapter is divided mainly into three parts. We first introduce the shearlet transform and discuss some of its properties. The second part of the chapter discusses the theory of compressed sensing. In the final part of the chapter, we give a brief background on SAR and formulate a discrete observation model for the spotlight mode SAR

2.1   The Shearlet Transform

The continuous wavelet transform provides a decomposition of a signal over dilated and translated versions of a fixed waveform ψ. Specifically, for a fixed $\psi \in L^2(\mathbb{R}^2)$, this is defined as the mapping W_ψ with domain $L^2(\mathbb{R}^2)$ such that for $x \in L^2(\mathbb{R}^2)$

$$W_\psi x(a, t) = \int_{\mathbb{R}^2} x(u) \, \overline{\psi_{a,t}(u)} \, du, \tag{2.1}$$

where $\psi_{a,t}(x) = a^{-1} \, \psi(a^{-1}(u - t))$, $a > 0$ and $t \in \mathbb{R}^2$. If the function ψ satisfies the admissibility or Calderòn condition

$$\int_0^\infty |\Psi(a\omega)|^2 \, \frac{da}{a^2} = 1 \quad \text{for a.e. } \omega \in \mathbb{R}^2$$

(where Ψ denotes the Fourier transform of ψ), then ψ is referred to as a *wavelet*, and any $x \in L^2(\mathbb{R}^2)$ can be recovered via the reproducing formula:

$$x = \int_0^\infty \int_{\mathbb{R}^2} \langle x, \psi_{a,t} \rangle \, \psi_{a,t} \, dt \, \tfrac{da}{a^4}.$$

Despite the success of wavelets in signal processing applications, it is known mathematically that traditional wavelets are not very effective in dealing with multidimensional signals containing discontinuities such as edges. In this section, we

briefly describe a recently developed multi-scale and multi-directional representation called *the shearlet transform* [3]. The shearlet transform combines the power of multi-scale methods with the ability to capture the geometry of multidimensional signals and provides an optimal representation for images containing edges.
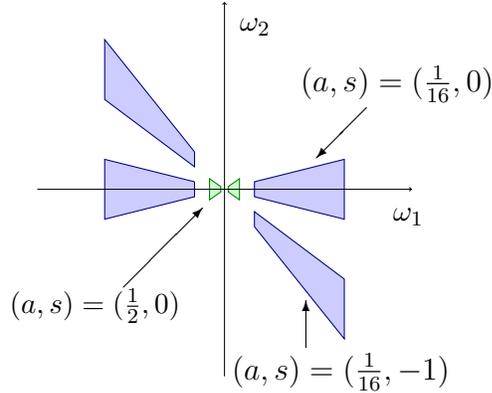
Figure 2.1: Frequency support of the shearlets for different values of a and s.

Consider the two-dimensional affine system

$$\{\psi_{a,s,t}(u) = |\det M_{a,s}|^{-\frac{1}{2}} \, \psi(M_{a,s}^{-1}u - t) : t \in \mathbb{R}^2\},$$

where

$$M_{a,s} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$$

is a product of a shearing and an anisotropic dilation matrix for $(a, s) \in \mathbb{R}^+ \times \mathbb{R}$. The generating functions ψ are such that, for $\omega_1 \neq 0$,

$$\Psi(\omega) = \Psi(\omega_1, \omega_2) = \Psi_1(\omega_1) \, \Psi_2\left(\frac{\omega_2}{\omega_1}\right), \tag{2.2}$$

where ψ_1 is a continuous wavelet for which $\Psi_1 \in C^\infty(\mathbb{R})$ with $\mathrm{supp}\,(\Psi_1) \subset [-2, -1/2] \cup [1/2, 2]$, and ψ_2 is chosen so that $\|\psi_2\| = 1$, $\Psi_2 \in C^\infty(\mathbb{R})$, $\mathrm{supp}\,(\Psi_2) \subset [-1, 1]$, with $\Psi_2 > 0$ on $(-1, 1)$. Then any $x \in L^2(\mathbb{R}^2)$ admits the representation

$$x(u) = \int_{\mathbb{R}^2} \int_{-\infty}^{\infty} \int_0^{\infty} \langle x, \psi_{a,s,t} \rangle \, \psi_{a,s,t}(u) \frac{da}{a^3} \, ds \, dt$$

for $a \in \mathbb{R}^+$, $s \in \mathbb{R}$, and $t \in \mathbb{R}^2$. The operator $\mathcal{SH}$ defined by

$$\mathcal{SH}x(a, s, t) = \langle x, \psi_{a,s,t} \rangle$$

is known as the *continuous shearlet transform* of $x \in L^2(\mathbb{R})$. The shearlet transform is a function of three variables: the scale a, the shear s and the translation t. In the frequency domain, $\Psi_{a,s,t}(\omega)$ has support in the set

$$\{(\omega_1, \omega_2) : \omega_1 \in [-\tfrac{2}{a}, -\tfrac{1}{2a}] \cup [-\tfrac{2}{a}, -\tfrac{1}{2a}], \left|\tfrac{\omega_2}{\omega_1} - s\right| \le \sqrt{a}\}.$$

Hence, each element $\Psi_{a,s,t}$ has support on a pair of trapezoids, at various scales, symmetric with respect to the origin and oriented along a line of slope s.

The collection of *discrete shearlets* is described by

$$\{\psi_{j,\ell,k} = |\det A_0|^{j/2} \, \psi(B_0^\ell A_0^j x - k) : j, \ell \in \mathbb{Z}, k \in \mathbb{Z}^2\},$$

where

$$B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

For the appropriate choices of ψ, the discrete shearlets form a Parseval frame (tight frame with bounds equal to 1) for $L^2(\mathbb{R}^2)$ [13], i.e., they satisfy the property

$$\sum_{j \in \mathbb{Z}, \ell \in \mathbb{Z}, k \in \mathbb{Z}^2} |\langle x, \psi_{j,\ell,k} \rangle|^2 = \|x\|^2.$$

The discrete shearlets described above provide a nonuniform angular covering of the frequency plane when restricted to the finite discrete setting for implementation. Thus, it is preferred to reformulate the shearlet transform with restrictions supported in the regions given by $\mathcal{D}_0 = \{(\omega_1, \omega_2) : |\omega_1| \ge 1/8, |\omega_2/\omega_1| \le 1\}$ and $\mathcal{D}_1 = \{(\omega_1, \omega_2) : |\omega_2| \ge 1/8, |\omega_1/\omega_2| \le 1\}$. Specifically, define

$$\Psi^{(0)}(\omega) = \Psi_1(\omega_1)\Psi_2(\tfrac{\omega_2}{\omega_1}), \quad \Psi^{(1)}(\omega) = \Psi_1(\omega_2)\Psi_2(\tfrac{\omega_1}{\omega_2})$$

where $\Psi_1, \Psi_2 \in C^\infty(\mathbb{R})$, supp $\Psi_1 \subset [-1/2, -1/16] \cup [1/16, 1/2]$ and supp $\Psi_2 \subset [-1, 1]$. In addition, we assume that

$$\sum_{j \ge 0} |\Psi_1(2^{-2j}\omega)|^2 = 1 \quad \text{for } |\omega| \ge \frac{1}{8},$$

and, for each $j \ge 0$,

$$\sum_{\ell=-2^j}^{2^j-1} |\Psi_2(2^j \omega - \ell)|^2 = 1 \quad \text{for } |\omega| \le 1.$$

Let

$$A_1 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

and choose $\Phi \in C_0^\infty(\mathbb{R}^2)$ to satisfy

$$|\Phi(\omega)|^2 + \sum_{d=0}^{1} \sum_{j \geq 0} \sum_{\ell=-2^j}^{2^j-1} |\Psi^{(d)}(\omega A_d^{-j} B_d^{-\ell})|^2 \, \chi_{\mathcal{D}_d}(\xi) = 1$$

for $\omega \in \mathbb{R}^2$, where $\chi_{\mathcal{D}}$ denotes the indicator function of the set $\mathcal{D}$.

With the functions ϕ and ψ as above, we deduce the following result.

Theorem 2.1.1 ([3]). *Let $\phi_k(x) = \phi(x - k)$ and $\psi_{j,\ell,k}^{(d)}(x) = 2^{\frac{3j}{2}} \psi^{(d)}(B_d^\ell A_d^j x - k)$. Then the collection of* shearlets $\{\phi_k : k \in \mathbb{Z}^2\}$ *together with*

$$\cup \{\psi_{j,\ell,k}^{(d)}(x) : j \geq 0, \; -2^j + 1 \leq \ell \leq 2^j - 2, \; k \in \mathbb{Z}^2, d = 0, 1\}$$

$$\cup \{\tilde{\psi}_{j,\ell,k}^{(d)}(x) : j \geq 0, \; \ell = -2^j, 2^j - 1, \; k \in \mathbb{Z}^2, d = 0, 1\},$$

is a Parseval frame for $L^2(\mathbb{R}^2)$, where $\tilde{\Psi}_{j,\ell,k}^{(d)} = \Psi_{j,\ell,k}^{(d)} \chi_{\mathcal{D}_d}$.

Based on ψ_1 and ψ_2, filters v_j and $w_{j,\ell}^{(d)}$ can be found so that $\langle x, \psi_{j,\ell,k}^{(d)} \rangle$ and $\langle x, \tilde{\psi}_{j,\ell,k}^{(d)} \rangle$ can be computed as

$$x * (v_j * w_{j,\ell}^{(d)})[k] \triangleq x * g_{j,\ell}^{(d)}[k],$$

where $g_{j,\ell}^{(d)} = v_j * w_{j,\ell}^{(d)}$ are the directionally-oriented filters. To simplify the notation, we suppress the superscript (d) and absorb the distinction between $d = 0$ and 1 by re-indexing the parameter ℓ so that it has double the cardinality. An M-channel filterbank whose filters correspond to $g_{j,\ell}$ can be implemented by using the techniques given in [14]. As a consequence, its implementation has a complexity of $O(N^2 \log_2(N))$ for an $N \times N$ image.

Notice that, just as in the continuous version, each element $\Psi_{j,\ell,k}$ is supported on a pair of trapezoids, and each trapezoid is contained in a box of size approximately $2^j \times 2^{2j}$ satisfying a parabolic scaling property. Their supports become increasingly thin as $j \to \infty$. and the elements $\Psi_{j,\ell,k}$ exhibit highly directional sensitivity since they are oriented along lines with slope given by $-\ell 2^{-j}$. These properties contribute to being able to establish the following theorem.

9

Theorem 2.1.2 ([15]). *Let x be C^2 away from piecewise C^2 curves, and let x_M^S be the approximate reconstruction of x using the M largest coefficients in the shearlet expansion. Then*

$$\|x - x_M^S\|_2^2 \leq C\,M^{-2}\,(\log M)^3.$$

The significance of this result is that a shearlet-based estimate yields a MSE approximation rate of $O(\epsilon^{4/3})$ as $\epsilon \to 0$, where ϵ is the noise level of the noisy image [16]. (This is achieved by choosing a threshold so that one reconstructs from the largest $M \approx \epsilon^{-2/3}$ noisy shearlet coefficients.) Similarly, one obtains the MSE approximation rate of wavelet thresholding as $O(\epsilon)$ for $\epsilon \to 0$.

The shearlet transform has similarities to the curvelet transform and the contourlet transform. Shearlets and curvelets, in fact, are the only two systems which are known mathematically to provide the rate of $O(M^{-2}\,(\log M)^3)$ using the M largest coefficients for images described as C^2 away from piecewise C^2 curves. The spatial-frequency tilings of curvelet and shearlet representations are completely different theoretically, yet the implementations of the curvelet transform corresponds to essentially the same tiling as that of the shearlet or contourlet transform.

Alternative discrete shearlet decompositions can be created by varying the support of the mother wavelet Ψ (which amounts to changes in Ψ_1) and changing the dilation matrices A_0 and A_1. Such changes produce different spatial-frequency tilings composed of regions of support that are restricted to pairs of various trapezoidal regions. When implemented in an undecimated form, the shearlet transform will produce a highly redundant decomposition consisting of the total number of paired trapezoidal regions considered.

As will be seen in Chapter 3 that an important advantage in the use of this redundant shearlet transform implementation for deconvolution is that it allows one to independently estimate each directionally-oriented frequency band with different amounts of regularization. This has not been done before and cannot be done by the current curvelet or contourlet transform implementations.

2.2 Compressed/Compressive Sensing

Compressed sensing is a new concept in signal processing and information theory where one measures a small number of non-adaptive linear combinations of the signal. These measurements are usually much smaller than the number of samples that define the signal. From these small number of measurements, the signal is then reconstructed by a non-linear procedure [6], [7]. In what follows, we present some fundamental premises underlying CS: sparsity, incoherent sampling and non-linear recovery. Excellent tutorials of CS can be found in [17], [18], [19], [20], [21], and [22].

2.2.1 Sparse representations

Consider an unknown finite length discrete time signal $\mathbf{x} \in R^N$. $\mathbf{x}$ can be viewed as an $N \times 1$ column vector with elements $x[n]$, $n = 1, ..., N$. Let's assume that the basis $\Psi = [\psi_1, ..., \psi_N]$ is an orthonormal. Then, any vector in $\mathbb{R}^N$ can be expressed as

$$\mathbf{x} = \sum_{i=1}^{N} \theta_i \psi_i \tag{2.3}$$

or more compactly $\mathbf{x} = \Psi\theta$, where θ is an $N \times 1$ column vector of coefficients. These coefficients are given by $\theta_i = \langle \mathbf{x}, \psi_i \rangle = \psi_i^T \mathbf{x}$ where $.^T$ denotes the transposition operation. If the basis Ψ provides a K-sparse representation of $\mathbf{x}$, then (2.3) can be rewritten as

$$\mathbf{x} = \sum_{i=1}^{K} \theta_{n_i} \psi_{n_i}, \tag{2.4}$$

where $\{n_i\}$ are the indices of the coefficients and the basis elements corresponding to the K nonzero entries. In this case, θ is an $N \times 1$ column vector with only K nonzero elements. That is, $\| \theta \|_0 = K$ where $\| . \|_p$ denotes the ℓ_p norm defined as

$$\| \mathbf{x} \|_p = \left(\sum_i | x_i |^p \right)^{\frac{1}{p}} \tag{2.5}$$

and the ℓ_0 norm is defined as the limit as $p \to 0$ of the ℓ_p norm

$$\| \mathbf{x} \|_0 = \lim_{p \to 0} \| \mathbf{x} \|_p^p = \lim_{p \to 0} \sum_i | x_i |^p . \tag{2.6}$$

In general, the ℓ_0 norm counts the number of non-zero elements in a vector

$$\| \mathbf{x} \|_0 = \sharp\{i : x_i \neq 0\}. \tag{2.7}$$

Typically, real-world signals are not exactly sparse in any orthogonal basis. Instead, they are compressible. A signal is said to be *compressible* if the magnitude of the coefficients, when sorted in a decreasing order, decays according to a power law [23],[24]. That is,

$$\mid \theta \mid_{(n)} \leq C.n^{-s}, \tag{2.8}$$

where $\mid \theta \mid_{(n)}$ is the nth largest value of θ, $s \geq 1$ and C is a constant. For a given L, the L-term linear combination of elements that best approximate $\mathbf{x}$ in an L_2-sense is obtained by keeping the L largest terms in the expansion

$$\mathbf{x}_L = \sum_{n=0}^{L-1} \theta_{(n)} \psi_{(n)}. \tag{2.9}$$

If θ obeys (3.5), then the error between $\mathbf{x}_L$ and $\mathbf{x}$ also obeys a power law as well [23], [24]

$$\| \mathbf{x}_L - \mathbf{x} \|_2 \leq CL^{-(s-\frac{1}{2})}. \tag{2.10}$$

In other words, a small number of vectors from Ψ can provide accurate approximations to $\mathbf{x}$.

2.2.2 Incoherent Sampling

In CS, the K-largest θ_i in (2.3) are not measured directly. Instead, $M \ll N$ projections of the vector $\mathbf{x}$ with a collection of vectors $\{\phi_j\}_{j=1}^M$ are measured as in $y_j = \langle x, \phi_j \rangle$. Arranging the measurement vector ϕ_j^T as rows in an $M \times N$ matrix Φ and using (2.3), the measurement process can be written as

$$\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\theta = \Theta\theta, \tag{2.11}$$

where $\mathbf{y}$ is an $M \times 1$ column vector of the compressive measurements and $\Theta = \Phi\Psi$ is the measurement matrix or the sensing matrix. Given an $M \times N$ sensing matrix Θ, the first question is to determine whether Θ is good for compressive sensing. Candes

and Tao introduced a necessary condition on Θ that guarantees a stable solution for both K-sparse and compressible signals [17], [25].

Definition 2.2.1. A matrix Θ is said to satisfy the RIP of order K with constants $\delta_K \in (0, 1)$ if

$$(1 - \delta_K) \parallel \mathbf{v} \parallel_2^2 \leq \parallel \Theta\mathbf{v} \parallel_2^2 \leq (1 + \delta_K) \parallel \mathbf{v} \parallel_2^2 \tag{2.12}$$

for any $\mathbf{v}$ such that $\parallel \mathbf{v} \parallel_0 \leq K$.

An equivalent description of RIP is to say that all subsets of K columns taken from Θ are nearly orthogonal. This in turn implies that K-sparse vectors cannot be in the null space of Θ. When RIP holds, Θ approximately preserves the Euclidean length of K-sparse vectors. That is,

$$(1 - \delta_{2K}) \parallel \mathbf{v_1} - \mathbf{v_2} \parallel_2^2 \leq \parallel \Theta\mathbf{v_1} - \Theta\mathbf{v_2} \parallel_2^2 \leq (1 + \delta_{2K}) \parallel \mathbf{v_1} - \mathbf{v_2} \parallel_2^2 \tag{2.13}$$

holds for all K-sparse vectors $\mathbf{v_1}$ and $\mathbf{v_2}$. A related condition known as incoherence, requires that the rows of Φ can not sparsely represent the columns of Ψ and vice versa.

Definition 2.2.2. The coherence between Φ and the representation basis Ψ is

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \leq i,j \leq N} \mid \langle \phi_i, \psi_j \rangle \mid, \tag{2.14}$$

where $\phi_i \in \Phi$ and $\psi_j \in \Psi$. The number μ measures how much two vectors in $\Theta = \Phi\Psi$ can look alike. The value of μ is between 1 and $\sqrt{N}$. We say that a matrix Θ is *incoherent* when μ is very small.

The incoherence holds for many pairs of bases. For example, it holds for the delta spikes and the Fourier bases. Surprisingly, with high probability, incoherence holds between any arbitrary basis and a random matrix such as Gaussian or Bernoulli [19], [26].

2.2.3 Recovery

Since, $M \ll N$, we have an underdetermined system of linear equations, which in general has infinitely many solutions. So our problem is ill-posed. If one desires

to narrow the choice to a well-defined solution, additional constraints are needed. One approach is to find a minimum-norm solution by minimizing the following optimization problem

$$\hat{\theta} = \arg \min_{\theta' \in \mathbb{R}^N} \parallel \theta' \parallel_2 \quad \text{subject to} \quad \mathbf{y} = \Theta\theta'. \tag{2.15}$$

The solution to the above problem is explicitly given by

$$\hat{\theta} = \Theta^\dagger \mathbf{y} = \Theta^*(\Theta\Theta^*)^{-1}\mathbf{y}, \tag{2.16}$$

where Θ^* is the adjoint of Θ. This solution, however, yields a non-sparse vector. The approach taken in CS is to instead find the sparsest solution.

The problem of finding the sparsest solution can be reformulated as finding a vector $\theta \in \mathbb{R}^N$ with a minimum possible number of nonzero entries. That is

$$\hat{\theta} = \arg \min_{\theta' \in \mathbb{R}^N} \parallel \theta' \parallel_0 \quad \text{subject to} \quad \mathbf{y} = \Theta\theta'. \tag{2.17}$$

This problem can recover a K-sparse signal exactly. However, this is an NP-hard problem [19]. It requires an exhaustive search of all $\binom{N}{K}$ possible locations of the nonzero entries in θ.

The main approach taken in CS is to minimize the ℓ_1 norm instead

$$\hat{\theta} = \arg \min_{\theta' \in \mathbb{R}^N} \parallel \theta' \parallel_1 \quad \text{subject to} \quad \mathbf{y} = \Theta\theta'. \tag{2.18}$$

Surprisingly, the ℓ_1 minimization yields the same result as the ℓ_0 minimization in many cases of practical interest. This program also approximates compressible signals. This convex optimization program is often known as Basis Pursuit (BP) [27]. The use of ℓ_1 minimization for signal restoration was initially observed by engineers working in seismic exploration as early as 1970s [21]. In last few years, a series of papers [6], [26], [28], [29], [24], [30], explained why ℓ_1 minimization can recover sparse signals in various practical setups.

2.2.3.1   Robust CS

In this section we examine the case when there are noisy observations of the following form

$$\mathbf{y} = \Theta\theta + \eta \tag{2.19}$$

14

where $\eta \in \mathbb{R}^M$ is the measurement noise or an error term. Note that η can be stochastic or deterministic. Furthermore, let's assume that $\|\eta\|_2 \leq \varepsilon$. Then, $\mathbf{x}$ can be recovered from $\mathbf{y}$ via θ by solving the following problem

$$\hat{\theta} = \arg \min_{\theta' \in \mathbb{R}^N} \| \theta' \|_1 \quad \text{subject to} \quad \| \mathbf{y} - \Theta\theta' \| \leq \varepsilon. \tag{2.20}$$

The problem (2.20) is often known as Basis Pursuit DeNoising (BPDN) [27]. In [30], Candes *at. el.* showed that the solution to (2.20) recovers an unknown sparse signal with an error at most proportional to the noise level.

Theorem 2.2.1. *[30] Let Θ satisfy RIP of order $4K$ with $\delta_{3K} + 3\delta_{4K} < 2$. Then, for any K sparse signal θ and any perturbation η with $\|\eta\|_2 \leq \varepsilon$, the solution $\hat{\theta}$ to (2.20) obeys*

$$\| \hat{\theta} - \theta \|_2 \leq \varepsilon C_K$$

with a well behaved constant C_K.

Note that for K obeying the condition of the theorem, the reconstruction from noiseless data is exact. A similar result also holds for stable recovery from imperfect measurements for approximately sparse signals (i.e compressible signals).

Theorem 2.2.2. *[30] Let Θ satisfy RIP of order $4K$. Suppose that θ is an arbitrary vector in $\mathbb{R}^N$ and let θ_K be the truncated vector corresponding to the K largest values of θ in magnitude. Under the hypothesis of Theorem 2.2.1, the solution $\hat{\theta}$ to (2.20) obeys*

$$\| \hat{\theta} - \theta \|_2 \leq \varepsilon C_{1,K} + C_{2,K} \frac{\| \theta - \theta_K \|_1}{\sqrt{K}}$$

with well behaved constants $C_{1,K}$ and $C_{2,K}$.

If θ obeys (3.5), then

$$\frac{\| \hat{\theta} - \theta_K \|_1}{\sqrt{K}} \leq C' K^{-(s-\frac{1}{2})}. \tag{2.21}$$

So in this case

$$\| \hat{\theta} - \theta_K \|_2 \leq C'' K^{-(s-\frac{1}{2})},$$

and for signal obeying (3.5), there are fundamentally no better estimates available. This, in turn, means that with only M measurements, we can achieve an approximation error which is almost as good as that one obtains by knowing everything about the signal θ and selecting its K-largest elements [30].

2.2.3.2 The Dantzig selector

In (2.19), if the noise is assumed to be Gaussian with mean zero and variance σ^2, $\eta \sim \mathcal{N}(0, \sigma^2)$, then the stable recovery of the reflectivity field is also possible by solving a modified optimization problem

$$\hat{\theta} = \arg\min_{\theta'} \|\, \theta' \,\|_1 \quad \text{s. t.} \quad \|\, \Theta^T(\mathbf{y} - \Theta\theta') \,\|_\infty \leq \varepsilon' \tag{2.22}$$

where $\varepsilon' = \lambda_N \sigma$ for some $\lambda_N > 0$ and $\|.\|_\infty$ denotes the ℓ_∞ norm. The above program is known as the Dantzig Selector [31].

Theorem 2.2.3. *[31] Suppose $\theta \in \mathbb{R}^N$ is any $K-$sparse vector obeying $\delta_{2K} + \vartheta_{K,2K} < 1$. Choose $\lambda_N = \sqrt{2\log(N)}$ in (2.22). Then, with a high probability, the solution to (2.22), $\hat{\theta}$ obeys*

$$\|\hat{\theta} - \theta\|_2^2 \leq C_1^2.(2\log(N)).K.\sigma^2, \tag{2.23}$$

with

$$C_1 = \frac{4}{1 - \delta_K - \vartheta_{K,2K}},$$

where $\vartheta_{K,2K}$ is the $K, 2K$-restricted orthogonal constant defined as follows

Definition 2.2.3. The K, K'-restricted orthogonality constant $\vartheta_{K,K'}$ for $K + K' \leq N$ is defined to be the smallest quantity such that

$$|\langle \Theta_T \mathbf{v}, \Theta_{T'} \mathbf{v}' \rangle| \leq \vartheta_{K,K'} \|\mathbf{v}\|_2 \|\mathbf{v}'\|_2 \tag{2.24}$$

holds for all disjoint sets $T, T' \subseteq \{1, ..., N\}$ of cardinality $|T| \leq K$ and $|T'| \leq K'$.

A similar result also exists for compressible signals (see [31] for more details).

The ℓ_1 minimization problem (2.22) is a linear program [31] while (2.20) is a second-order cone program (SOCP) [27], which can can be solved using interior

point methods [32]. Log-barrier and primal dual methods can also be used [33], [34] to solve SOCPs. Note, the optimization problems (2.18), (2.20), and (2.22) minimize convex functionals, hence a global minimum is guaranteed.

In the following sections, we describe other CS related reconstruction algorithms.

2.2.3.3 Iterative thresholding algorithms

A Lagrangian formulation of the problem (2.20) is the following

$$\hat{\theta} = \arg\min_{\theta'} \|\mathbf{y} - \Theta\theta'\|_2^2 + \lambda\|\theta'\|_1. \tag{2.25}$$

There exists a mapping between λ from (2.25) and ε from (2.20) so that both problems (2.20) and (2.25) are equivalent. Several authors have proposed to solve (2.25) iteratively [35], [36], [37], [38]. This algorithm iteratively performs a soft-thresholding to decrease the ℓ_1 norm of the coefficients θ and a gradient descent to decrease the value of $\|\mathbf{y} - \Theta\theta\|_2^2$. The following iteration is usually used

$$\mathbf{y}^{n+1} = T_\lambda(\mathbf{y}^n + \Theta^*(\theta - \Theta\mathbf{y}^n)), \tag{2.26}$$

where T_λ is the element wise soft-thresholding operator

$$T_\lambda(\alpha) = \begin{cases} \alpha + \frac{\lambda}{2}, & \text{if } \alpha \leq \frac{-\lambda}{2} \\ 0, & \text{if } |\alpha| < \frac{\lambda}{2} \\ \alpha - \frac{\lambda}{2}, & \text{if } \alpha \geq \frac{\lambda}{2}. \end{cases} \tag{2.27}$$

The iterates $\mathbf{y}^{n+1}$ converge to the solution of (2.20), $\hat{\theta}$ if $\|\Theta\|_2 < 1$ [36]. Similar results can also be obtained using hard-thresholding instead of the soft-thresholding methods described in (2.26) [37].

Other methods for solving (2.25) have also been proposed. See for instance GPSR [39], SPGL1 [40], Bregman iterations [41], split Bregman iterations [42], SpaRSA [43], and references therein.

2.2.3.4 Greedy Pursuits

In certain conditions, greedy algorithms such as matching pursuit [44], orthogonal matching pursuit [45], [46], gradient pursuits [47], and stagewise orthogonal matching pursuit [48] can also be used to recover sparse (or in some cases compressible) θ from (2.19). In particular, a greedy algorithm known as, CoSaMP, is well supported by theoretical analysis and provides the same guarantees as some of the optimization based approaches [49].

2.2.3.5 Other algorithms

Recently, there has been a great interest in using ℓ_p minimization with $p < 1$ for compressive sensing [50]. It has been observed that the minimization of such a nonconvex problem leads to recovery of signals that are much less sparse than required by the traditional methods [50]. Other related algorithms such as FOCUSS and reweighted ℓ_1 have also been proposed in [51] and [52].

2.2.4 Sensing Matrices

Most of the sensing matrices in CS are produced by taking i.i.d. random variables with some given probability distribution and then normalizing their columns. These matrices are guaranteed to perform well with high probability. In what follows, we present some commonly used sensing matrices in CS [30], [26], [17].

- *Random matrices with i.i.d. entries:* Consider a matrix Θ with entries drawn independently from the Gaussian probability distribution with mean zero and variance $1/M$. Then the conditions for Theorem 2.2.1 hold with a high probability when

$$K \leq CM/\log(N/M).$$

- *Fourier ensemble:* Let Θ be an $M \times N$ matrix obtained by selecting M rows, at random, from the $N \times N$ discrete Fourier transform matrix and renormalizing the columns. Then with a high probability, the conditions for Theorem 2.2.1
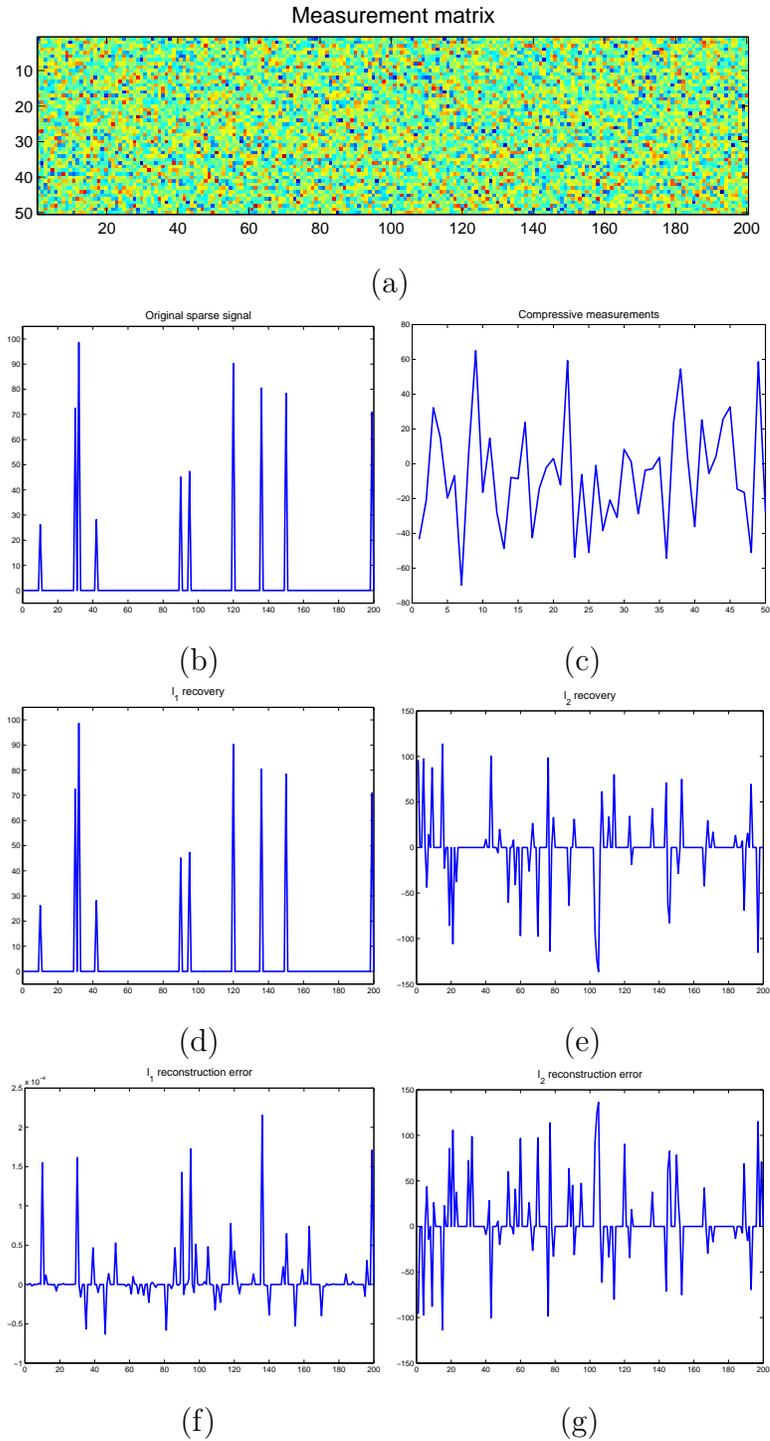
Figure 2.2: 1D sparse signal recovery example from random Gaussian measurements. (a) Compressive measurement matrix. (b) Original sparse signal. (c) Compressive measurements. (d) ℓ_1 recovery. (e) ℓ_2 recovery. (f) ℓ_1 reconstruction error. (g) ℓ_2 reconstruction error.
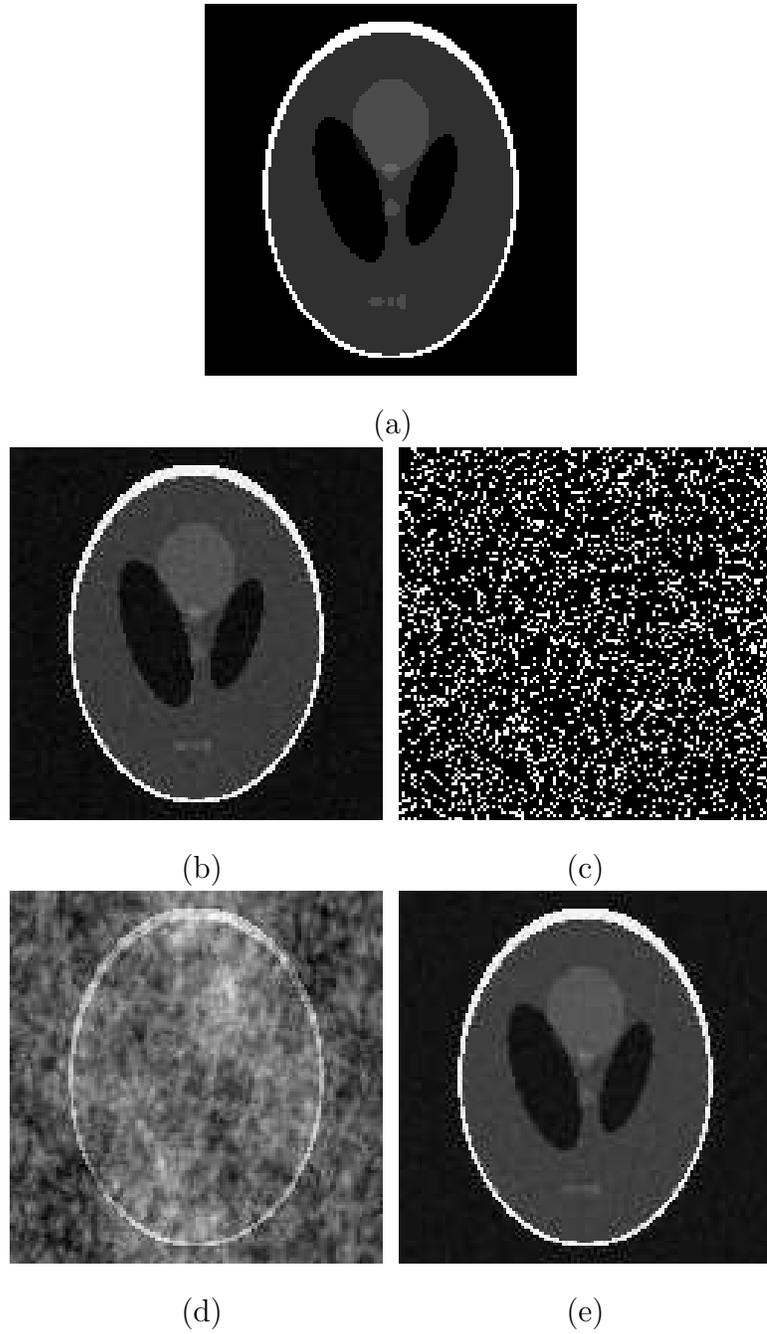
Figure 2.3: 2D sparse image recovery example from random Fourier measurements. (a) Original image. (b) Original image contaminated by additive white Gaussian noise with signal-to-noise ratio of 20 dB. (c) Sampling mask in the Fourier domain. (d) ℓ_2 recovery. (e) ℓ_1 recovery.

holds provided that

$$K \leq C \frac{M}{(\log(N))^6}.$$

- *General orthogonal ensembles:* Suppose Θ is obtained by selecting M rows from an $N \times N$ orthonormal matrix Ξ and renormalizing the columns. If the rows are selected at random, then the conditions for Theorem 2.2.1 hold with a high probability when

$$K \leq C \frac{1}{\mu^2} \frac{M}{(\log(N))^6},$$

where μ is defined in (2.14).

2.2.5   Numerical Examples

We end this section by considering the following two examples. In the first example, a 1D signal $\mathbf{x}$ of length 200 with only 10 nonzero elements is undersampled using a random Gaussian matrix $\mathbf{A}$ of size 50×200 as shown in Fig. 2.2(a). Here, the sparsifying transform is simply the identity and the observation vector $\mathbf{y}$ is of length 50. Having observed $\mathbf{y}$ and knowing $\mathbf{A}$ the signal $\mathbf{x}$ is then recovered by solving the following optimization problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}' \in \mathbb{R}^N} \| \mathbf{x}' \|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}'. \tag{2.28}$$

As can be seen from Fig. 2.2(d), indeed the solution to the above optimization problem recovers the sparse signal exactly from highly undersampled observations. Whereas, the minimum norm solution (i.e. by minimizing the ℓ_2 norm), as shown in Fig. 2.2(e), fails to recover the sparse signal. The errors corresponding the ℓ_1 and ℓ_2 recovery are shown in Fig. 2.2(f) and Fig. 2.2(g), respectively.

In the second example, we reconstructed an undersampled Shepp-Logan phantom image of size 128×128 in the presence of additive white Gaussian noise with signal-to-noise ratio of 30 dB. For this example, we used Haar wavelets as a sparsifying transform. So the observations can be written as $\mathbf{y} = \mathbf{M}\mathbf{F}\boldsymbol{\Psi}\mathbf{x_0} + \eta$, where $\mathbf{y}, \mathbf{M}, \mathbf{F}, \Psi, \mathbf{x_0}$ and η are the noisy compressive measurements, the restriction operator, Fourier transform operator, the Haar transform operator, the sparse coefficients

vector and the noise vector with $\|\eta\|_2 \leq \varepsilon$, respectively. The image was reconstructed via $\mathbf{x_0}$ by solving the following optimization problem

$$\hat{\mathbf{x_0}} = \arg \min_{\mathbf{x_0}' \in \mathbb{R}^N} \| \mathbf{x_0'} \|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{MF\Psi x_0}'\| \leq \varepsilon. \tag{2.29}$$

For this example, we used only 15% of the random Fourier measurements. The reconstruction from ℓ_2 and ℓ_1 minimization is shown in Fig. 2.2(d) and Fig. 2.2(e), respectively. This example shows that, it is possible to obtain a stable reconstruction from the compressive measurements in the presence of noise. For both of the above examples we used SPGL1 [40] algorithm.

In [7], [6], a theoretical bound on the number of samples that need to be measured for a good reconstruction has been derived. However, it has been observed by many researchers [53], [30], [26], [24], [17] that in practice samples in the order of two to five times the number of sparse coefficients suffice for a good reconstruction. Our experiments also support this claim.

2.3   SAR Imaging

In this section, we give a brief background on SAR and formulate a discrete observation model for the spotlight mode SAR [54], [55], [56], [57].

2.3.1   Introduction to SAR

Synthetic Aperture Radar (SAR) is a radar imaging modality that is capable of producing high resolution images of the stationary surface targets and terrain. The main advantages of SAR imaging are its ability to operate at night and in adverse weather conditions, hence overcoming limitations of both optical and infrared systems. The basic idea of SAR is as follows: as the radar moves along its path, it transmits pulses at microwave frequencies at an uniform pulse repetition interval (PRI) which is defined as 1/PRF, where PRF is the pulse repetition frequency. The reflected energy at any instant can be modelled as a convolution of the pulse waveform with the ground reflectivity function . Each received pulse is pre-processed

and passed on to an image formation processor. The image formation processor produces an image that is a two dimensional mapping of the illuminated scene [58] (i.e. the reflectivity function) [58]. A depiction of the SAR image formation process is shown in Fig. 2.4.
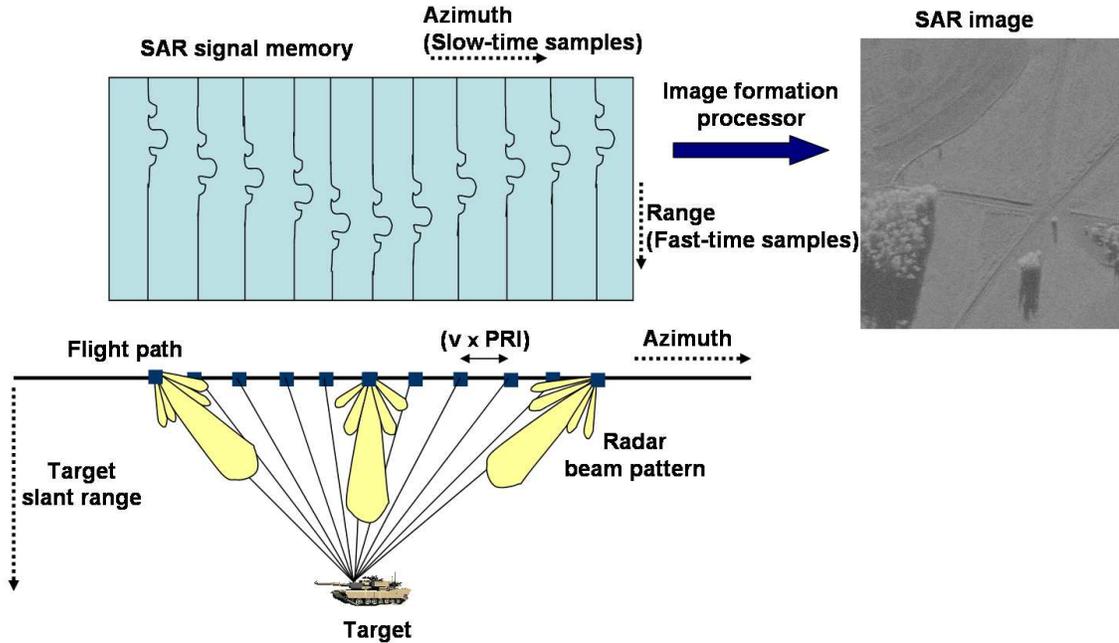
Figure 2.4: Spotlight SAR data collection in 2D.

The two dimensional image formed is interpreted in the dimensions of range and cross-range or azimuth. The range is the direction of signal propagation and the cross-range is the direction parallel to the flight path. Sometimes the range and the cross-range samples are referred to as fast-time and slow-time samples, respectively.

2.3.1.1   Resolution

In order to compare the effects of each type of processing, in this section we define the notion of *resolution*. The resolution in either dimension of a system is defined as the minimum separation of two target responses that can be distinguished or resolved as separate by the system [59], [60]. The range resolution, δ_r, for SAR depends on an effective transmitted pulse length or on a signal bandwidth. It is

given by [59]

$$\delta_r = \frac{c}{2\beta},\tag{2.30}$$

where β is the signal bandwidth and c is the speed of light. In a real beam radar, as shown in Fig. 2.5(a), the cross-range resolution is approximated by [61]

$$\delta_{cr}^r \approx \frac{k\lambda R}{D},\tag{2.31}$$

where λ, D and R are the radar wavelength, the width of the radar antenna, and the range to a target scene, respectively and k depends on the antenna design. For ideal aperture, k is usually 0.89 [61]. We will assume that k is approximately equal to one and ignore it in subsequent calculations. The cross-range resolution given by (2.31) is not practical for imaging purposes. To see this, let us consider the following parameters. If $\lambda = 0.02\ m$ (x band radar), $D = 1.5\ m$, $R = 150\ km$, then $\delta_{cr}^r = 2000\ m$. This resolution is too coarse for imaging purposes.
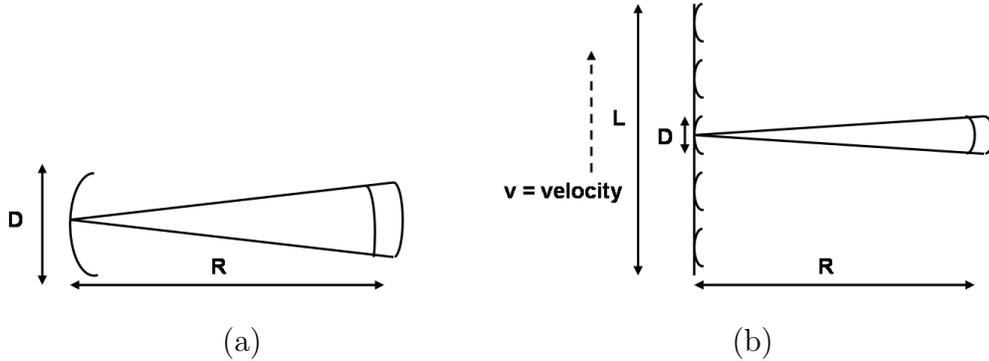
Figure 2.5: (a) Real beam radar. (b) Synthetic aperture radar.

On the other hand, SAR achieves the fine cross-range resolution by synthesizing the effect of a large antenna using multiple observations from a small antenna. Consider the synthetic aperture setup shown in Figure 2.5(b). The corss-range resolution for a synthetic aperture of length L is given by [59]

$$\delta_{cr}^{SAR} \approx \frac{\lambda R}{2L} = \frac{\lambda R}{2vT_a},\tag{2.32}$$

where v is the sensor velocity and T_a is the aperture time. In comparison with the real beam radar, for $L = 7\ km$, the cross-range resolution for a SAR using the same parameters as before is given by $\delta_{cr}^{SAR} = 0.21\ m$.

2.3.1.2 SAR modes

(a)   (b)
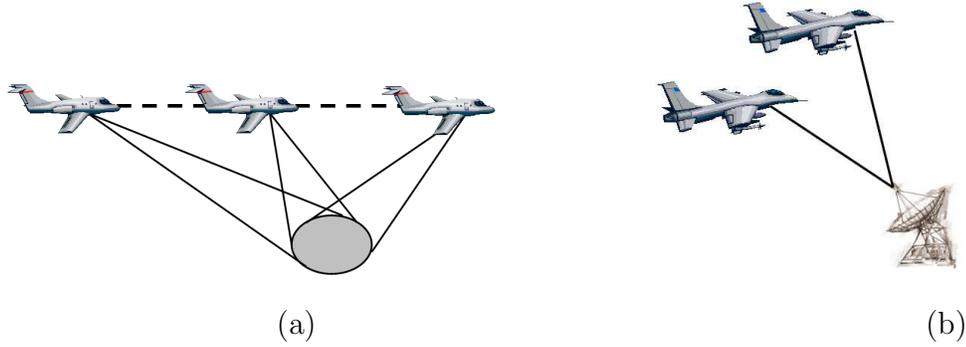
Figure 2.6: SAR modes. (a) Spotlight (b) ISAR

There are four common modes of SAR: scan, stripmap, spotlight and inverse SAR (ISAR). In this dissertation, we will mainly focus on the spotlight mode SAR and ISAR. In spotlight mode SAR, as shown in Fig. 2.6(a), the radar sensor steers its antenna beam to continuously illuminate the terrain patch being imaged. It can provide higher resolution than the stripmap and scan mode SAR because it maps a scene at multiple viewing angles during a single pass [58]. In ISAR, Fig. 2.6(b), the radar is stationary and the target is moving. The angular motion of the target with respect to the radar can be used to form an image of the target. Differential Doppler shifts of adjacent scatters on a target are observed and the target's reflectivity function is obtained through the Doppler frequency spectrum [62].

2.3.1.3 PRF Constraints

The PRF determines to what extent the observed ranges and doppler frequencies will be (un)ambiguous [59], [61], [63]. If the PRF is too high then, an echo from a distant target may return after the transmitter has transmitted another pulse. In this case, the radar will be unable to distinguish between pulses, because it would be impossible to determine whether the returned pulse is the echo of the pulse just transmitted or the echo of the preceding pulse. This situation is referred to as range ambiguity.

The roundtrip travel time of the transmitted and reflected back from a target located at range R is given by $\frac{2R}{c}$, where c is the speed of light. Since most radars can not simultaneously transmit and receive, the radar must wait for the trailing edge of the returned pulse to be received before it can transmit again. For a point target, the required minimum PRI for range unambiguous processing is given by [59], [63]

$$PRI_{\min} = \frac{2R}{c} + T, \tag{2.33}$$

where T is the transmitted pulse length. In many cases, T is very small compared with $\frac{2R}{c}$. So the minimum PRI can be approximated by $\frac{2R}{c}$. Hence, the maximum PRF for unambiguous range is

$$PRF_{\max} = \frac{c}{2R}. \tag{2.34}$$

A second constraint on the PRF is due to the Doppler bandwidth. The PRF must equal or exceed the maximum Doppler shift of the echo signals. The lower limit on the PRF is [59], [63]

$$PRF_{\min} = \frac{2v}{D}, \tag{2.35}$$

where D is the radar antenna length and v is the radar platform velocity. Hence, the PRF is constrained by the following inequality

$$\frac{2v}{D} \leq PRF \leq \frac{c}{2R}. \tag{2.36}$$

2.3.1.4 Spotlight Mode SAR Phase Histories

In this section, we give a brief description of the tomographic formulation of the spotlight-mode SAR [54], [55], [56], [57]. The ground plane geometry in spotlight-mode SAR is shown in Fig. 2.3.1.4. The radar, which transverses the flight path, continuously points in the direction of a ground patch. At locations corresponding to each increments of θ, high-bandwidth pulses are transmitted. The returns from the ground patch are received and processed to image the reflectivity profile $g(x, y)$ [54].
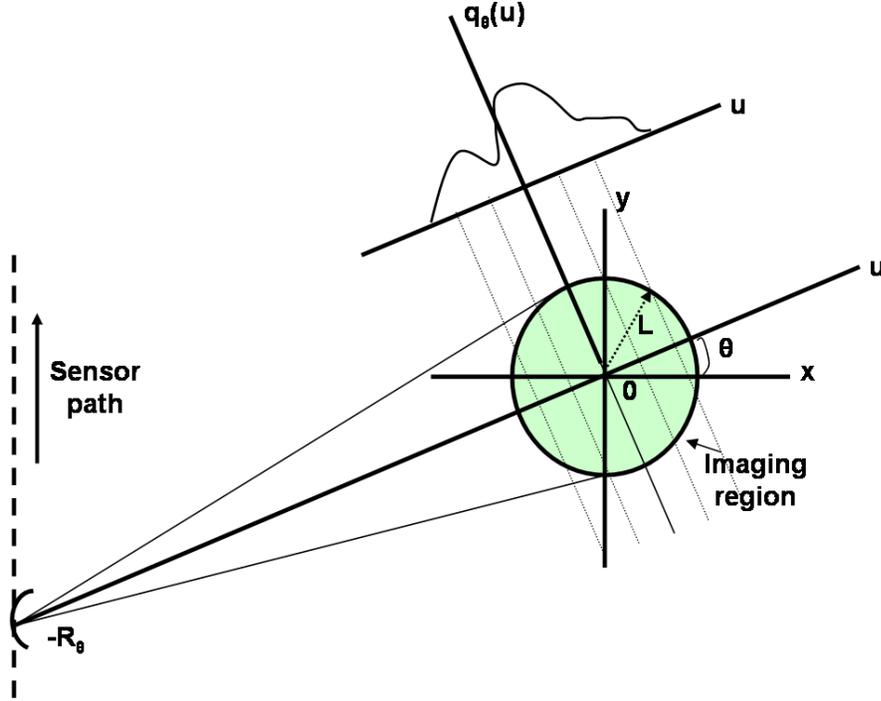
Figure 2.7: Ground plane geometry in spotlight mode SAR.

A radar can transmit linear FM (LFM) chirp pulses described mathematically as the real part of

$$s(t) = \begin{cases} e^{j(\omega_0 t + \alpha t^2)}, & |t| \le \frac{T}{2} \\ 0, & \text{otherwise} \end{cases} \tag{2.37}$$

where $\omega_0, 2\alpha$ and T are the carrier frequency, the chirp rate and the pulse length, respectively. The return signal from a differential area centered on the point (x_0, y_0) at a distance R_0 from the radar will be

$$p(t) = |g(x_0, y_0)| \cos\left[\omega_0\left(t - \frac{2R_0}{c}\right) + \alpha\left(t - \frac{2R_0}{c}\right)^2 + \angle g(x_0, y_0)\right] dx\ dy \tag{2.38}$$

$$= Re\left\{g(x_0, y_0)s\left(t - \frac{2R_0}{c}\right)\right\} dx\ dy \tag{2.39}$$

where $\angle g$ denotes the phase of g, c is the speed of light, $Re\{g\}$ denotes the real part of g and $\frac{2R_0}{c}$ denotes the two way travel time from radar to target. Also, we have neglected the propagation attenuation. Now let's consider points on the ground patch that are equidistant from the radar. These points lie on an arc. Since $R \gg L$,

27

this arc can be approximated as a straight line (see Fig. 2.3.1.4) [54]. Let $q_\theta(u)$ be the line integral given by

$$q_\theta(u) = \int \int_D \delta(u - x\cos(\theta) - y\sin(\theta))g(x,y)dx\ dy \tag{2.40}$$

where $D = \{(x,y)|x^2 + y^2 \leq L^2\}$. This line integral represents a sum of reflectivities at distance $R + u$ to the radar at an angle θ. Using (2.40), and by superposition, the return signal from a differential line of scatterers normal to u axis at $u = u_0$ is given by

$$p_1(t) = Re\left\{q_\theta(u_0)s\left(t - \frac{2(R + u_0)}{c}\right)\right\}du. \tag{2.41}$$

So, the return from the entire ground patch can be approximated by

$$\tilde{p}_\theta(t) = Re\left\{\int_{-L}^{L} q_\theta(u)s\left(t - \frac{2(R + u)}{c}\right)du\right\}. \tag{2.42}$$

Using (2.37), (2.42) can be rewritten as

$$\tilde{p}_\theta(t) = Re\left\{\int_{-L}^{L} q_\theta(u)\exp\left\{j\left(\omega_0\left(t - \frac{2(R + u)}{c}\right) + \alpha\left(t - \frac{2(R + u)}{c}\right)^2\right)\right\}du\right\} \tag{2.43}$$

where

$$t \in \left[\frac{-T}{2} + \frac{2(R + L)}{c}, \frac{T}{2} + \frac{2(R - L)}{c}\right].$$

Mixing $\tilde{p}_\theta(t)$ with the reference chirp signal

$$\exp\left[-j(\omega_0(t - \tau_0) + \alpha(t - \tau_0)^2)\right]$$

where $\tau_0 = \frac{2R}{c}$ and low pass filtering gives [54]

$$\hat{p}_\theta(t) = \int_{-L}^{L} q_\theta(u)\exp\left\{j\frac{4\alpha u^2}{c^2}\right\}\exp\left\{-j\frac{2}{c}(\omega_0 + 2\alpha(t - \tau_0))u\right\}du. \tag{2.44}$$

We will assume that the quadratic phase term $\{j\frac{4\alpha u^2}{c^2}\}$ in (2.44), known as the Residual Video Phase (RVP), is negligible and can be ignored [54]. In cases when the RVP is significant, a pre-processing procedure known as diskewing can be applied to remove this phase effect [58]. Also, we will assume that τ_0 is known. In practice, autofocus techniques are applied when τ_0 is partially known. The observed signal after these approximations and assumptions is given by

28
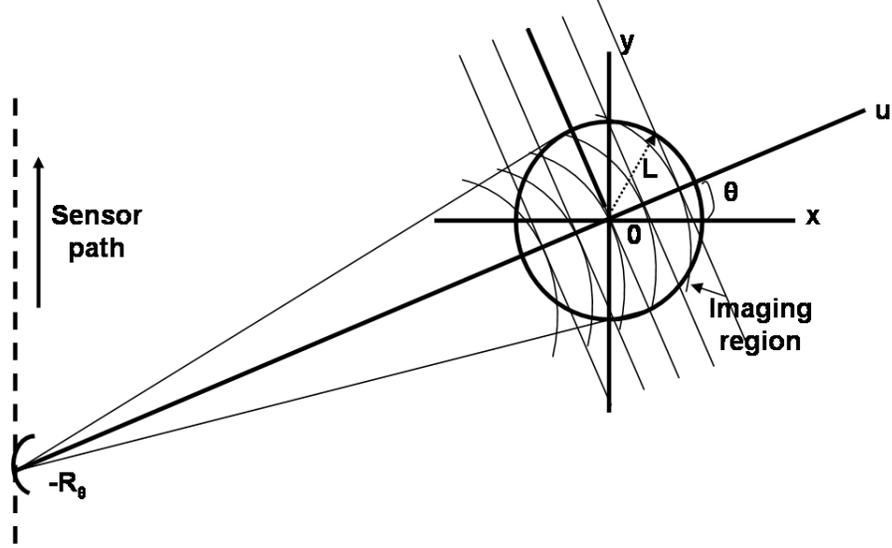
Figure 2.8: Approximation of arcs by straight lines.

$$r_\theta(t) = \int_{-L}^{L} q_\theta(u) \exp\left\{-j\frac{2}{c}(\omega_0 + 2\alpha(t - \tau_0))u\right\} du. \tag{2.45}$$

Let $\Omega(t) = \frac{2}{c}(\omega_0 + 2\alpha(t - \tau_0))$, then (2.45) can be written as

$$r_\theta(t) = \int_{-L}^{L} q_\theta(u) \exp\left\{-j\frac{2}{c}\Omega(t)u\right\} du. \tag{2.46}$$

$r_\theta(t)$ can be identified as a 1D Fourier transform of the projection $q_\theta(u)$. Substituting (2.40) into (2.46) gives

$$r_\theta(t) = \int_{-L}^{L} \int\int_D \delta(u - x\cos(\theta) - y\sin(\theta))g(x,y) \exp\left\{-j\Omega(t)u\right\} \, dx \, dy \, du$$

$$= \int\int_D g(x,y) \exp\left\{-j\Omega(t)[x\cos(\theta) + y\sin(\theta)]\right\} dx \, dy. \tag{2.47}$$

Hence, $r_\theta(t)$ can be viewed as a bandlimited slice at θ from the 2D Fourier transform of the field $g(x,y)$. Taking advantage of a formulation of (2.47) that re-expresses $r_\theta(t)$ as $(\mathcal{A}_\theta g(x,y))(t)$, where $\mathcal{A}_\theta$ is the continuous observation kernel, we can setup a matrix-based formulation of the imaging acquisition problem [57], [56]. Let $r_{\theta_i}(t_j)$ be the fast-time samples at the ith observation angle θ_i at times t_j of the continuous observation $r_{\theta_i}(t)$. Let $\mathbf{r}_{\theta_i}$ be the vector of these samples and A_{θ_i} be a discretized

approximation to the kernel $\mathcal{A}_{\theta_i}$ and $\mathbf{g}$ be a lexicographically ordered vector of unknown sampled reflectivity field of length $\tilde{N}^2$. Then, one can write

$$\begin{bmatrix} \mathbf{r}_{\theta_1} \\ \mathbf{r}_{\theta_2} \\ \vdots \\ \mathbf{r}_{\theta_P} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\theta_1} \\ \mathbf{A}_{\theta_2} \\ \vdots \\ \mathbf{A}_{\theta_P} \end{bmatrix} \mathbf{g} \tag{2.48}$$

or $\mathbf{r} = \mathbf{A}\mathbf{g}$, where $\mathbf{r} = [\mathbf{r}_{\theta_1}, ..., \mathbf{r}_{\theta_P}]^T$, $\mathbf{A} = [\mathbf{A}_{\theta_1}, ..., \mathbf{A}_{\theta_P}]^T$, P is the total number of slow-time samples used to form the image, and $\mathbf{A}$ is of size $P^2 \times \tilde{N}^2$. For the sake of simplicity, we will assume that the number of phase histories and the number of samples per phase history are the same. The collection of phase histories, $\mathbf{r}$, defines a set of samples in the Fourier space of the scene on an annular region, as shown in Figure 2.9. Since this system relates the reflectivity profiles to the measurements directly, polar to rectangular sampling is not required [64]. Assuming the presence of additive measurement noise η, the discrete SAR observation model can be expressed as [57], [64]

$$\mathbf{r} = \mathbf{A}\mathbf{g} + \eta. \tag{2.49}$$

Having observed $\mathbf{r}$ and knowing $\mathbf{A}$, the objective is to recover the complex reflectivity field $\mathbf{g}$. In the following section, we briefly describe some of the commonly used spotlight-mode SAR image formation algorithms.

2.3.1.5 SAR image formation methods

2.3.1.6 Polar Format Algorithm

As discussed earlier, in spotlight-mode SAR, the collection of phase histories, $\mathbf{r}$ provides a set of samples in the Fourier sapce of the scene on a polar wedge. In Polar Format Algorithm (PFA), the collection of phase histories is first interpolated to a Cartesian grid. Then a window such as Hamming or Hanning is applied to control the sidelobes. Finally, a 2D inverse Fourier transform is applied and the magnitude of the image is displayed for viewing [58]. PFA is one of the most commonly used
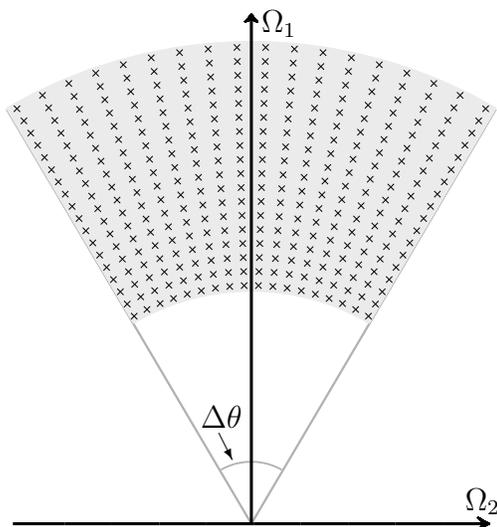
Figure 2.9: Graphical representation of the phase histories on a Polar wedge.

algorithms for SAR image formation.

2.3.1.7 Filtered Backprojection Algorithm

Another popular spotlight-mode SAR image reconstruction method, based on the tomographic formulation of SAR [54], is the Filtered Backprojection algorithm (FBPA) [65]. In this method, each radial slice in the Fourier space is considered as the 1D Fourier transform of the projection of the field at the corresponding angle. The procedure is based on two steps. The first is the filtering of the projections and the second is called backprojection.

2.3.1.8 Reconstruction Methods based on Regulariztion

In [66], iterative reweighted methods such as FOCUSS [51] were applied for SAR image formation. It was argued that such methods can provide reconstructions with reduced clutter, enhanced prominent scatterers and reduced speckle compared to images formed by traditional Fourier methods.

In [57], an estimation theoretic technique based on ℓ_1-norm regularization was presented to obtain the sparse estimate of the reflectivity field. It was shown by

examples that such a method can provide SAR images with reduced sidelobes and improved resolution.

A regularized ℓ_p-norm based feature enhanced SAR image formation method was proposed in [64], [56]. Extensions of half-quadratic regularization methods [67], [68] were used to minimize non-quadratic functionals. The following benefits were claimed for such regularization based method: sidelobe reduction, increased resolution, ease of region segmentation, and speckle suppression.

2.3.1.9   Other Algorithms

Algorithms such as range migration and chirp scaling can also be used for the spotlight-mode SAR image reconstruction [58]. Spectral estimation based methods have also been proposed for SAR image estimation [69].

2.3.2   An Example

In this section, we provide an example of SAR image formation of point scatterers using the PFA. The parameters used in our simulation are summarized in Table 2.1. In Fig. 2.10(a), we show the locations of the point targets used in our simulation. In Fig. 2.10(b) and Fig. 2.10(c), we show the support of the data on a polar wedge and the resulting k-space samples, respectively. The reconstructed SAR image of the point scatterers by using the PFA is shown in Fig. 2.10(d).

Table 2.1: Parameters used in our simulation

Center frequency	3.80×10^9 Hz
Pulse width	1×10^{-6} sec
Bandwidth	1.34×10^8 Hz
Chirp rate	1.34×10^{14} Hz/sec
$\Delta\theta$	$2.1°$
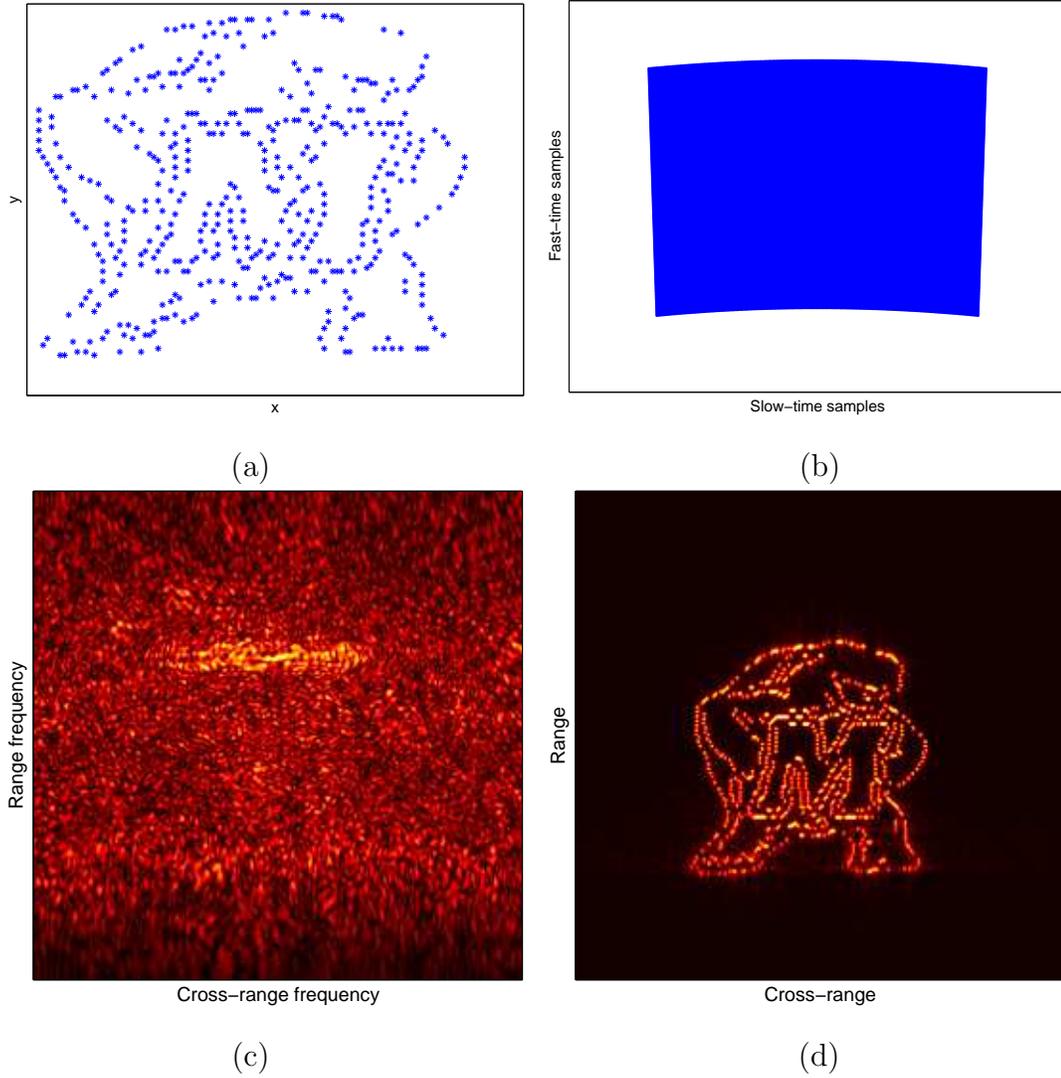
Figure 2.10: (a) Point targets. (b) Corresponding phase histories on a polar wedge. (c) Phase histories in the k-space. (d) Reconstructed image using the PFA.

Chapter 3

Shearlet Approach to Deconvolution

3.1   Introduction

In image restoration, the goal is to best estimate an image that has been degraded. Examples of image degradation include the blurring introduced by camera motion as well as the noise introduced from the electronics of the system. In the case when the degradations can be modelled as a convolution operation, the process of recovering the original image from the degraded blurred image is commonly called deconvolution. The process of deconvolution is known to be an ill-posed problem. Thus, to get a reasonable image estimate, a method of reducing/controlling noise needs to be utilized.

Wavelets are popular for image representation and are used in a wide variety of image processing applications such as compression, and image restoration [70], [71]. The main reason for wavelets' success can be explained by their ability to sparsely represent one-dimensional signals which are smooth away from point discontinuities. By sparse representation, we mean that most of the signal's energy can be captured by a few of the transform coefficients. This is quantified by the decay rate of the nonlinear approximation error. Indeed, it can be shown that the best nonlinear M-term wavelet expansion (reconstruction from the M largest coefficients) for this type of signal has the rate of decay that is the best achievable [72, 2]. It is understood that, higher the decay rate, better the signal estimate from the noisy data will be from that representation.

It is because of this optimality property of wavelet representations that wavelet-based deconvolution routines have been proposed. However, wavelet representations are actually not optimal for all types of images. Specifically, two dimension, if we model images as piecewise smooth functions that are smooth away from a C^2 edge[1],

<hr>

[1]C^2 is the space of functions that are bounded and 2-times continuously differentiable.

the standard 2D wavelets do not reach the best possible rate. In particular, the approximation error for a wavelet representation decays as $O(M^{-1})$ as M increases [2]. As a result, denoising estimates based on 2D wavelets tend to have small unwanted artifacts and complex decision metrics or schemes need to be utilized to try to improve the quality of the estimate. Multi-directional representations such as shearlets [73, 3] provide nearly the optimal approximation rate for these types of images (the optimal rate being $O(M^{-2})$ as M increases [74]) and the corresponding denoising estimates do not suffer from the same types of artifacts [75]. Although related transforms such as contourlets [76], [77], and curvelets [78], [16], [79] share similar properties, in this work we utilize properties unique to an implementation of the shearlet transform that offer advantages for the purpose of deconvolution.

The concept of using a sparse representation to achieve good estimates for deconvolution has been suggested before (see for example [80] and [81]). However, particular features concerning implementations of such representations that contribute to performance presented here have not been previously considered. Our shearlet-based deconvolution has the unique ability for a multi-scale and anisotropic regularization inversion to be done before noise suppression. Furthermore, for a given regularization parameter its adaptive noise suppression surpasses similar schemes. This is an important consideration since in some case it may not be possible to find the optimal regularization parameter.

In the implementation stage, to deal with boundary effects, some concepts in the literature have centered around the idea of noise shrinkage either before or after the application of the deconvolution procedure (see [80] and [82]). However, to carry out such schemes effectively, one needs a transform that can be implemented in a non-recursive formulation as is done in this work with the shearlet transform. Otherwise, error estimates made by one set of coefficients will highly influence estimates made on a different but dependent set of coefficients. In addition, to be effective in regularizing the approximate deconvolution process, a nonsubsampled (redundant) transform should be utilized. This redundancy not only provides for more effective measurements based on the use of auxiliary functions such the GCV function, but it

also greatly aids in estimation [77, 1, 83, 84]. As will become apparent, these desired features are obtained by using an M-channel shearlet transform implementation [10], [4].

3.1.1 The Image Deconvolution Problem

Since a digitally recorded image is a finite discrete data set, an image deconvolution problem is formulated as a matrix inversion problem. Without loss of generality, assume the recorded arrays are of size $N \times N$. Let γ denote an $N \times N$ array of samples from a zero mean additive white Gaussian noise (AWGN) with variance σ^2. Given the $N \times N$ arrays y and x, representing the observed image and the image to be estimated, respectively, the matrix deconvolution problem can be described as

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\gamma}, \tag{3.1}$$

where $\mathbf{y}$, $\mathbf{x}$, and $\boldsymbol{\gamma}$ are $N^2 \times 1$ column vectors representing the arrays y, x, and γ lexicographically ordered, and H is the $N^2 \times N^2$ matrix that models the blur operator. In the case when H is a block-circulant-circulant-block matrix [85], the problem can be described as

$$y(n_1, n_2) = (x \circledast h)(n_1, n_2) + \gamma(n_1, n_2), \tag{3.2}$$

where $0 \leq n_1, n_2 \leq N - 1$, $\circledast$ denotes circular convolution, and h denotes the point spread function (PSF) of a linear space-invariant system. Equation (3.2) in the discrete Fourier transform (DFT) domain can be written as

$$Y(k_1, k_2) = \mathcal{H}(k_1, k_2)X(k_1, k_2) + \Gamma(k_1, k_2), \tag{3.3}$$

where $Y(k_1, k_2), \mathcal{H}(k_1, k_2), X(k_1, k_2)$ and $\Gamma(k_1, k_2)$ are the 2D DFTs of y, h, x, and γ, respectively, for $-N/2 \leq k_1, k_2 \leq N/2 - 1$. The conditioning of this system is determined by the ratio of the largest to smallest magnitudes of the $\mathcal{H}$ values. Typically, $|\mathcal{H}(k_1, k_2)|$ contains values at or near zero which makes the system ill-conditioned.

In general, to regularize the inversion of the convolution operator, a representation that diagonalizes the convolution operator (matrix) is needed in order to appropriately control the approximation. In particular, if H is a block-circulant-circulant block matrix, H is diagonalizable by a Fourier basis. This means that an estimate of the image can be found by filtering the diagonal components of the Fourier diagonalization of H in order to approximately invert H. For instance, let $\Upsilon(k_1, k_2)$ be a filter that is nearly one when $H(k_1, k_2)$ is large, is small when $H(k_1, k_2)$ is nearly zero, and such that $\Upsilon(k_1, k_2)\mathcal{H}(k_1, k_2)^{-1}$ is defined everywhere. Then, a Fourier-based estimate can be given as $\Upsilon(k_1, k_2)\mathcal{H}(k_1, k_2)^{-1}Y(k_1, k_2)$. This in turn means that the image is estimated from a Fourier representation.

However, if our image is considered as a piecewise smooth function that is smooth away from a C^2 edge, then the decay rate of the nonlinear approximation from a Fourier representation is $O(M^{-1/2})$ as M increases. Yet, for this type of image the decay rate of the nonlinear approximation from a wavelet representation is $O(M^{-1})$ as M increases. This means that estimating the image from the perspective of removing noise, a wavelet based estimate would perform better than the one from a Fourier basis. In short, the ability to get a good estimate depends on balancing a representation that is effective for regularizing the inversion of the convolution operator and a representation that is effective for estimating the image from colored noise by means of the approximate inversion of the operator. (See [81] for further discussion).

3.1.2 Historical Perspective

Deconvolution methods can be separated into two major categories: direct and iterative.

Direct Methods. Some of these methods are based on filtering the singular value decomposition (SVD) such as Tikhonov, truncated SVD (TSVD), and Wiener filtering [86]. Increased performance of such direct methods can be attributed to the inclusion of the wavelet-based estimators. One such technique called the Wavelet-Vaguelette deconvolution (WVD) was proposed in [80]. In this work, functionals

called *vaguelettes* are used to simultaneously deconvolve and compute the wavelet coefficients. However, the scheme does not provide good estimates for all convolution operators. To overcome this limitation, Kalifa *et al.* proposed a wavelet packet based method that matches the frequency behavior of certain convolution operators [87]. Additional wavelet-based techniques have been proposed in [82], [88], [89], and [90].

An improved hybrid wavelet-based regularized deconvolution algorithm that works with any ill-conditioned convolution system was developed in [1]. This *Fourier-Wavelet Regularized Deconvolution* (ForWaRD) method employs Fourier-domain regularized inversion followed by wavelet-domain noise shrinkage to minimize the distortion of spatially localized features in the image. An extension in terms of curvelets, known as ForCuRD, was proposed in [91].

Iterative Methods. Some of the better-known basic iterative methods are the Conjugate Gradient algorithm [86], Richardson-Lucy [92], [93], and Landweber [94]. Many extensions and improvements over these methods have been made that include the use of wavelets, or other sparse representations such as curvelets. Some of these are [95, 83, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105]. Additional techniques may be found within these references. Note some of the these promising techniques make use of one-norm transform-domain sparsity promotion. Such methods seem to retain edge information well.

Among the direct methods, the local polynomial approximation (LPA) algorithm [106] outperformed some of the best existing deconvolution methods such as [1] and [107] in terms of improvement in signal-to-noise-ratio and was established to be state-of-the-art. In this work, we will only focus on direct methods for comparison because some applications may desire direct methods, and also since some of these iterative methods can use the estimates provided from such techniques as the initial starting point.

3.2 Generalized Cross Validation (GCV) for Shearlet Thresholding

In this section, we describe a shearlet-thresholding scheme based on a GCV

function for the purpose of noise reduction [108]. One of the major advantages of this GCV method is that it obtains nearly the optimal thresholding without knowing the noise variance. It depends only on the data and automatically adjusts the shrinkage parameter according to the data. A similar GCV method for wavelet thresholding has been proposed in [109, 110, 111]. Note that, although we are suggesting the use of a GCV function, it is also possible to adapt the new SURE approach [112] for this task.

Suppose

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\gamma}, \tag{3.4}$$

where the vectors $\mathbf{y}$, $\mathbf{x}$ and $\boldsymbol{\gamma}$ represent respectively the observation, the original image and the colored noise that is assumed to be second order stationary (i.e. the mean is constant and the correlation between two points depends only on the distance between them). Corresponding to a threshold τ, define the soft-threshold function $T_\tau(x)$ to be equal to $x - \tau \mathrm{sign}(x)$ if $|x| > \tau$ and zero otherwise. We will show that nearly optimal threshold values $\tau_{j,\ell}$ can be obtained by finding the values minimizing a GCV function which is dependent on each scale j and direction ℓ.

Just as in the case of wavelets, to obtain results similar to those in [111], it is not necessary for the shearlet coefficients to be uncorrelated at any moment; however, it is necessary that the noise be second order stationary [110]. If the noise process $\boldsymbol{\gamma}$ is stationary, then using the multi-scale and multi-directional structure of shearlets, we obtain the following lemma.

Lemma 3.2.1. *If $\langle \boldsymbol{\gamma}, \psi_{j,\ell,k} \rangle$ represents a shearlet coefficient of a random vector $\boldsymbol{\gamma}$ at scale j, direction ℓ, and location k, then the variance of this coefficient, $E\left[|\langle \boldsymbol{\gamma}, \psi_{j,\ell,k} \rangle|^2\right]$, depends only on the scale j and direction ℓ.*

Proof: It follows from the fact that we are using a filter bank with appropriate directional filters and if $\boldsymbol{\gamma}$ is a discrete stationary random process which is an input to a shift-invariant filter, $g_{j,\ell}$, corresponding to scale j and direction ℓ then the output is a convolution of $\boldsymbol{\gamma}$ with $g_{j,\ell}$ which is also stationary [113]. ∎

By Lemma 3.2.1, the shearlet transform of stationary correlated noise is stationary within each scale and directional component. We can use this property to choose a different threshold for each resolution and directional component. If $\mathbf{y}_{j,\ell}$ represents the vector of shearlet coefficients of $\mathbf{y}$ at scale j and direction ℓ, then we can write

$$R(\tau) = \sum_j \sum_\ell \frac{L_{j,\ell}}{L} R_{j,\ell}(\tau_{j,\ell}), \tag{3.5}$$

where $L_{j,\ell}$ is the number of shearlet coefficients on scale j and direction ℓ, L is the total number of shearlet coefficients, and

$$R_{j,\ell}(\tau_{j,\ell}) = \frac{1}{L_{j,\ell}} \|T_\tau(\mathbf{y}_{j,\ell}) - \mathbf{x}_{j,\ell}\|^2. \tag{3.6}$$

Since, all the components in (3.5) are positive, minimizing the mean squared error or risk function $R(\tau)$ is equivalent to minimizing $R_{j,\ell}(\tau_{j,\ell})$ for all j and ℓ. An argument similar to that used in [111] leads to the following GCV functions:

$$GCV_{j,\ell}(\tau_{j,\ell}) = \frac{\frac{1}{L_{j,\ell}} \| T_\tau(\mathbf{y}_{j,\ell}) - \mathbf{y}_{j,\ell} \|^2}{\left[\frac{L_{j,\ell,0}}{L_{j,\ell}}\right]^2}, \tag{3.7}$$

where $L_{j,\ell,0}$ is the total number of shearlet coefficients that were replaced by zero. We now have the following result:

Theorem 3.2.2. *The minimizer of $GCV_{j,\ell}(\tau_{j,\ell})$ is asymptotically optimal for the minimum risk threshold $R_{j,\ell}(\tau_{j,\ell})$ for scale j and directional component ℓ.*

Thus, by using the values $\tau_{j,\ell}$ that minimize $GCV_{j,\ell}$ for each j and ℓ, a shearlet-based denoised estimate will likely be close to the ideal non-noise corrupted image.

An important feature about the use of our non-subsampled shearlet transform implementation is that it facilitates the use of asymptotic methods such as those based on GCV functions. A subsampled transform implementation would cause the number of coefficients to decrease as the levels progress, so that thresholds found by minimizing the individual GCV functions will be less likely to correspond to the actual threshold values that minimize the risk functions for each frequency band.
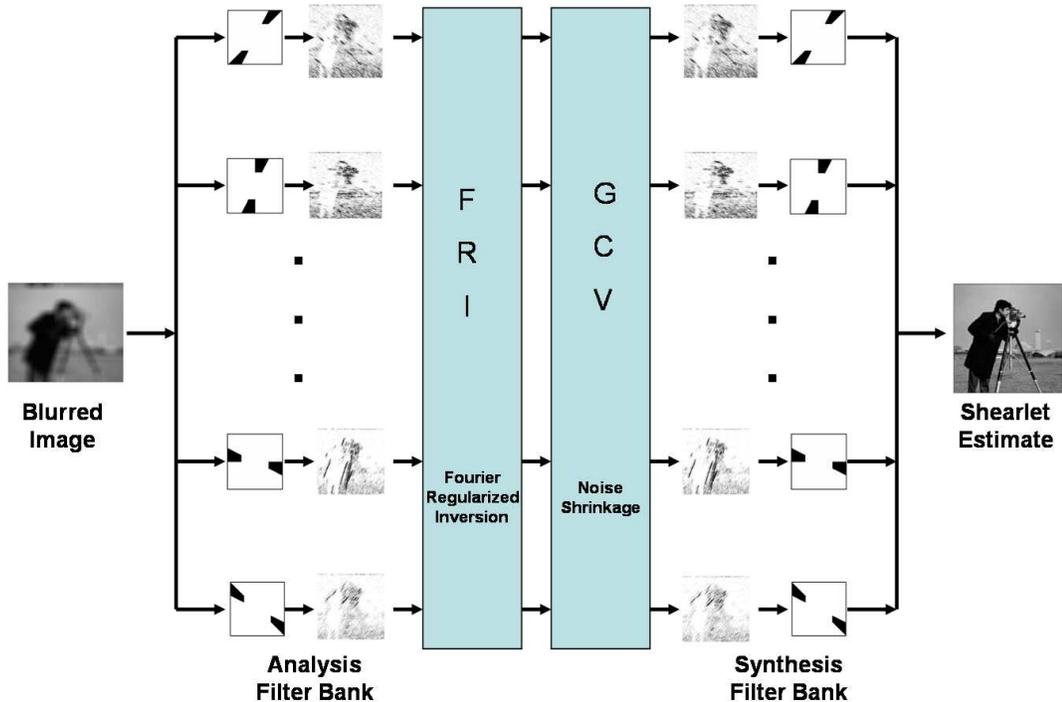
Figure 3.1: Fourier-shearlet regularized deconvolution.

3.3   Shearlet-based Deconvolution

Having established a method for obtaining a good image estimate when the image is corrupted by colored noise, let us now focus on how we are to use this method as part of a deconvolution routine. Since our blurring model is described by (3.2), a suitable pseudo-inverse estimate can be found by regularizing the convolution operator from a discrete Fourier basis. Using the regularized inverse operator

$$H_\alpha(k_1, k_2) = \frac{\overline{\mathcal{H}}(k_1, k_2)}{\mid \mathcal{H}(k_1, k_2) \mid^2 + \alpha} \tag{3.8}$$

for some regularizing parameter $\alpha \in R^+$, an image estimate in the Fourier domain is given by

$$X_\alpha(k_1, k_2) = Y(k_1, k_2) H_\alpha(k_1, k_2), \tag{3.9}$$

for $-N/2 \leq k_1, k_2 \leq N/2 - 1$. This type of regularization applied is often referred to as Tikhonov-regularization [114]. When an estimate of the power spectral density

(PSD) can be accurately determined from a method such as that proposed in [115], a Wiener-based solution can be found by using

$$H_\alpha(k_1, k_2) = \frac{\overline{\mathcal{H}}(k_1, k_2)}{\mid \mathcal{H}(k_1, k_2) \mid^2 + \alpha \frac{\sigma^2}{P_{\hat{x}}(k_1, k_2)}}, \tag{3.10}$$

where $\alpha \in \mathbb{R}^+$ and $P_{\hat{x}}(k_1, k_2)$ is the estimated PSD of the image for $-N/2 \leq k_1, k_2 \leq N/2 - 1$.

Taking advantage of the shearlet decomposition, we can adaptively control the regularization parameter to be the best suited of each frequency supported trapezoidal region. Let $G_{j,\ell}$ denote the DFT of the shearlet filter $g_{j,\ell}$ for a given scale j and direction ℓ. The shearlet coefficients of an estimate of the image for a given regularization parameter $\alpha_{j,\ell}$ can be computed in the Fourier domain as

$$X_{\alpha_{j,\ell}}(k_1, k_2) = Y(k_1, k_2)G_{j,\ell}(k_1, k_2)H_{\alpha_{j,\ell}}(k_1, k_2) \tag{3.11}$$

for $-N/2 \leq k_1, k_2 \leq N/2 - 1$.

The remaining aspect of the deconvolution problem is transformed into a denoising problem in the presence of colored noise. This can be dealt with by thresholding the estimated shearlet coefficients using the GCV determined previously without having to know the noise variance explicitly. An important advantage in using the GCV is that after the Fourier regularized inversion (FRI), the method automatically adjusts the shrinkage parameter according to the data. We summarize the shearlet based deconvolution method in Figure 3.1.

Let $\widehat{X}_\alpha$ denote the result of the inversion of the shearlet transform after the shearlet coefficients of X_α have been thresholded. We want to choose an α that minimizes the shearlet-based mean-squared-error (MSE) $\|x - \hat{x}\|_2^2$. However, since x is unknown, α is chosen to be a minimizer of the cost function

$$C(\alpha) = \sum_{k_1} \sum_{k_2} \frac{|\mathcal{H}(k_1, k_2)|}{|\mathcal{H}(k_1, k_2)|^2 + \eta} \left| \mathcal{H}(k_1, k_1)\widehat{X}_\alpha(k_1, k_2) - Y(k_1, k_2) \right|^2 \tag{3.12}$$

where $\eta = N^2\sigma^2/\|y - \mu(y)\|_2^2$, $\mu(y)$ is the mean of y, and the sum is taken over all values of k_1 and k_2 from $-N/2$ to $N/2 - 1$ inclusive. In other words, we choose α such that the estimate agrees with the observation y based on weighing

$|\mathcal{H}(k_1, k_1)\widehat{X}_\alpha(k_1, k_2) - Y(k_1, k_2)|^2$ counter-balanced by an approximation to $1/\mathcal{H}(k_1, k_2)$ for each frequency index (k_1, k_2). This is just an extension (shearlet-based estimate vs. wavelet-based estimate) of the cost function originally suggested in [1]. This optimization function assumes the noise variance σ^2 to be known. This is not a problem since the noise variance can easily be estimated by using a median estimator on the finest scale of the wavelet coefficients of y [2], [71].

In this case, there is a great advantage in using the GCV for the shearlet thresholding as the variance of the colored noise at each location and scale dependent on α does not have to be estimated. In addition, a GCV-based thresholding routine produces better results over schemes based on estimating the standard deviation of the noise throughout the decomposition.

If we define $Y_{j,\ell}(k_1, k_2) = Y(k_1, k_2)G_{j,\ell}(k_1, k_2)$ for $-N/2 \leq k_1, k_2 \leq N/2 - 1$, then the optimal $\alpha_{j,\ell}$ for each thresholded shearlet coefficient $\widehat{X}_{\alpha_{j,\ell}}$ can be found by minimizing the cost function $K(\alpha_{j,\ell})$ equal to

$$\sum_{k_1} \sum_{k_2} \frac{|\mathcal{H}(k_1, k_2)|}{|\mathcal{H}(k_1, k_2)|^2 + \eta} \left| \mathcal{H}(k_1, k_1)\widehat{X}_{\alpha_{j,\ell}}(k_1, k_2) - Y_{j,\ell}(k_1, k_2) \right|^2, \qquad (3.13)$$

where $\eta = N^2\sigma^2/\|y_{j,\ell} - \mu(y_{j,\ell})\|_2^2$ and $y_{j,\ell}$ is the inverse DFT of $Y_{j,\ell}$.

The use of this optimization function to find α has been shown to be satisfactory with many of the examples tested. The L-curve method could be adopted to estimate α (see [86] for details on the method). This could prove to be more reliable and does not require any estimate of the noise variance. It is also possible to use the optimization function $\|H\widehat{\mathbf{x}}_\alpha - \mathbf{y}\|_2^2/(trace(I - H\tilde{H}_\alpha))$ where $\tilde{H}_\alpha$ denotes the regularized Tikhonov inverse of H and I denotes the identity matrix. Such an optimization function weighs the fidelity of the estimate $\widehat{x}_\alpha$ against the given data y and inversely weighs it against a measure of how far away the regularized inverse operator is from an idealized inversion operator. This generalized cross validation function can be derived by using similar arguments to that given in [108].

We summarize the main steps of the shearlet-based deconvolution algorithm as follows:

Given $\alpha_{j,\ell}$, for some j and ℓ.

- Use the shearlet filter $G_{j,\ell}$ and apply the regularized filter (3.9) or (3.10) to Y to obtain $X_{\alpha_{j,\ell}}$.

- Apply the GCV based shearlet shrinkage to $x_{\alpha_{j,\ell}}$ to obtain $\widehat{x}_{\alpha_{j,\ell}}$.

Repeat process for a different value of $\alpha_{j,\ell}$ until $K(\alpha_{j,\ell})$ is minimized for each j and ℓ.

After each shearlet coefficient $\widehat{x}_{\alpha_{j,\ell}}$ that minimizes $K(\alpha_{j,\ell})$ is found, form the final estimate by applying the inverse shearlet transform.

The values $\alpha_{j,\ell}$ that minimize K can be found by using either a sequence of possible values or by using a minimization routine.

Although we have described the most general case of regularizing each shearlet coefficient separately, in some cases for efficiency it may be preferred to use a common regularization parameter α. In such a case, the algorithm is implemented using the coast function $C(\alpha)$ instead of $K(\alpha_{j,\ell})$.

3.4   Experimental Results

In this section, we present the results of our proposed algorithm and compare them with some of the recent multiscale wavelet and wavelet-like deconvolution methods described in [1, 106] and [91][2]. In these experiments we use the improvement in signal-to-noise-ratio (ISNR) to measure the performance of the routines tested using the images shown in Figure 3.2. The ISNR is defined as

$$ISNR = 10 \log_{10} \left(\frac{\|x - y\|_2^2}{\|x - \hat{x}\|_2^2} \right). \tag{3.14}$$

---

[2]The implementation of the ForWaRD algorithm is available at www.dsp.rice.edu/software and the implementation of the LPA-ICI algorithm is available at www.cs.tut.fi/∼lasip/2D.

For an image of size $m \times n$, the BSNR is defined in decibels as

$$BSNR = 10 \log_{10} \left(\frac{\|(x \circledast h) - \mu(x \circledast h)\|_2^2}{mn\sigma^2} \right), \qquad (3.15)$$

where $\mu(x \circledast h)$ denotes the mean of $x \circledast h$.

(a)   (b)

(c)   (d)

Figure 3.2: Images used for different experiments. (a) *Cameraman* image, (b) *Barbara* image, and (c) *Lena* image, (d) *Peppers* image.

For all shearlet transform implementations, we used 1, 8, 8, 16, and 16 directions in the scales from coarse to fine which corresponds to the same decomposition tested in [3]. We apply the GCV based shrinkage to the outputs from each of the 48 filters except the output corresponding to the coarsest scale. Experiments have

shown that increasing the number of directions every scale usually results in better estimates.

In the case when wavelets are used for image denoising, it was shown in [116] that a Wiener-based wavelet shrinkage filter typically improves upon the mean square error performance over that of hard/soft thresholding. By Wiener-based shrinkage, we mean to weigh the wavelet coefficients $w_{j,l}$ as $w_{j,l}|w'_{j,l}|^2/(|w'_{j,l}|^2 + \beta_{j,\lambda})$, where $w'_{j,l}$ are the wavelet coefficients from another denoised estimate, and $\beta_{j,\lambda}$ are scale-dependent regularization parameters. The performance of the proposed method is improved by using a similar Wiener shrinkage filter. In this case, the shearlet coefficients with a slightly different decomposition (three decomposition levels) are filtered using the initial shearlet-based estimate. Several experiments have shown that the final estimate is mostly driven by how successful the initial estimate is, so that even the use of a wavelet decomposition instead of an alternate shearlet decomposition can provide just as an effective estimate.

In the first set of tests, we consider the setup of [1], where a *Cameraman* image is blurred by a 9×9 uniform box-car blur. The AWGN variance, σ^2, is chosen with a BSNR of 40 dB. A comparison of different methods in terms of ISNR is shown in Table 7.1 under the Experiment 1 column. The shearlet-based method yields a value 7.89 dB which is better than the values obtained by any of the other methods.

In Figure 3.3, we display a few of the $GCV_{j,\ell}$ and $R_{j,\ell}$ curves for different scales and directions obtained from Experiment 1. In [1], after the Fourier shrinkage, the leaked colored noise variance was estimated at each scale and was used to shrink the wavelet coefficients. Similarly, we can estimate the colored noise variance, $\sigma_{j,\ell}$, at different scales and directions at the output of the shearlet filter bank as follows:

$$\sigma_{j,\ell}^2 = \sigma^2 \parallel H_{\alpha_{j,\ell}}(k_1, k_2) G_{j,\ell}(k_1, k_2) \parallel_2^2$$
$$= \sum_{k_1} \sum_{k_2} \sigma^2 |H_{\alpha_{j,\ell}}(k_1, k_2) G_{j,\ell}(k_1, k_2)|^2.$$

The threshold values $\lambda_{j,\ell}$ are determined by $\lambda_{j,\ell} = \kappa_{j,\ell} \sigma_{j,\ell}$, where $\kappa_{j,\ell}$ is a scale and direction dependent threshold [1],[2]. For comparison, the estimated $\lambda_{j,\ell}$ are plotted in Figure 3.4 along with the actual minimum values obtained by minimizing

46

the $GCV_{j,\ell}$ and $R_{j,\ell}$ functions for experiment 1. Figures 3.3 and 3.4 indicate that both $GCV_{j,\ell}$ and $R_{j,\ell}$ have approximately the same minimum values. However, in some cases the estimated $\lambda_{j,\ell}$ values are very different than the values obtained by minimizing the $GCV_{j,\ell}$ and $R_{j,\ell}$ curves.

(a) (b) (c)
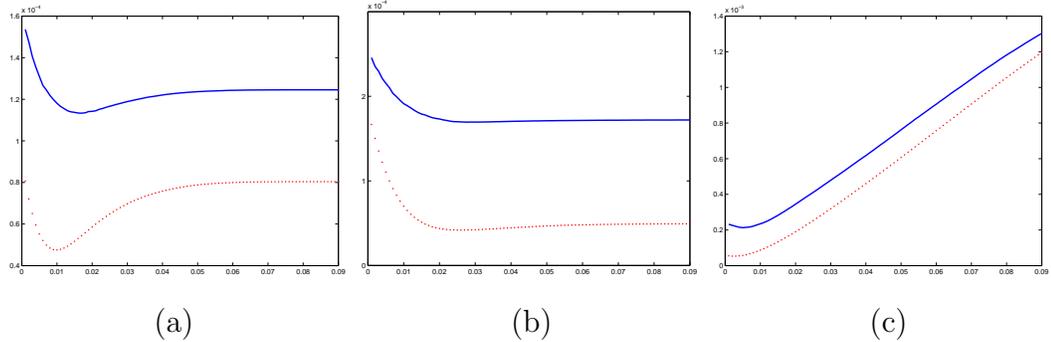
Figure 3.3: $R_{j,\ell}$ and $GCV_{j,\ell}$ as functions of the threshold τ for different scales and directions from Experiment 1. Solid (blue) line corresponds to the $GCV_{j,\ell}$ and dotted (red) line corresponds to $R_{j,\ell}$. (a) $j = 2, \ell = 13$, (b) $j = 1, \ell = 4$, (c) $j = 3, \ell = 6$.

In Figure 3.5, we show the signal-to-noise-ratio (SNR) performance of the shearlet-based deconvolution (black-line) compared to ForWaRD (dotted-line) and ForCuRD (gray-line) as a function of blur SNR (BSNR). In this illustration, we used the 9×9 box-car blur on the *Lena* image shown in Figure 3.2(c). As explained previously, an estimate based on a shearlet or curvelet decomposition decays faster than that of an estimate based on a wavelet decomposition as a function of noise level for images that are smooth away from C^2 edges. Figure 3.5 displays a similar correspondence in decay rates for the proposed shearlet-based deconvolution and the ForCuRD scheme over the wavelet-based deconvolution scheme (ForWaRD). Since the performance is measured in terms of SNR instead of MSE, it is expected that shearlet and curvelet-based estimates will decay slower as a function of noise level given in terms of BSNR.

In the second set of experiments performed over the *Cameraman* image, we replicate the experimental setup of [107]. The point spread function of the blur operator is given by: $h(n_1, n_2) = (1 + n_1^2 + n_2^2)^{-1}$, for $n_1, n_2 = -7, ..., 7$, and the

(a)

(b)

(c)

(d)
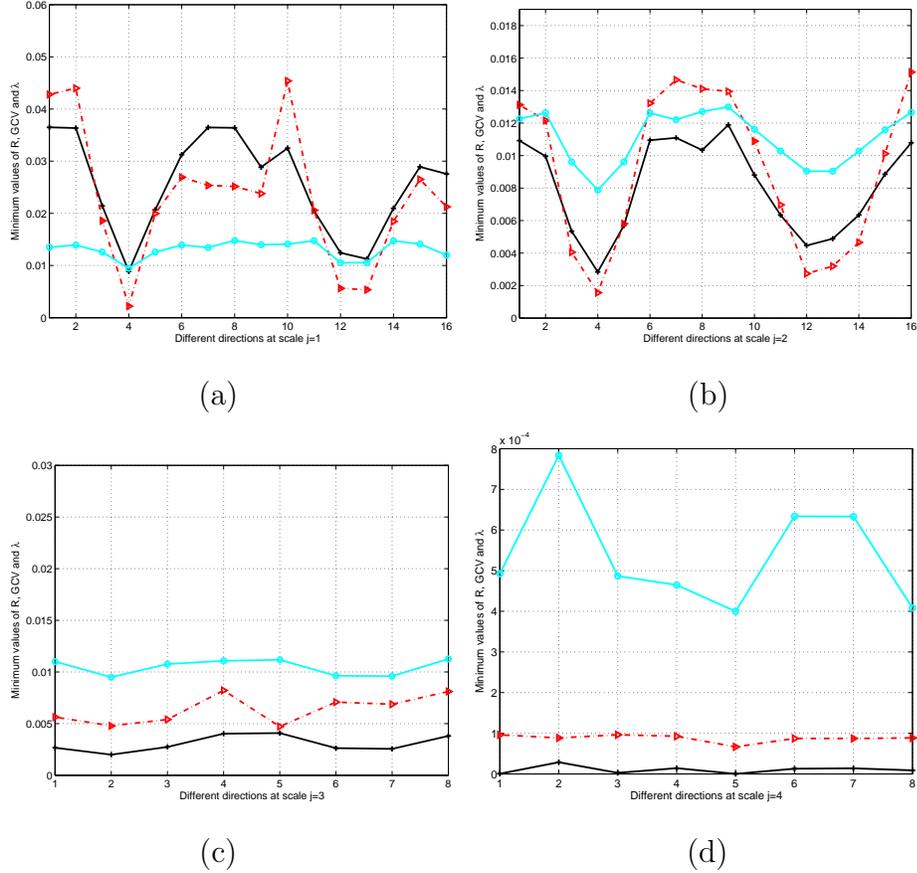
Figure 3.4: Minimum values obtained by minimizing $R_{j,\ell}$ (black), and $GCV_{j,\ell}$ (dot-dash) values along with λ (gray) values obtained by the method described in [1], for different scales and directions from Experiment 1. As can be seen from the figure, minimum values of $GCV_{j,\ell}$ are approximately equal to the minimum values of $R_{j,\ell}$. (a) At scale $j = 1$, there are 16 directions. (b) At scale $j = 2$, there are 16 directions. (c) At scale $j = 3$, there are 8 directions. (d) At scale $j = 4$, there are 8 directions.
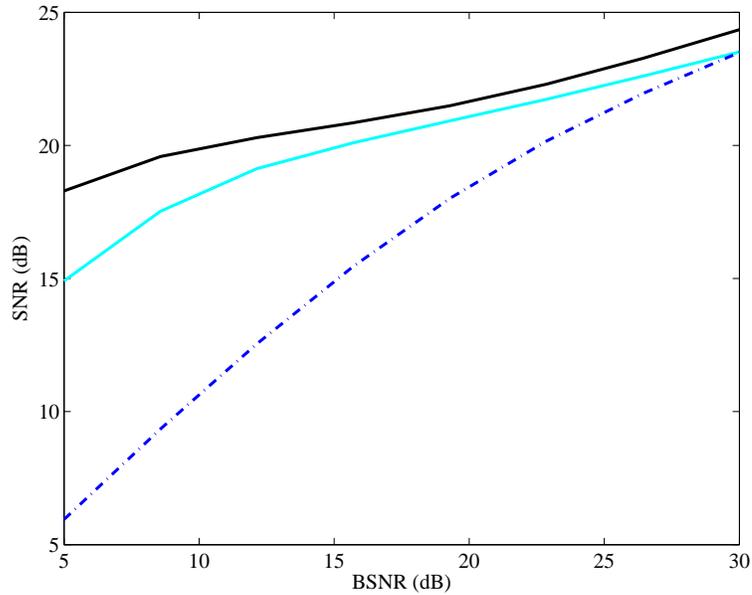
Figure 3.5: SNR performance of ShearDec (black-line) compared to ForCuRD (gray-line) and ForWaRD (dotted-line) as a function of BSNR.

noise variances are $\sigma^2 = 2$ and $\sigma^2 = 8$. The SNR improvements are summarized in Table 7.1 under the Experiment 2 and Experiment 3 columns, for $\sigma^2 = 2$ and $\sigma^2 = 8$, respectively. Again, the shearlet-based deconvolution algorithm outperforms the other methods in terms of ISNR.

In the third set of tests, the original image of *Lena* is blurred by a Gaussian PSF defined as

$$h(i,j) = Ke^{-\frac{i^2+j^2}{2\eta^2}},$$

Table 3.1: ISNR for different experiments.

Experiments

Methods	1	2	3	4	5	6
ShearDec	7.89	7.55	5.67	5.84	1.74	5.23
LPA-ICI	7.84	7.31	5.54	5.77	1.00	4.17
ForCuRD	7.28	6.84	5.09	5.66	1.03	4.53
ForWaRD	7.40	6.75	5.07	5.60	1.12	3.91

for $i, j = -5, ..., 5$, where K is a normalizing constant ensuring that the blur is of unit mass, and η^2 is the variance that determines the severity of the blur. In this experiment we chose $\eta = 2$ and the noise variance, σ^2, with a BSNR of 40 dB. We report the simulation results under the Experiment 4 column of Table 7.1. Again, our proposed method outperforms the best performing methods known for this problem setup.

In the fifth experiment, we use the blur filter considered in [106]. The original image of *Barbara* is blurred by a 5×5 separable filter with weights $[1, 4, 6, 4, 1]/16$ in both the horizontal and vertical directions and then contaminated with AWGN with $\sigma = 5$. The details of the images obtained by the different methods are shown in Figure 3.7. Again, the shearlet-based algorithm performs the best in terms of ISNR and captures the details better than any of the other methods.

In light of the robustness to noise of the shearlet-based method showed in Figure 3.5, we replicated the set-up similar to that given in [1]. In this case, the *Peppers* image is blurred by a 9×9 uniform box-car blur and the AWGN added is such that the BSNR=30 dB. We tested the ForWaRD method (ISNR = 3.91 dB), the For-CuRD method (ISNR = 4.53 dB), the LPA-ICI method (ISNR = 4.17 dB), and the shearlet-based method (ISNR = 5.23 dB) using the same regularization parameter α (not necessarily optimal) for each routine (experiment 6). Close-ups of some of the results are shown in Figure 3.9. The Fourier regularized inversion estimate used in all three algorithms is shown in Figure 3.9 (c). This experiment presents an important comparison in robustness to noise suppression and an indication of the shearlet-based algorithm's high default tolerance level when the regularization parameter is not chosen optimally.

Plots of the validation functions $C(\alpha)$ and $\|H\widehat{\mathbf{x}}_\alpha - \mathbf{y}\|_2^2/(trace(I - H\tilde{H}_\alpha))$ are shown for various values of α in Figure 3.8. For comparison, a plot of $\|\widehat{x}_\alpha - x\|_2^2$ is also given. In the experiment set up for this comparision, we used the *Peppers* image blurred by a 9×9 uniform box-car blur and the AWGN added with corresponding

(a)

(b)

(c)

(d)
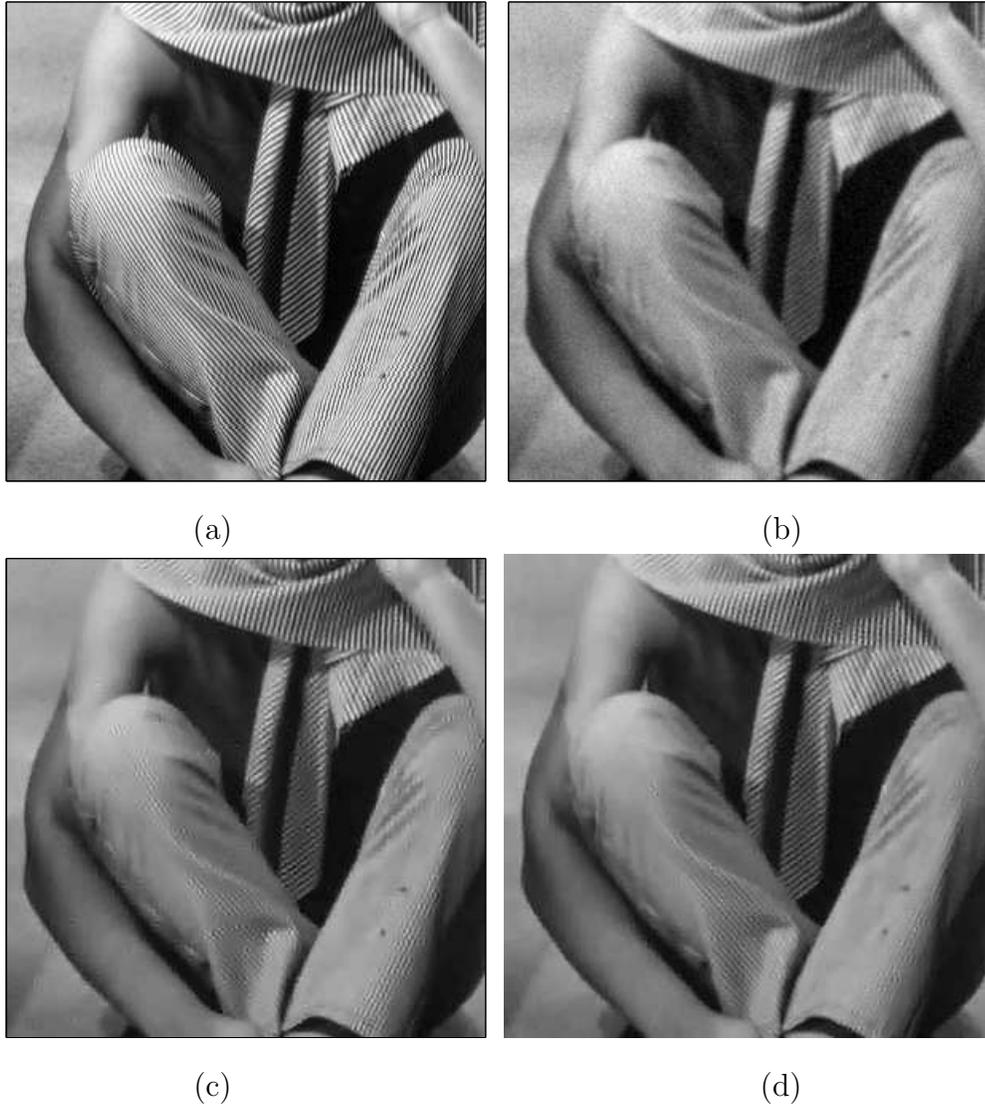
Figure 3.6: Details of the image deconvolution experiment with a *Barbara* image.
(a) Original image. (b) Noisy blurred image, $\sigma = 5$ dB. (c) ForWaRD estimate,
ISNR = 1.12 dB. (d) ForCuRD estimate, ISNR= 1.03 dB.

<div align="center">(e)   (f)</div>
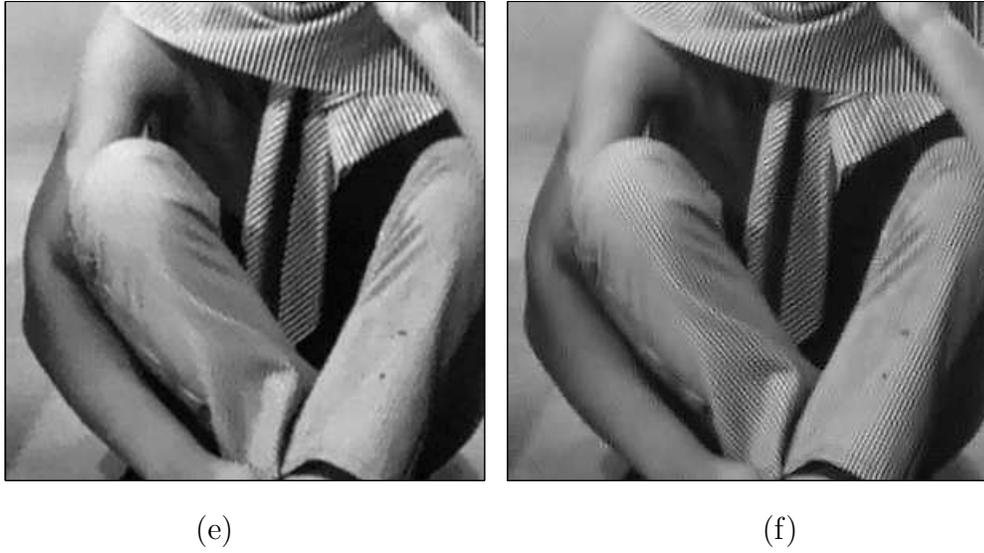
Figure 3.7: Details of the image deconvolution experiment with a *Barbara* image. (e) LPA-ICI algorithm, ISNR = 1.0 dB. (f) shearlet-based estimate, ISNR=1.74 dB.
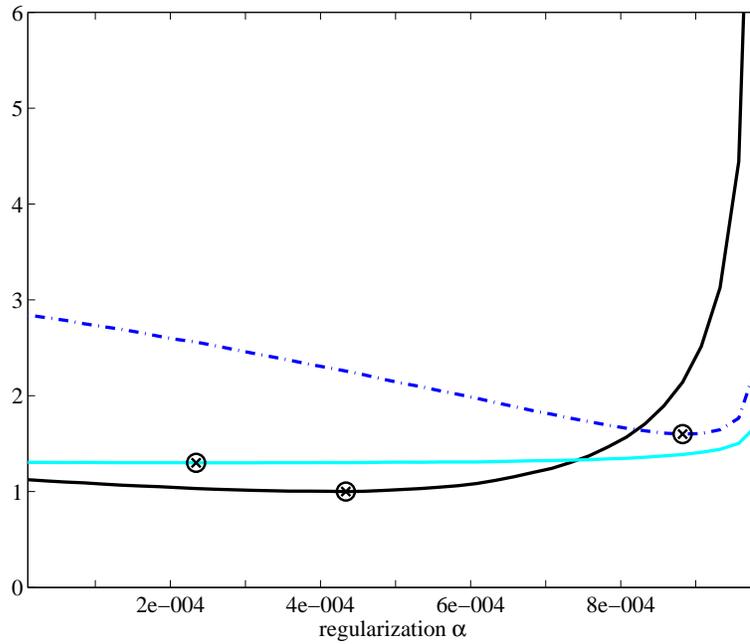
Figure 3.8: Plots of $\|H\widehat{\mathbf{x}}_\alpha - \mathbf{y}\|_2^2/(trace(I - H\tilde{H}_\alpha))$, $C(\alpha)$, and $\|\widehat{x}_\alpha - x\|_2^2$ for various values of α are displayed as a black line, a gray line, and a dash-dotted line, respectively. The output values have been rescaled for illustration purposes. The locations of their minimal value are marked with a circled 'X'.

<div align="center">52</div>
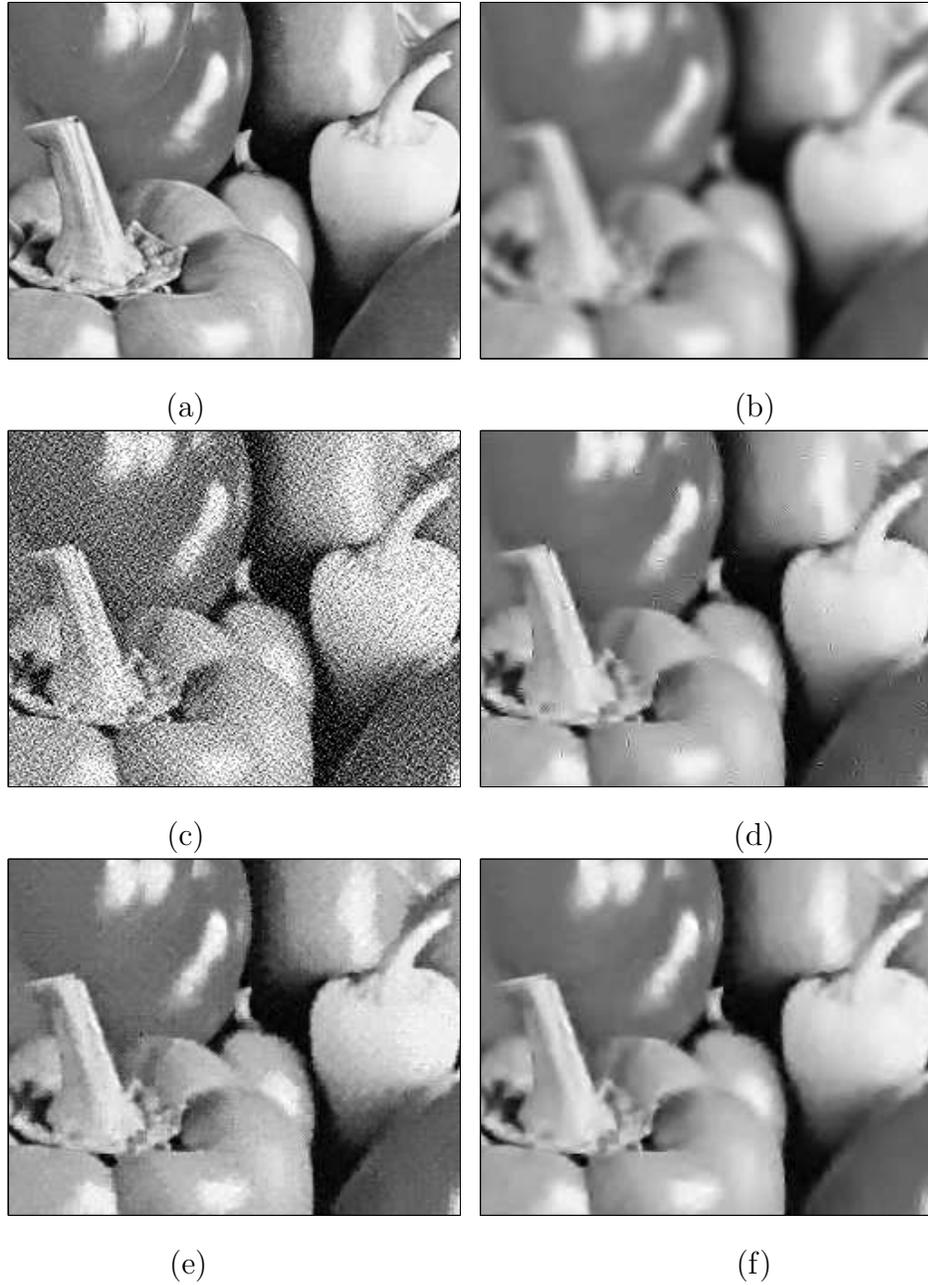
Figure 3.9: Details of the image deconvolution experiment with a *Peppers* image. (a) Original image. (b) Noisy blurred image, BSNR=30 dB. (c) Regularized inversion estimate, ISNR=-11.03 dB. (d) ForWaRD estimate based on result shown in (c), ISNR= 3.91 dB. (e) LPA-ICI estimate based on result shown in (c), ISNR = 4.17 dB. (f) shearlet-based estimate based on result shown in (c), ISNR=5.23 dB.

BSNR=40 dB. The results give an indication that the validation function

$$\frac{\|H\widehat{\mathbf{x}}_\alpha - \mathbf{y}\|_2^2}{(trace(I - H\tilde{H}_\alpha))} \tag{3.16}$$

can be just as effective as $C(\alpha)$ for finding estimates of the optimal α when the standard deviation of the noise is not estimated.

3.5   Chapter Summary

We have studied an effective shearlet-based deconvolution algorithm which utilizes the power of a Fourier representation to approximately invert the convolution operator and a redundant shearlet representation to provide an image estimate. The multi-scale and multi-directional aspects of the shearlet transform provide a better estimation capability over that of the wavelet transform or wavelet-like transforms for images exhibiting piecewise smooth edges. In addition, we have adapted a method of automatically determining the threshold values for the shearlet noise shrinkage without knowing the noise variance by using a generalized cross validation function. Demonstrations show this method to outperform many of the state-of-the-art methods that have been previously compared to the ForWaRD algorithm in the literature.

Chapter 4

Gradient-based Recovery Methods from Partial Fourier
Measurements

Several scanning technologies such as MRI and SAR will greatly benefit from being able to reconstruct an image from a reduced sampling set. The theory of CS points to new ways to reconstruct such images from highly undersampled data. In this chapter, we study a reconstruction method that recovers images assumed to have a sparse representation in a gradient domain by using partial measurement samples that are collected in the Fourier domain. A key improvement of this technique is that it makes use of a robust generalized Poisson solver that greatly aids in achieving a significantly improved performance over similar proposed methods. Experiments also demonstrate that this new technique is more flexible to work with either random or restricted sampling scenarios better than its competitors [5].

4.1 Introduction

Reconstruction of imagery from an incomplete set of samples from a Fourier representation is an important goal to improve scanning technologies. Solutions to such a problem would allow a significant reduction in collection time and improve the capacity to collect very time sensitive events. For example, an MRI scanner records data in a Fourier representation so methods, capable of reconstructing from such partial samples, would greatly reduce a patient's exposure time.

Many MRI and SAR images are sparse or compressible in some transform domain such as those from a wavelet, gradient or Fourier transform. Images such as angiograms are inherently sparse in the pixel domain or gradient domain. For instance, if the image is piecewise constant then a gradient representation would only contain nonzero values near boundary positions.

Maleh *et. al* presented in [117] an improved recovery algorithm that can

recover sparse gradient images from partial Fourier data using orthogonal matching pursuit (OMP) [46]. It is known as the GradientOMP method. The GradientOMP method, however, has some drawbacks. In many MRI data collection scenarios, Fourier samples are collected along curves in Fourier space (k-space) and there is usually no reason to discard any samples collected within such curves. This violates GradientOMP's working premise that the sampling set is the union of two radial slices through the origin perpendicular to one another and randomly chosen points in the plane. While this may be approriate for 3D MRI and 2D spectroscopy when the image plane is perpendicular to the collection trajectories, a versitile method should be able to recover imagery from 2D MRI data as well.

In this chapter, we present a new method that recovers images that can be sparsely represented by gradients using highly undersampled Fourier measurements. This new method is less restrictive on sampling constraints than the GradientOMP and is more robust in the estimation process. A key element in the improvement is that besides using a different/better nonlinear estimator to estimate the gradient field, we solve a generalized Poisson equation to obtain the estimated image. Such a Poisson based solution [118][119] is much more robust to imperfections in the estimate of the gradients and does not require any constraint on the type of data collections that can be made.

4.2 Background

Let $x \in \mathbb{C}^{l \times l}$ and Ω be some randomly chosen set of frequencies of size $|\Omega| = M$ and $W_\Omega : \mathbb{C}^{l \times l} \to \mathbb{C}^M$ to denote the partial Fourier transform operator restricted to Ω. Define the total-variation (TV) operator as

$$TV(x)_{n,m} = \sqrt{|p(n,m)|^2 + |q(n,m)|^2} \tag{4.1}$$

where p and q are the discrete gradients of an image x, i.e.

$$p(n,m) = x(n, m+1) - x(n,m) \tag{4.2}$$

and

$$q(n, m) = x(n + 1, m) - x(n, m) \qquad (4.3)$$

for $0 \leq m, n \leq l - 1$. Given the assumption that x has sparse gradients; that is $\| TV(x) \|_0 = K \ll N = l^2$, and $Y = W_\Omega x \in \mathbb{C}^M$ (representing the partial Fourier measurements of x), the general problem is to estimate x. It was shown in [7] that for an overwhelming percentage of sets Ω with cardinality obeying $|\Omega| = M \geq AK \log(N)$, the minimizer to

$$x_{rec} = \arg \min_{x' \in \mathbb{C}^N} \| TV(x') \|_1 \quad \text{s. t.} \quad W_\Omega x' = W_\Omega x \qquad (4.4)$$

is unique and recovers x, where A is some constant.
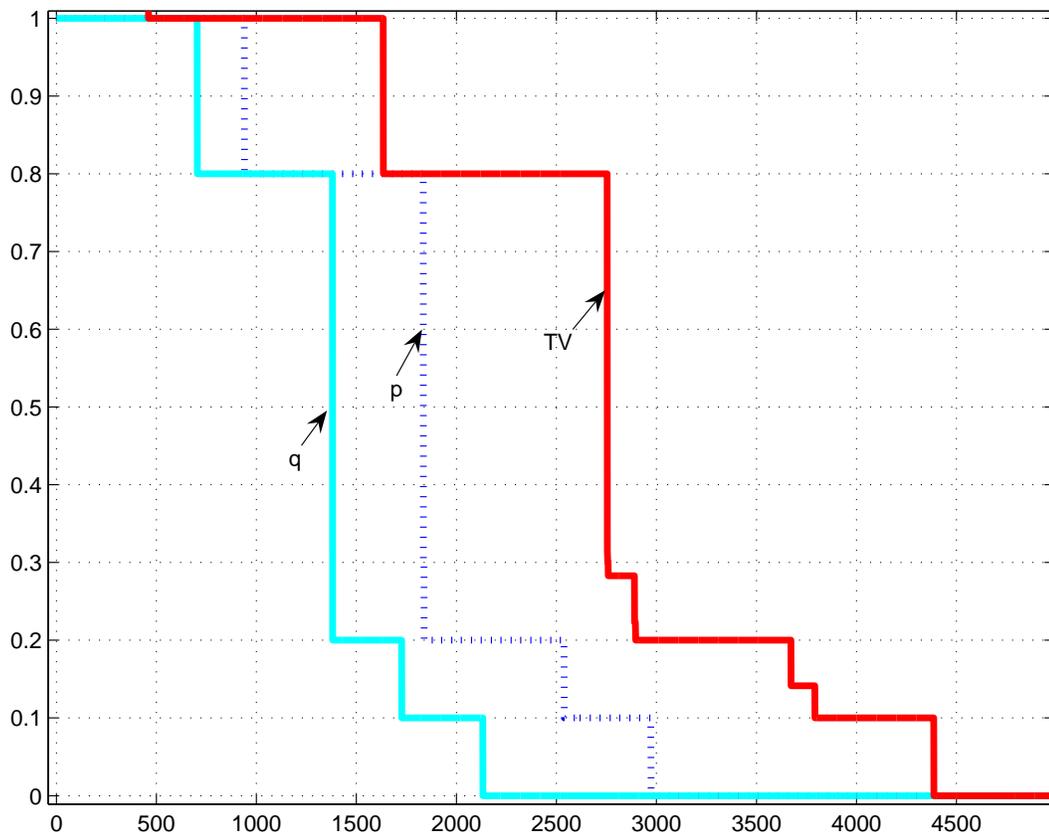
Figure 4.1: The magnitude of TV (red), p (dotted blue) and q (cyan) coefficients in decreasing order for the Shepp-Logan Phantom image (see Figure 4.2(a)).

In this chapter, we show that instead of reconstructing an image by the TV minimization, one can reconstruct the image by separately reconstructing the gradi-

ents and then solving for the image. This allows one to reconstruct the image with a far fewer number of measurements than required by the TV minimization method (see Figure 4.2). Figure 4.1 presents an important comparison in the sparsity of p, q and the TV measure. The plots of the sorted absolute values of the coefficients of the gradients p,q and the TV measure for the Shepp-Logan Phantom image indicate that p and q decay much faster than the TV measure. In fact, it is easy to see from the expression of TV, (4.1), that the coefficients of p and q will always decay faster than the coefficients of TV. This means our method can take advantage of this and be able to reconstruct an image with far fewer measurements than that required by using the TV-based method. Note the assumption that the image of interest has sparse gradients with the use of the TV constraint for estimation has been suggested before [7], [53], [120], [121].

4.3   Image Gradient Estimation

By using the properties of the Fourier transform, we can directly express the partial gradients $P = W_\Omega p$ and $Q = W_\Omega q$, in the Fourier domain from $Y = W_\Omega x$ as

$$P = (e^{\frac{j2\pi k_2}{l}} - 1)Y \ \text{ and } \ Q = (e^{\frac{j2\pi k_1}{l}} - 1)Y, \tag{4.5}$$

where $(k_1, k_2) \in \Omega$. Taking into account the presence of additive noise during the measurement process, gradients can be reconstructed by solving the following two constrained optimization problems

$$p_{rec} = \arg\min_{p'} \| \, p' \, \|_1 \ \text{ s. t. } \ \| \, W_\Omega p' - P \, \|_2 \leq \varepsilon \tag{4.6}$$

$$q_{rec} = \arg\min_{q'} \| \, q' \, \|_1 \ \text{ s. t. } \ \| \, W_\Omega q' - Q \, \|_2 \leq \varepsilon \tag{4.7}$$

where ε is a noise statistic that controls the fidelity of the reconstruction. It can be shown that the solutions to (4.6) and (4.7) will recover the unknown sparse gradients p and q with an error proportional to the noise level [30]. That is, $\|p_{rec} - p\|_2 \leq C\varepsilon$, and $\|q_{rec} - q\|_2 \leq C\varepsilon$, where the constant C may depend on the restricted isometry constant [30].

There has been a number of approaches suggested for solving BPDN problems such as (4.6) and (4.7). In our approach, we shall employ a highly efficient algorithm that is suitable for large scale applications known as the *spectral projected gradient* (SPGL1) algorithm [40].

4.4 Image Reconstruction from Gradients

Once the gradients have been estimated, the problem is to obtain the desired image reconstruction. The gradient field (p, q) of a scalar surface or an image should be integrable (conservative). That is, the integral along any closed curve should be equal to zero. However, this is often not the case when inherent noise during the estimation process contaminates the gradient field.

One of the ways of enforcing integrability is by using the Simchony, Chellappa and Shao (SCS) method [122]. It can be described as follows [118], [119], [122]. Let $(p(n, m), q(n, m))$ be the estimated non-integrable gradient field. Define the curl and divergence operators as

$$C_{p,q} = \texttt{curl}(p, q) = \frac{\partial p}{\partial n} - \frac{\partial q}{\partial m} \tag{4.8}$$

and

$$D_{p,q} = \texttt{div}(p, q) = \frac{\partial p}{\partial m} + \frac{\partial q}{\partial n}. \tag{4.9}$$

The objective is to obtain a surface $\hat{x}(m, n)$ estimate (reconstructed image) such that its gradient field $(\hat{p}, \hat{q})$ satisfies the integrability constraint $C_{\hat{p},\hat{q}} = 0$ and minimizes

$$J(\hat{x}) = \int \int [(\hat{p} - p)^2 + (\hat{q} - q)^2] \, dm \, dn. \tag{4.10}$$

The error function $E = (\hat{p} - p)^2 + (\hat{q} - q)^2$ is a function of $\hat{x}$, and the extremum points are given by the Euler-Lagrange equation:

$$\frac{\partial E}{\partial \hat{x}} - \frac{d}{dn} \frac{\partial E}{\partial \hat{p}} - \frac{d}{dm} \frac{\partial E}{\partial \hat{q}} = 0. \tag{4.11}$$

Solving the above equation gives the Poisson equation

$$0 - 2\frac{\partial}{\partial n}(\hat{p} - p) - 2\frac{\partial}{\partial m}(\hat{q} - q) = 0.$$

$$\nabla^2 \hat{x} = \mathtt{div}(p, q), \tag{4.12}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial n^2} + \frac{\partial^2}{\partial m^2} \tag{4.13}$$

is the Laplacian operator.

Since the least squares based approaches, such as the one described above, are known not to be robust to noise and outliers, they result in a loss of salient features in the image. Thus, we incorporate in our proposed algorithm an improved Poisson based approach known as the *affine transformation on gradients using diffusion tensors* [118].

4.4.1 Affine Transformation of Gradients using Diffusion Tensors

Motivated by the application of anisotropic diffusion in image enhancement and denoising, a generalization of divergence based equation for image reconstruction has been suggested [118]. Given an image I, the heat equation estimates the image by solving the following PDE

$$I_t = \mathtt{div}(\nabla I), \tag{4.14}$$

where ∇I is the gradient of I and I_t is the gradient at time t. For anisotropic restoration, Perona and Malik [123] proposed to solve the following PDE by using a function f

$$I_t = \mathtt{div}(f(\|\nabla I\|)\nabla I), \tag{4.15}$$

to stop diffusion across edges. This way sharp edges can be preserved in the final estimate. Weickert proposed a generalized divergence based equation for image restoration

$$I_t = \mathtt{div}(D\nabla I), \tag{4.16}$$

where D is a symmetric positive definite matrix at each pixel [124]. Similarly, for image reconstruction, one can define a modified Poisson equation using D.

Given that

$$D(n, m) = \begin{bmatrix} d_{11}(n, m) & d_{12}(n, m) \\ d_{21}(n, m) & d_{22}(n, m) \end{bmatrix} \tag{4.17}$$

be a 2×2 symmetric positive-definite matrix associated to each pixel location, a generalized relation is then

$$\mathtt{div}\left(D\begin{bmatrix}\hat{p}\\\hat{q}\end{bmatrix}\right)) = \mathtt{div}\left(D\begin{bmatrix}p\\q\end{bmatrix}\right). \tag{4.18}$$

This equation is the associated Euler-Lagrange equation of the functional

$$J(\hat{x}) = \int\int d_{11}(\hat{p}-p)^2 + (d_{12}+d_{21})(\hat{p}-p)(\hat{q}-q) + d_{22}(\hat{q}-q)^2 \, dm \, dn. \tag{4.19}$$

Let x_D denote the lexicographically ordered column vector of $\mathtt{div}(d_{11}p+d_{12}q, d_{21}p+d_{22}q)$, then the general Poisson equation (an alternative formulation of the general relation) can be simply expressed as

$$\nabla_D^2 \hat{x} = x_D, \tag{4.20}$$

where ∇_D^2 is the weighted Laplacian kernel based on the diffusion tensor D (see [118] for more details). The image can then efficiently be recovered as $\hat{x} = L_D^{-1}u_D$, where the matrix L_D is the sparse weighted Laplacian matrix [118]. Note that this method does not require any restriction on the amount of information gleaned from a partially sampled Fourier data set.

We now summarize our proposed algorithm as follows:

Given partial Fourier data $Y = W_\Omega x$,

1. Find partial Fourier gradients P and Q using (4.5).

2. To recover p and q, solve (4.6) and (4.7), respectively.

3. Reconstruct the desired image using the method of *affine transformation of gradients using diffusion tensors.*

4.5   Experiments

In this section, we give some numerical examples to illustrate the performance of our algorithm and compare it with the GradientOMP [117], the *TV* minimization method [125], and the minimum energy reconstruction by zero-filling [53].

In the first example, we reconstructed a 512×512 Shepp-Logan phantom image after 94.6% random undersampling. As can be seen from Figure 4.2, the reconstruction by zero-filling, GradientOMP and the TV minimization fail to reconstruct the image perfectly, while our proposed algorithm reconstructs the image exactly (MSE $= 3.2e^{-6}$).
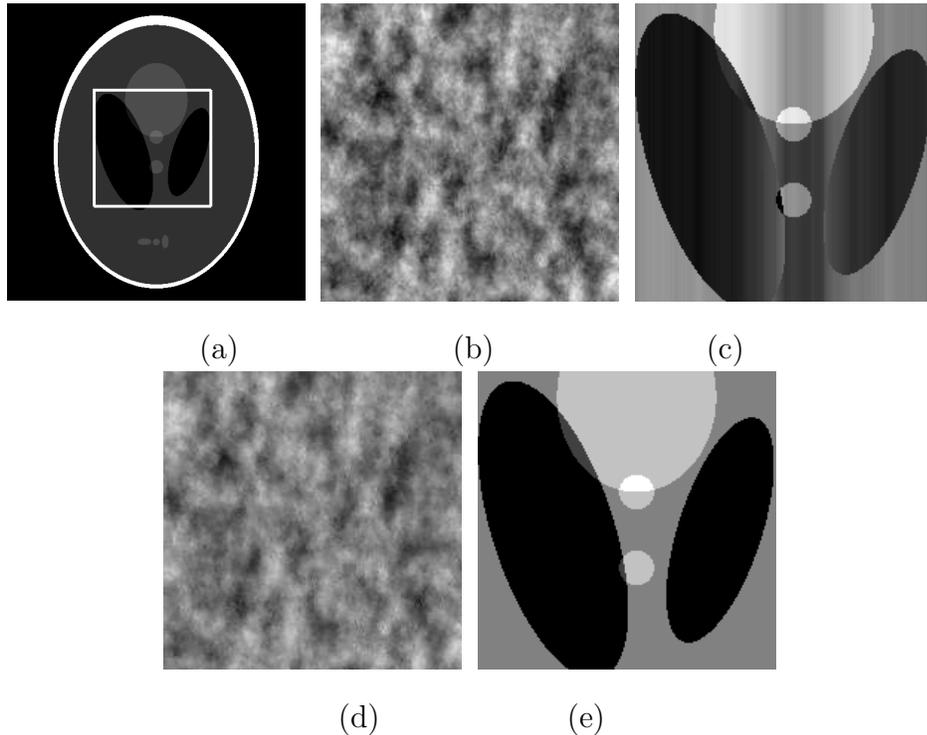
Figure 4.2: (a) 512×512 Shepp-Logan Phantom example. Reconstructed by (b) zero-filling, (c) GradientOMP, (d) TV method, and (e) our proposed method.

In the second example, Figure 4.3, we show the reconstruction of a 256×256 MRI of the brain after 68.5% undersampling using Cartesian radial sampling. Both GradientOMP and zero-filling provide poor image reconstructions, while our method performs significantly better.

Finally, to compare our algorithm in terms of robustness to noise suppression, we tested a 128×128 Shepp-Logan phantom image with additive white Gaussian noise with signal-to-noise ratio of 20 dB. Only 35% of the Fourier coefficients were used. As can be seen from the Figure 4.4, reconstruction by the GradientOMP

suffers from the vertical streaking artifacts while our method removes these artifacts and provides more improved reconstruction.

(a)   (b)   (c)

(d)   (e)
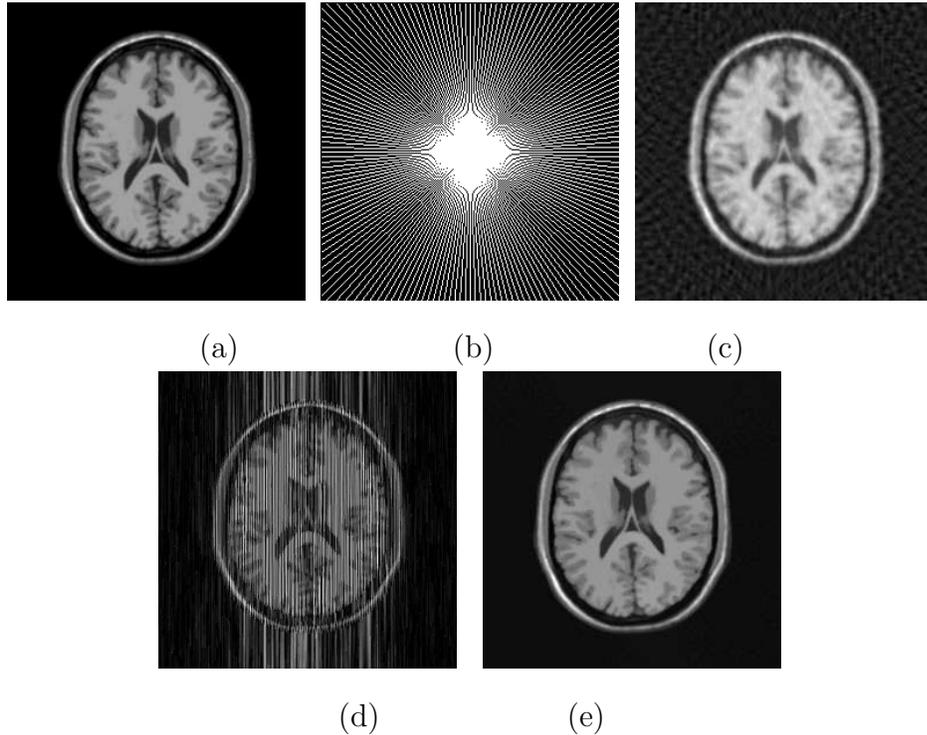
Figure 4.3: (a) Original MRI image. (b) Radial k-space trajectories. Reconstructed by (c) zero-filling, (d) GradientOMP, and (e) our method.

4.5.1 Remark

Experiments have shown that image reconstruction by the GradientOMP always suffers from the vertical streaking artifacts regardless of the type of sampling being used. Reconstruction by the TV method is effective when the measurements are given along the radial lines, however it fails miserably when the coefficients are chosen randomly [117]. Our method is robust to work with either random or restricted sampling and provides much better reconstruction from fewer measurements than the TV method or the GradientOMP.

Recently, there has been a great interest in using ℓ_p minimization with $p < 1$ for compressive sensing [120], [50]. It has been observed that the minimization of
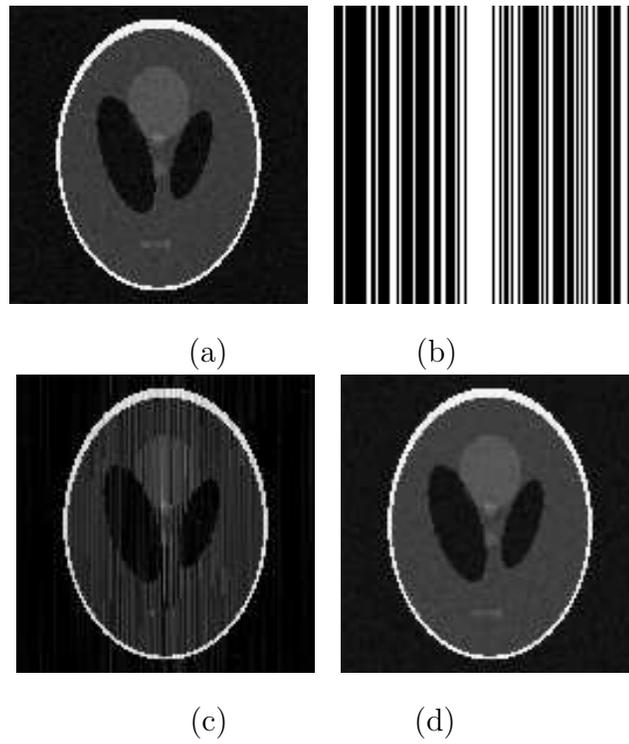
Figure 4.4: (a) Noisy 128×128 phantom image (SNR 20dB). (b) Sampling pattern. Reconstructed image by (c) GradientOMP and (c) by our method.

such a nonconvex problem leads to recovery of signals that are much less sparse than required by traditional methods [120]. Our method is flexible enough that, instead of solving problems (4.6) and (4.7) using SPGL1, the gradients can be reconstructed by using any of these nonconvex algorithms. It is also possible to reconstruct the gradients from even fewer measurements by adapting the reweighted ℓ_1 minimization algorithm [52]. OMP can also be used to recover the gradients. In this case, our method can be viewed as an improved GradientOMP.

4.6 Chapter summary

We studied a new method of recovering an image when partially sampled data is collected in the Fourier domain and the image is considered to be sparsely represented by gradients. This entails using a robust general Possion solver to obtain the reconstructed image. Various experiments have shown a great improvement in quality of reconstruction as well as a robustness to various sampling scenarios over similar type schemes.

Chapter 5

Compressive SAR Imaging

Since a SAR image is a map of the spatial distribution of the reflectivity function of stationary targets and terrain, many SAR images can be sparse or compressible in some representation such as those from wavelets or the complex wavelet transform[1]. The recently introduced CS theory states that it is possible to recover such compressible images from a small number of random measurements provided that the undersampling results in noise like artifacts in the transform domain and an appropriate nonlinear recovery scheme is used [6], [7].

In this chapter, we introduce a new SAR image formation algorithm empirically derived based on the theory of CS that reduces the number of transmitted and/or received waveforms. It will be demonstrated that if the SAR image is assumed to be compressible in some transform domain, then one can reconstruct a good estimate of the reflectivity profile using this new image formation algorithm that relies on using a far fewer number of waveforms than the conventional systems do and requires no changes to a radar system hardware to work. It is also the case, that the radar community has considered similar concepts that we are presenting such as that provided in [126]. Yet our method enhances some of these suggestions and provides a framework along with general reconstruction techniques. By using concepts provided by CS theory we are able to propose an imaging system that should pave the way for many new applications that are highly desirable. In addition, not only do we demonstrate this concept of SAR imaging using CS theory with real data but we point out some of its valued benefits never before realized.

---

[1]SAR images are often characterized by the multiplicative noise known as speckle. Speckle makes the compressibility of the SAR images difficult. However, the underlying reflectivity map without the speckle has compressibility as good as many natural images.

5.0.1 Previous Compressive Radar Related work

Inspired by the CS theory, more efficient schemes for sensing signals at much lower sampling rate than required by the traditional Nyquist sampling theorem have been proposed. This sub-Nyquist acquisition framework is often known as the analog-to-information (A2I) conversion [127, 128, 129]. Motivated by the CS theory one such compressive radar imaging scheme based on A2I was proposed in [130]. It was argued that such a radar system can eliminate the need for the matched filter in the radar receiver and reduce the required receiver analog-to-digital conversion bandwidth [130]. A similar acquisition and imaging system for Ground Penetrating Radar (GPR) was proposed in [131],[132]. Instead of sampling the radar returns at the Nyquist rate, linear projections of the echo signals with random vectors were taken as measurements. It was shown that, using only a small subset of the measurements, sparser and sharper target images could be obtained compared to the standard backprojection method [131],[132]. In [133], a high-resolution radar was proposed based on CS by transmitting specially designed waveforms. A similar concept was also proposed in [134], where the theory of CS with random convolution was used by transmitting random noise like signals. Also, in [135], to reduce the amount of stored SAR data, a method based on CS theory was proposed. Recently, a method of imaging a scene of sparse targets from the perspective of scattering theory and CS has been proposed in [136].

Our approach in this work, is in contrast to some of the above mentioned compressive radar related algorithms that have only considered using CS as part of the A2I conversion or transmitting specially designed waveforms. Note our method also requires no changes to a system's hardware to work unlike many other schemes that propose using CS theory for imaging.

5.0.2 Compressive Sampling for SAR

The design of a CS undersampling scheme for SAR entails the selection of phase histories such that the resulting mutual incoherence is small. Some of the

results about CS are based on the fact that the k-space samples are obtained randomly. However, sampling a truly random subset of the phase histories in SAR is usually impractical for existing hardware. In this section, we consider two compressed sensing k-space undersampling schemes for SAR. Since, the PRF essentially determines the slow-time sampling rate, our CS undersampling schemes are based on modifying the PRF of the radar. Implementation of such schemes is very simple and requires a minor change to the PRF scheduling of the radar. To analyze the severity of artifacts introduced by these undersampling schemes, we adapt the notion of the Point Spread Function (PSF) with regards to CS theory from [53].

Note that in practice the sparsity of the image will typically mean the percentage of transform coefficients needed to form an acceptable reconstruction. The acceptable reconstruction will depend on specific applications in mind such as whether it will be used for target identification or situational awareness. Thus, before methods suggested here are fielded, systematic studies will be needed that will depend on their intended use.

5.0.2.1   Point Spread Function (PSF)

To analyze the severity of artifacts introduced by these undersampling schemes, we adapt the notion of the Point Spread Function (PSF) with regards to CS theory from [53]. Let $\mathbf{r}$ denote the collection of phase histories and $\tilde{S} = \mathbf{R_S}\mathbf{r} = \mathbf{R_S}\mathbf{A}\mathbf{g} = \Phi\mathbf{g}$ represent the k-space measurements obtained by incorporating one of the slow-time undersampling schemes described above. Here, $\mathbf{R_S}$ represents the restriction operator that selects the phase histories from the SAR model. Furthermore, we assume that $\mathbf{g}$ has a sparse representation (or is compressible) in some basis Ψ, so that $\mathbf{g} = \Psi v$. In this case, the sensing matrix Θ is given by $\Theta = \Phi\Psi$.

In [53], Lustig *et al.* proposed that the Transform Point Spread Function (TPSF) be used to measure the incoherence of a sampling scheme. It is defined as follows

$$TPSF(m;n) = \frac{\langle \Phi\Psi\mathbf{e}_m, \Phi\Psi\mathbf{e}_n \rangle}{\|\Phi\Psi\mathbf{e}_m\|_2 \|\Phi\Psi\mathbf{e}_n\|_2}, \tag{5.1}$$

where $\mathbf{e}_m$ is the mth vector of the natural basis having 1 at the mth location and zeros elsewhere. Ideally, for $m \neq n$, the $TPSF(m;n)$ should be much smaller than 1 and should be noise like. This implies that the aliasing artifacts introduced by undersampling produces relatively small uncertainty in resolving m^{th} transform coefficient from the nth coefficient. In the case, when Ψ is the identity matrix, we call the resulting TPSF simply the PSF. The maximum of the sidelobe-to-peak ratio (SPR) or the maximum off diagonal entry in TPSF, $\max_{m \neq n} |TPSF(m;n)|$ is used as a measure of severity of the aliasing artifacts [53]. Note that the SPR is also equal to the mutual coherence. In what follows, we propose using two types of slow-time undersampling and analyze their PSFs to establish their viability.

5.0.2.2   Random slow-time undersampling

As discussed earlier, as the sensor advances along its path, pulses are transmitted and received by the radar. The pulses are transmitted at every $PRI = \frac{1}{PRF}$. Undersampling methods based on sampling at regular intervals produce strong aliases. Random changes to the PRI can break up the periodicity of the aliasing artifacts and can convert strong aliases to random noise like artifacts [137, 138]. For this reason, instead of transmitting pulses with a regular PRI, we propose to transmit fewer pulses than traditional systems at random intervals. This amounts to undersampling the 2D phase histories along the slow-time axis randomly.

5.0.2.3   Jittered slow-time undersampling

Jittered undersampling is based on a regular undersampling which is perturbed slightly by random noise. The effect of jitter in one dimension was analyzed by Balakrishnan in [139]. He analyzed time jitter in which the nth sample is jittered by an amount ζ_n so that it occurs at time $nP_T + \zeta_n$, where P_T is the sampling period. He reported that if the ζ_n are uncorrelated then the following happens: high frequencies are attenuated, the energy lost to the attenuation appears as uniform noise, and the basic structure of the spectrum does not change [137, 138, 139]. Jittered sampling in

2D was generalized and applied in computer graphics in [137, 138]. Also, Hennenfent and Herrmann in [140] have successfully applied 2D jittered undersampling in the context of CS for seismic data processing. Inspired by the properties of the jittered sampling, we propose to apply jittered undersampling in slow-time as well.

5.0.2.4 PSF Analysis

In this section, we analyze the aliasing artifacts introduced from the slow-time undersampling schemes by the PSF. In the case, a sparsifying transform is used, TPSF can be studied to analyze the aliasing artifacts. We use the PSF of pure 2D random sampling where samples are chosen randomly from a Cartesian grid, as a standard for comparison [53]. Figure 5.1 shows the PSFs for random 2D undersampling, random slow-time undersampling and jittered slow-time undersampling. The height of the red line measures the effect of the aliasing artifacts. The higher the line more severe are the aliasing artifacts. It is clear from Figure 5.1, that by undersampling the phase histories along the slow-time axis, we are mainly exploiting 1D sparsity in 2D. Therefore, the artifacts introduced by random or jittered slow-time undersampling are not as good as the one obtains with truly 2D random k-space undersampling. Nevertheless, these aliasing artifacts can be removed using a nonlinear reconstruction technique promoting sparsity as suggested in [140, 53].

5.0.3 SAR image reconstruction

Given the partial k-space measurements of the reflectivity map to be imaged, in this section, we show how the nonlinear recovery can be used to reconstruct the SAR image.

From the previous discussion, in the presence of additive measurement noise, we can write the partial k-space measurements as

$$\tilde{S} = \mathbf{R_S}\mathbf{A}g + \eta = \mathbf{R_S}\mathbf{A}\Psi v + \eta = \Phi\Psi v + \eta = \Theta v + \eta, \tag{5.2}$$

where η is an arbitrary noise vector with $\|\eta\|_2 \leq \varepsilon$, and $\Phi = \mathbf{R_S}\mathbf{A}$. Note the model in (5.2) may not be completely accurate as speckle is multiplicative but this is a
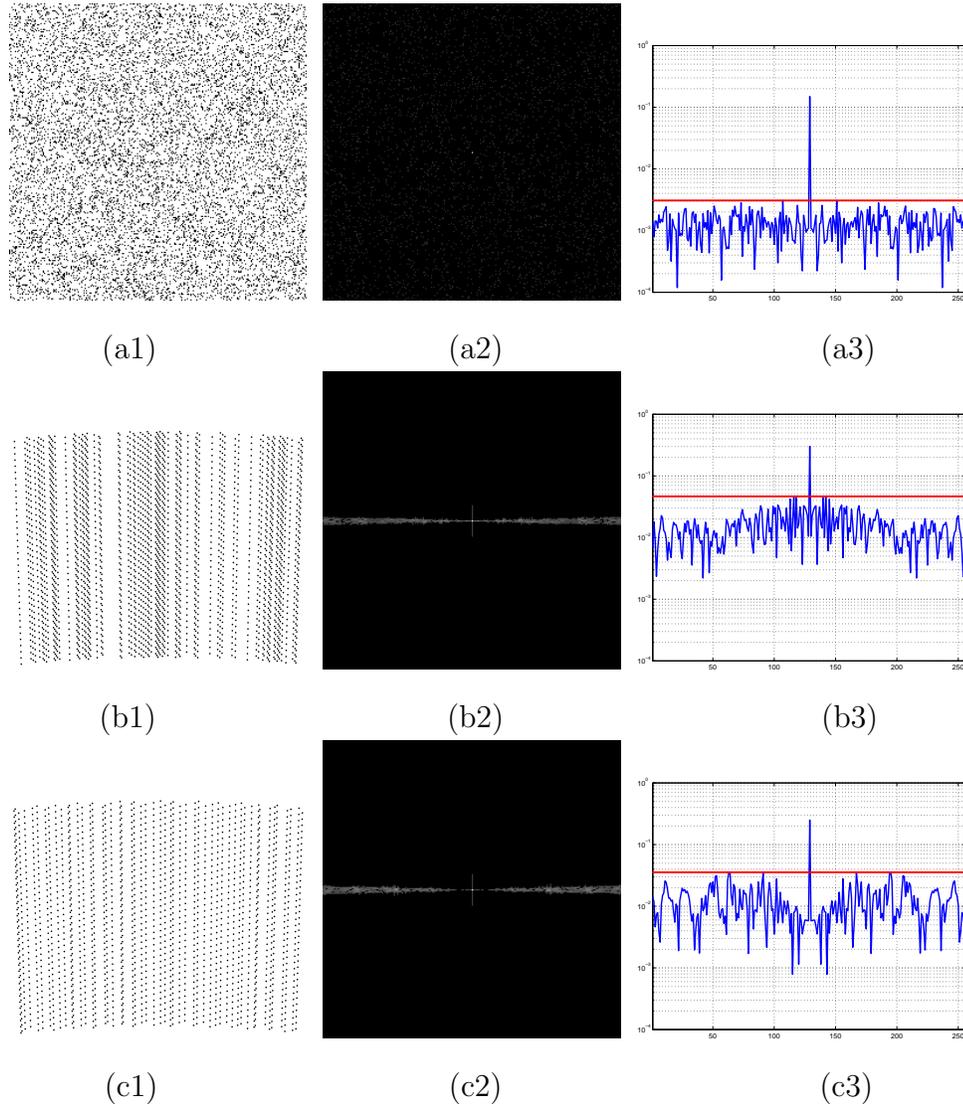
Figure 5.1: PSFs of various SAR undersampling schemes. (a1), (b1), (c1) Random 2D points, Random slow-time undersampling, and Jittered slow-time undersampling, respectively. (a2), (b2), (c2) Image plots of the 2D PSFs corresponding to (a1), (b1), and (c1), respectively for all m and a fixed n. (a3), (b3), (c3) 1D slice through the 2D PSFs in (a2), (b2), and (c2), respectively.

commonly used model for SAR (see [64] and [57]). The reflectivity map $\mathbf{g}$ can be estimated via v by solving the following ℓ_1 minimization problem

$$v_{rec} = \arg\min_{v'} \| v' \|_1 \quad \text{s. t.} \quad \| \tilde{S} - \Phi\Psi v' \|_2 \leq \varepsilon. \tag{5.3}$$

It was shown in [141] that if $\|v\|_0 < \frac{1}{4}\left(\frac{1}{\mu}+1\right)$ then the solution to (5.3) obeys

$$\|v_{rec} - v\|_2^2 \leq \frac{4\varepsilon^2}{1 - \mu(4\|v\|_0 - 1)}. \tag{5.4}$$

It is very difficult to prove any general claim that Θ satisfies a RIP or a mutual incoherence property for any particular sampling scheme. This remains an open problem. Thus, the best method for establishing mutual incoherence is to study the TPSF for a particular proposed sampling scheme given a particular scanning scenario. This presents no real obstacle in applications especially since many pre-calculations can be and are done before scanning begins. Note that, in the context of CS, the TPSF has successfully been used to characterize the incoherence of different sampling schemes arising in magnetic resonance [53] and in photo-acoustic tomography [142].

5.0.3.1 Speckle

Many coherent imaging modalities such as SAR, sonar, holography and ultrasound often suffer form a multiplicative noise known as speckle. Speckle appears when objects illuminated by coherent radiation have surfaces that are rough compared with the imaging wavelength. It is caused by the constructive and destructive interference of the coherent returns scattered by small reflectors within each resolution cell [143],[55]. Speckle noise can make the detection and interpretation difficult for automated as well as human observers. In some cases, it maybe important to remove speckle to improve applications such as segmentation. To deal with speckle, instead of minimizing (5.3), we propose to minimize a modified problem

$$v_{rec} = \arg\min_{v'} \|v' \|_1 + \alpha TV(|\Psi v'|) \quad \text{s. t.}$$

$$\parallel \tilde{S} - \Phi\Psi v' \parallel_2 \leq \varepsilon \qquad (5.5)$$

for some user specified $\alpha \geq 0$ and TV is the Total Variation [144] defined as

$$TV(x) = \parallel \nabla(x) \parallel_1 .$$

By adding the TV constraint along with the ℓ_1 constraint, we require the magnitude of the underlying complex SAR reflectivity field to be sparse in both the transform Ψ and finite difference domains. The assumption that the underlying reflectivity field is piecewise smooth has been used before for image restoration under speckle noise [64], [145], [146] [144]. Note that no stability results have been proven for the minimization by the TV method. However, empirical experiments in [53], [147] have shown that TV minimization provides the reconstruction as good as BP. Our experiments have shown that in practice (5.5) provides much better reconstruction in the presence of high speckle.

5.0.4 Applications

The idea of transmitting waveforms at a non-uniform PRI for SAR has been suggested before [126]. This method, however, suffers from the smearing of the image in the cross-range dimension due to the randomness of the PRI. By solving the basis pursuit denoising problem (5.3), or (5.5), we are able to not only reconstruct the image as good as some of the traditional SAR reconstruction methods do from the full data but even at a lower sampling rate. Thus, our method of compressing the synthetic aperture offers many advantages.

5.0.4.1 Counter-countermeasure

In many military applications of SAR, the user encounters scenarios where the adversary uses a transmitting radar to send out a signal within the band of the SAR system transmitter to confuse the SAR receiver. This process is called the Electronic Countermeasure (ECM)[148],[63]. The ECM causes the SAR system to receive and process erroneous information, which obscures targets or features of interest.

The ECM schemes used to jam a SAR system often relies on estimating the radar signal parameters such as PRI by exploiting the multiple transmissions of the signal at each synthetic aperture position. An elementary electronic counter-countermeasure (ECCM) is to jitter the PRF [148]. Each outgoing pulse is either delayed or not depending on a sequence generated by a random number generator. A simpler implementation of changing the interpulse period is to drop pulses randomly. Hence, our compressive aperture method can offer strong countermeasures resistance [126, 148].

5.0.4.2   Reduction in data

In many SAR systems, radar data is directly stored on board and then transmitted to the ground in some reduced form. Our system has the potential to significantly reduce the amount of data to be stored and transmitted [149, 135].

Based on the CS theory, Bhattacharya *et al.* in [135], proposed a method of compressing the raw SAR data by using a simple encoder with a 2D FFT and a random sampler. The decoding was done by using one of the CS recovery algorithms. Similarly, one can also encode the raw SAR data by using our compressed aperture method.

5.0.4.3   Obtaining wider swaths

In [150], Stoyle proposed a satellite imaging method that obtains an image of a wide swath of a planet's surface. He argued that by transmitting pulses at random PRIs, it is possible to image a much wider swath than possible by the conventional methods (see [150] for more details). Our method can also achieve the same task with an additional advantage of reduced data transmission.

5.0.5   Experimental Results

In the following sections, we present results of our proposed CS SAR methods on synthetic and real SAR data.

5.0.5.1 Phase transition diagrams

We evaluate the performance of our compressed synthetic aperture radar methods by generating phase transition diagrams [151]. A phase transition diagram provides a way of checking ℓ_0/ℓ_1 equivalence, indicating how sparsity and indeterminancy affect the success of ℓ_1 minimization [151, 152]. Let $\delta = \frac{M}{N}$ be a normalized measure of undersampling factor and $\rho = \frac{K}{M}$ be a normalized measure of sparsity. A plot of the pairing of the variables δ and ρ describes a two-dimensional phase space $(\delta, \rho) \in [0, 1]^2$. In the following experiments, the values of δ and ρ ranged through 30 equispaced points in the interval $[0.03, 1]$ and $N = 900$. At each point on the grid, corresponding to a CS SAR model for certain values of K, M and N (in this case $N = 900$), we recorded the mean number of coordinates at which original and reconstruction differed by more than 10^{-3}, averaged over 30 independent realizations. In our approach, we employed a highly efficient algorithm that is suitable for large scale applications known as the *spectral projected gradient* (SPGL1) algorithm [40] for solving BP and BPDN problems such as (5.3).

In Figures 5.2(a) and (b), we show the phase transition diagrams corresponding to the slow-time random and jittered undersampling, respectively, for the case when there is no complex white Gaussian measurement noise ($\eta = 0$) and $\Psi = I$ in (5.2). In Figures 5.2(c) and (d), we show the phase transition diagrams for the slow-time random and jittered undersampling, respectively, for the case when the complex white Gaussian measurement noise η has been added with signal-to-noise-ratio of 20 dB in (5.2). These plots indicate that the original target scene (image) can be recovered well as long as it is sparse enough and enough measurements are taken.

5.0.5.2 Reconstruction from the compressive measurements

In this section, we demonstrate the performance and applicability of our compressive imaging algorithm on synthetic and real SAR data and compare it with the PFA. Note that the approximation underlying the PFA is generally poorer compared with the other algorithms. However, it is one of the most commonly used
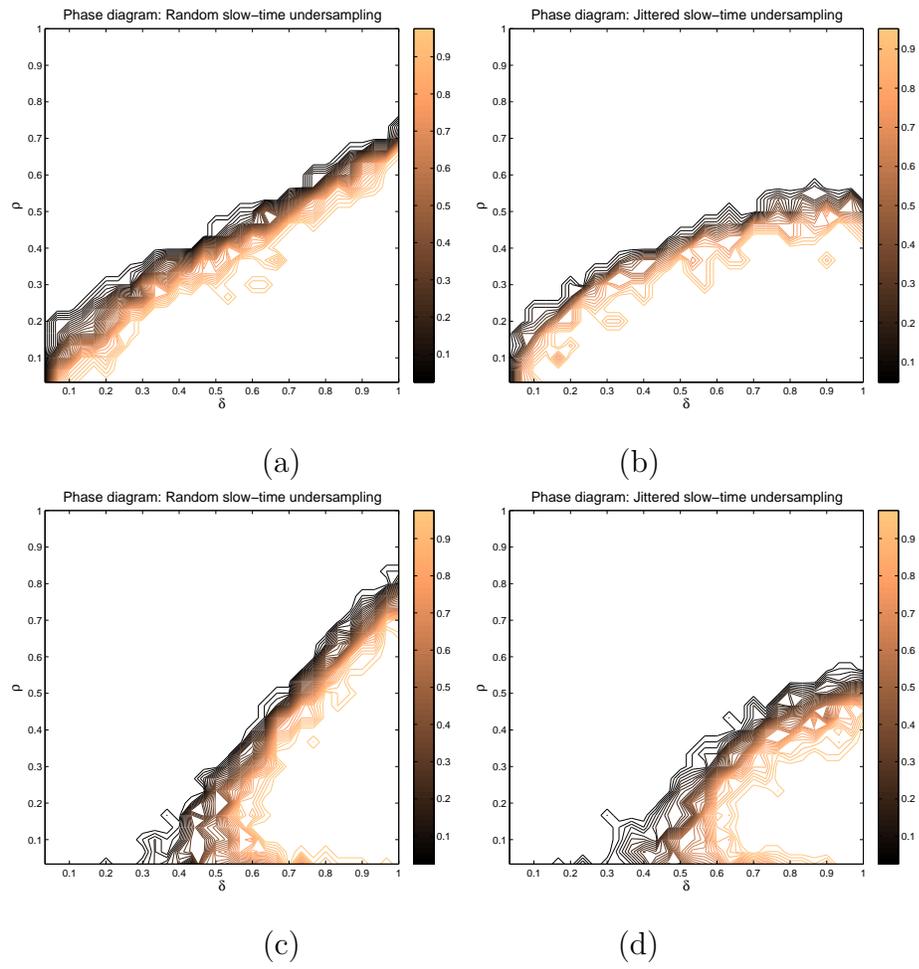
Figure 5.2: Phase transition diagrams corresponding to Random slow-time undersampling (a) without noise, (c) with noise and Jittered slow-time undersampling (b) without noise and (d) with noise.

reconstruction methods for SAR which is why it is chosen for comparison.
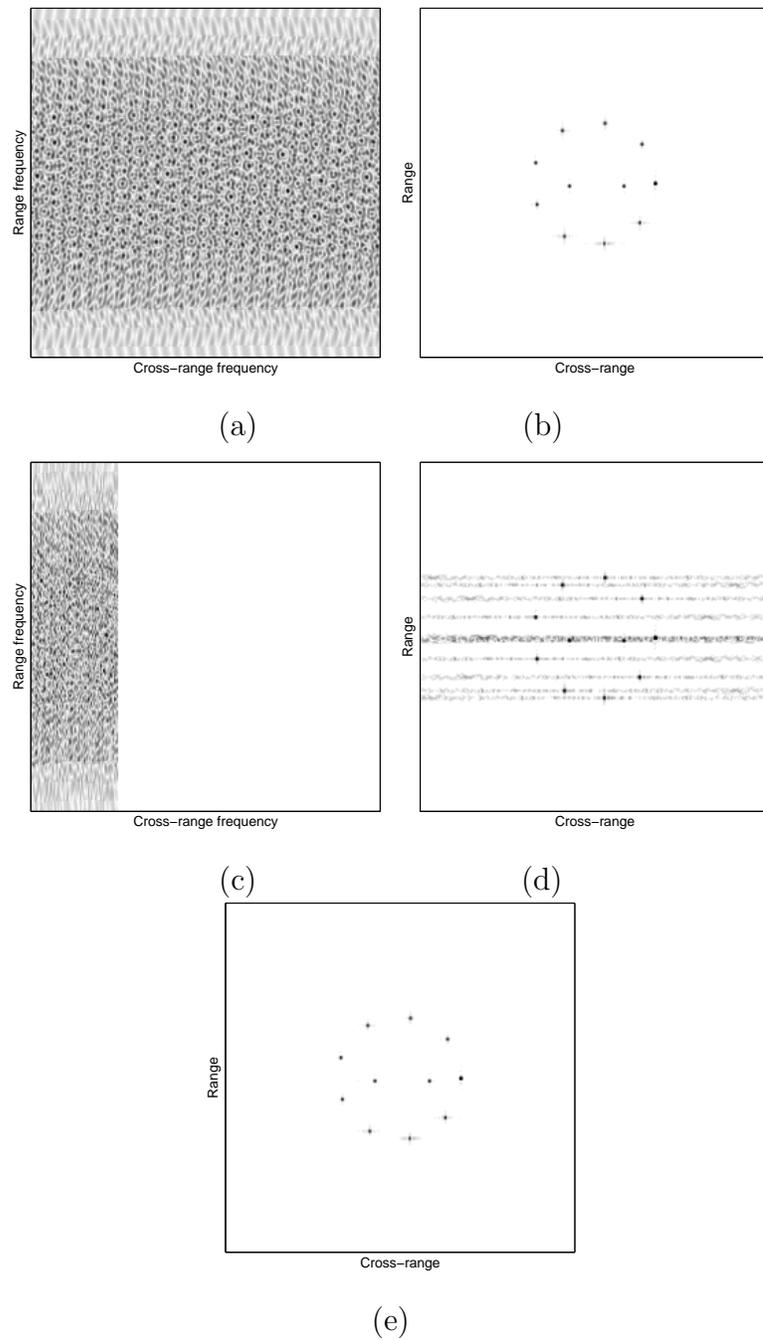
Figure 5.3: Point targets example. (a) Simulated full data. (b) Traditional PFA based reconstruction from the full simulated data. (c) Random slow-time undersampled phase histories (25% of data used). (d) Reconstructed by the PFA from the compressive measurements. (e) Reconstructed image using our method.

In the first example, we used eleven point targets with unit magnitude to generate the phase histories. The control parameters used in our simulation are summarized in Table 1. These determine all other variable parameters (see [58] for details). The range resolution is given by $\rho_r = \frac{c}{2B} = 1.1m$ and the cross-range resolution is given by $\rho_{cr} = \frac{\lambda}{2\Delta\theta} = 1.1m$. In Figure 5.3(a), we show the generated phase histories from the eleven point targets. Figure 5.3(b) shows the reconstruction of the point targets from the full simulated data using the PFA [58]. In Figure 5.3(c), we show the phase histories obtained after applying 75% random slow-time undersampling (25% of data used). Figure 5.3(d), shows the traditional reconstruction from the compressive measurements using the PFA. Figure 5.3(e) shows the reconstructed image using our proposed method. It is clear from Figure 5.3 that our method produces image from the compressive measurements as good as the PFA does from the full simulated data.

Table 5.1: Parameters used in the first example

Center frequency	3.80×10^9 Hz
Pulse width	1×10^{-6} sec
Bandwidth	1.34×10^8 Hz
Chirp rate	1.34×10^{14} Hz/sec
$\Delta\theta$	$2.1°$

In the second experiment, we used the ISAR data collected on a SAAB 9000 car using System Planning Corporation's Mark V radar[2]. We reconstructed the image after 60% jittered slow-time undersampling was applied to the data (40% of data used). As can be seen from Figure 5.4, the reconstructed image from the compressed measurements, shown in Figure 5.4(e), is identical to the one reconstructed from the full measurements, shown in Figure 5.4(a). Figure 5.4(d) shows how the traditional reconstruction fails to recover the ISAR image from the compressive measurements shown in Figure 5.4(c).

For the following two experiments, we used the SAR images from the MSTAR

---

[2]specifications of the radar can be found at www.sysplan.com/Radar
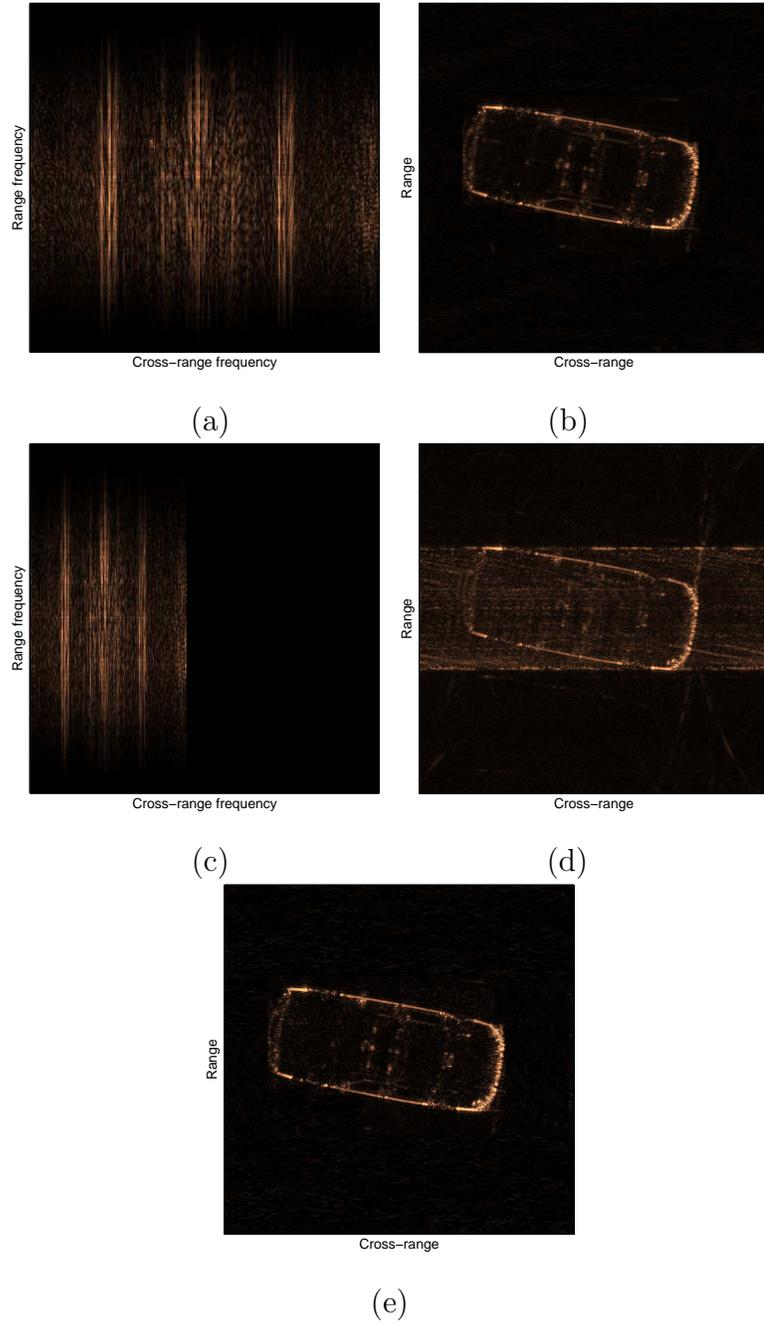
Figure 5.4: SAAB 9000 car ISAR example. (a) Full measured data. (b) Traditional reconstruction from the full data. (c) Jittered slow-time undersampled phase histories (40% of data used). (d) Traditional reconstruction from the compressive measurements in (c). (e) Reconstructed image using our proposed method.

public target database [153]. We simulated SAR phase histories using the acquisition method described in [56]. We used Daubechies 4 wavelet as a sparsifying transform for this experiment. The reconstruction from only 50% of the jittered slow-time undersampled data using the PFA, (5.3) and (5.5) is shown in Figure 5.6(d), (b) and (c), respectively. In Figure 5.6(e) and (f) we show how much speckle has been removed by solving (5.3) and (5.5), respectively. This experiment shows that it is possible to reconstruct and despeckle simultaneously from the compressive measurements. The value of α was chosen to be 3.8 after experimenting with different values for α and checking its performance.

Reconstructed SAR images with different methods in the case when only 50% of the random slow-time undersampled phase histories used are shown in Fig. 5.5. Fig. 5.5(a) shows the conventional SAR image where speckle is clearly visible. Fig. 5.5(b), (c) and (e) are the reconstructed images from the compressive measurements using the PFA, (5.3) and (5.5), respectively. In Fig. 5.5(d) and (f) we show how much speckle has been removed (i.e the residual) by solving (5.3) and (5.5), respectively.

In the final experiment, we used the ISAR dataset of a B-747 [62]. Figure 5.7(a)-(c) show the reconstructed image from the traditional method with full data, traditional method with partial data and our method from the partial data, respectively. The reconstruction was done using 70% random slow-time undersampled phase histories (30% of data used). As can be seen from Figure 5.7 that our compressive imaging method provides a good reconstruction compared to the traditional method based on the full measurements.

Note that in our experiments the additive noise was either non-existent or almost negligible. In the case when noise was non-existent, ε was chosen to be 0. For the other cases, ε was set to 10^{-2}, which was a value we determined by trial and error.

It has been observed by many researchers [30], [53], [26] that in practice, Fourier samples in the order of three to five times the number of sparse coefficients suffice for a good reconstruction. Our experiments also support this claim. In our

(a)

(b)

(c)

(d)

(e)

(f)
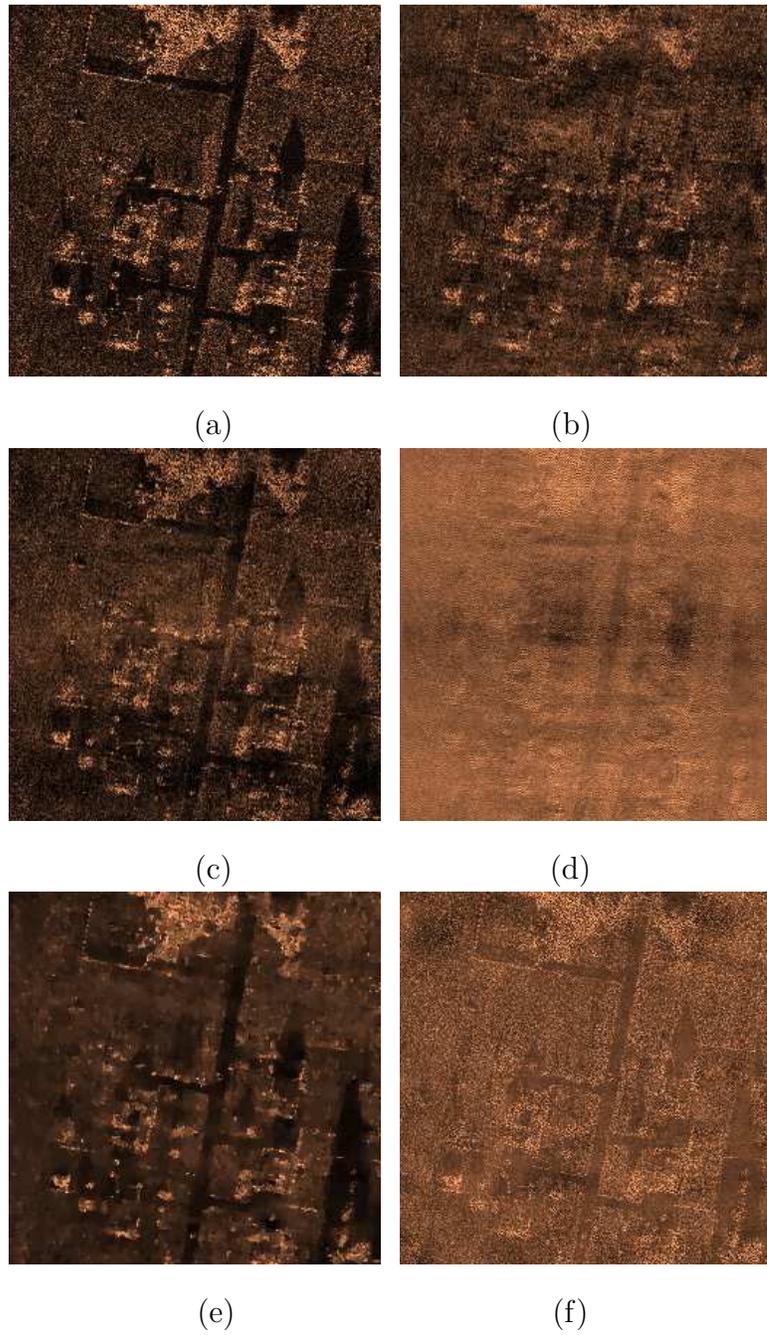
Figure 5.5: (a) Conventional SAR image. (b) Reconstructed by using PFA. (c) Reconstructed by solving (5.3). (d) The residual (i.e (a)-(c)). (e) Reconstructed by solving (5.5). (f) The residual (i.e. (a)-(e)).
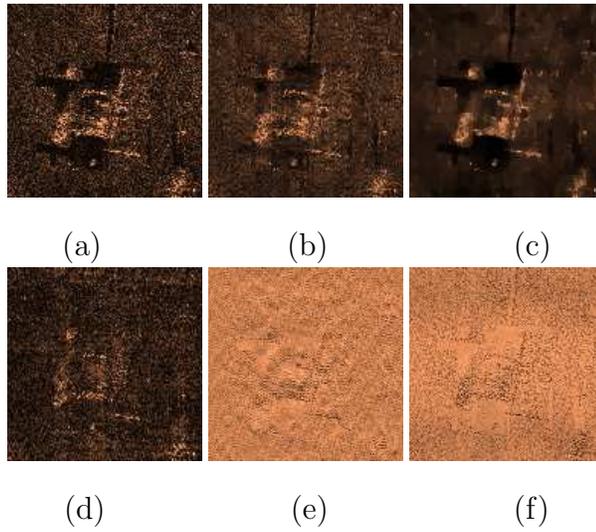
Figure 5.6: MSTAR example. (50% of data used) (a) Traditional reconstruction with the full data. (b) Reconstructed by solving (5.3). (c) Reconstructed by solving (5.5). (d) Reconstructed by using PFA. (e) The residual (i.e (a)-(b)). (f) The residual (i.e. (a)-(c)).
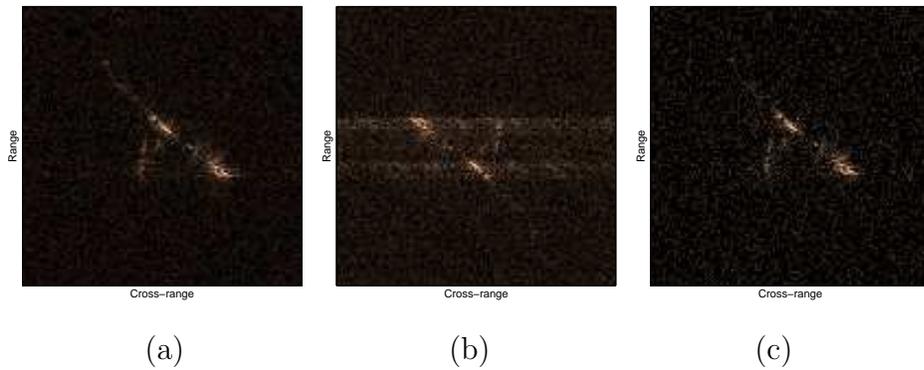
Figure 5.7: B-747 example. (a) Traditional reconstruction with the full data. (b) Traditional reconstruction with the partial data (30% of data used). (c) Reconstructed image using our proposed method.

experiments, we have noticed that taking more measurements generally improves the quality of reconstruction especially when the presence of speckle is high.

5.0.6 Chapter Summary

We have utilized CS theory and demonstrated that it is possible to compress the synthetic aperture for radar imaging. Most importantly, not only can our suggested undersampling be used in novel collection schemes to produce high quality images but many new applications such as signals intelligence, imaging much wider swaths, and reduced storage constraints are possible because of it.

Chapter 6

Dictionary-based Face Recognition

6.1   Introduction

One principle for designing recognition algorithms for visual imagery is to encode objects as sparse representations from an over-complete bases or dictionary [154]. An area of promise for sparse representation is face recognition. In face recognition, each face would be represented by a small number of atoms selected from a dictionary. The goal is that the set of atoms that represent each face is different from other faces. The ability to accomplish this goal depends on the method for selecting atoms in the dictionary. The atoms in a dictionary should produce representations that are stable across naturally occurring changes in pictures of each face.

To successfully develop a face recognition algorithm based on the principles of sparse representation a number of challenges have to be addressed. Among these are constructing a dictionary of atoms designed for discrimination that is robust to natural changes in the appearance of faces and developing computationally efficient methods. This chapter addresses the following problems in developing a face recognition algorithm based on the principles of sparse representation.

- A technique for generating dictionaries adapted to face recognition. The dictionary is designed to be robust to changes in illumination.

- Faces are encoded by a small number of atoms and hence the representation is sparse.

- A two phase computational efficient algorithms for generating sparse encodings.

  - The first phase enrolls a person into the face recognition system. The

atoms are leaned to represent the person.

- – The second phase performs recognition. A novel face is projected onto the appropriate atoms.

- A computational efficient algorithm for generating the sparse representation for each person. The sparse encoding can be generated from any number of images of a person.

6.2   Related work

Ideas from the sparse representation theory have been previously applied to face recognition. Phillips [155] proposed matching pursuit filters for face feature detection and identification. The filters are designed through a simultaneous decomposition of a training set into a 2D wavelet expansion designed to discriminate among faces. It was shown that the resulting algorithm was robust to facial expression and the surrounding environment.

Compressive sensing has shown it is possible to efficiently compress signals using a sparse representation [7],[6]. In turn, this has led to a resurgence in interest in the principles of sparse representation for recognition. Recently, Wright *et al.* [156] introduced an algorithm, called Sparse Repersentation-based Classification (SRC), based on sparse representation and compressed sensing. This work was later extended to handle pose and illumination variations in [157], [158] and for iris recognition in [159]. Nagesh and Li [160] presented an expression invariant face recognition method based on ideas from distributed compressed sensing and joint sparsity models. Also, Li *et al.* [161] presented a face recognition method based on sparse representation for recognizing 3D face meshes under expressions using low-level geometric features.

Despite the success of some of the above mentioned methods harnessing sparsity for face recognition, they suffer from some limitations. For instance, in SRC, for good recognition performance, the training images are required to be extensive enough to span the conditions that might occur in the test set. For example, to

be able to handle illumination variations in the test image, more and more training images are needed in the gallery. But in most realistic scenarios, the gallery contains only a single or a few images of each subject and it is not practical to assume the availability of multiple images of the same person under different illumination conditions. Another limitation of this approach is that the large size of the matrix due to the inclusion of the large number of gallery images can tremendously increase the computational load as well as storage needs which can make the real-time processing very difficult.

The second strand of related work is dictionary learning for sparse representation. It has been shown that for image restoration, learning dictionaries from the data instead of using pre-specified ones, usually leads to state-of-the-art results [162], [163], [164]. While these dictionaries are often trained to obtain good reconstruction, training dictionaries with a specific discriminative criteria has also been considered. For instance, linear discriminant analysis (LDA) based basis selection and feature extraction algorithm for classification using wavelet packets was proposed by Etemand and Chellappa in [165]. Recently, similar algorithms for simultaneous sparse signal representation and discrimination have also been proposed in [166], [167], and [168]. In [168], a framework for signal classification is proposed that combines a discriminative method with a reconstructive method using LDA and a pre-defined dictionary. A similar algorithm called supervised simultaneous orthogonal matching pursuit (SSOMP) is presented in [167]. The ideas presented in [167] and [168] are based on a pre-defined dictionary. In contrast with these methods, in [166], a framework that learns a non-parametric dictionary which is efficient for simultaneous sparse representation as well as class discrimination is presented. Other methods have also been proposed for learning discriminative dictionaries. See [169], [170], [171], [172], [173], [174] and the references therein for more details.

One of the most significant problems in face recognition is illumination variation [175]. The performance of most existing face recognition algorithms is highly sensitive to illumination variation. Many methods have been proposed to deal with the illumination problem in face recognition. Some of them include illumination

cone methods [176],[177], spherical harmonic-based methods [178], [179],[180], quotient images [181], [182] gradient faces [183], and logarithmic total variation method [184]. In our approach, to deal with illumination variations that might occur in the test image, we use a relighting framework based on a robust albedo estimation method presented in [185]. We generate multiple images with different lighting of the same subject and use it for training. In this setting, as will become apparent, our proposed method has the ability to recognize face images with good accuracy even in the case when only a single or a very few images are provided for training.

6.3 Dictionary-based Recognition

In face recognition, given labeled training images, the objective is to identify the class of a novel probe face image. Suppose that we are given C distinct classes and a set of m_i training images per class, $i \in \{1, \cdots, C\}$. We identify an $l \times q$ grayscale image as an N-dimensional vector, $\mathbf{x}$, which can be obtained by stacking its columns, where $N = l.q$. Let

$$\mathbf{B}_i = [\mathbf{x}_{i1}, \cdots, \mathbf{x}_{im_i}] \in \mathbb{R}^{N \times m_i} \tag{6.1}$$

be an $N \times m_i$ matrix of training images corresponding to the ith class. Similarly, we define a new matrix

$$\mathbf{A} = [\mathbf{B}_1, ..., \mathbf{B}_C] \in \mathbb{R}^{N \times M} \tag{6.2}$$

$$= [\mathbf{x}_{11}, ..., \mathbf{x}_{1m_1} | \mathbf{x}_{21}, ..., \mathbf{x}_{2m_2} | ...... | \mathbf{x}_{C1}, ..., \mathbf{x}_{Cm_C}],$$

as concatenation of training samples from all the classes, where $M = \sum_{i=1}^{C} m_i$.

For recognition, there has been a plethora of techniques proposed to explore the structure of the matrix $\mathbf{A}$. It has been observed that since human faces have similar overall configuration, the facial images can be described by a relatively low dimensional subspace. Dimensionality reduction subspace methods such as PCA [186], LDA [187], and many others [175], [178],[188] have been proposed for the task of face recognition. Images of the same person can vary significantly due to variations in the illumination conditions, expression, and pose. To recognize a

person across illumination variations, it is required to characterize the set of images under all illumination conditions. It has been found that the set of images under all possible illumination conditions can be well approximated by a 9-dimensional linear subspace [188]. Based on this result, Lee *et al.* [188] showed that there exists a configuration of 9 light source directions such that the subspace formed by the images taken under these nine sources is effective for recognizing faces under a wide range of lighting conditions.

In what follows, we propose a reconstructive approach to dictionary learning for face recognition. Our formulation is based on exploiting the structure of each $\mathbf{B}_i$ based on simultaneous sparse signal representation [189],[190],[191]. We learn multiple dictionaries, one per each class such that each learned dictionary provides an economic representation for its corresponding class and a poor reconstruction for the other classes.

6.3.1 Learning Class Specific Reconstructive Dictionaries

Designing dictionaries based on training is recent approach to dictionary design which is strongly motivated by the advances in the sparse representation theory [190],[191],[164]. We now briefly describe the method of optimal directions (MOD) [190] and the K-SVD [191] algorithms for learning dictionaries for face images. Note that the K-SVD algorithm has been applied to compressing face images in [192].

Given a set of examples $\mathbf{B} = [\mathbf{x}_1, \cdots, \mathbf{x}_m]$, the goal of the K-SVD and MOD algorithms is to find a dictionary $\mathbf{D}$ and a sparse matrix Γ that minimize the following representation error

$$(\hat{\mathbf{D}}, \hat{\Gamma}) = \arg\min_{\mathbf{D}, \Gamma} \|\mathbf{B} - \mathbf{D}\Gamma\|_F^2 \text{ subject to } \forall i \ \|\gamma_i\|_0 \leq T_0 \tag{6.3}$$

where γ_i represent the columns of Γ and the ℓ_0 sparsity measure $\|.\|_0$ counts the number of nonzero elements in the representation. Here, $\|\mathbf{A}\|_F$ denotes the Frobenius norm defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{ij} \mathbf{A}_{ij}}$. Both MOD and K-SVD are iterative methods designed to minimize the representation error in (6.3) and they alternate between sparse-coding and dictionary update steps. First, a dictionary $\mathbf{D} \in \mathbb{R}^{N \times P}$ with ℓ_2

normalized columns is initialized. For example, this can be done by randomly selecting face images from the gallery set. Then, the main iteration is composed of the following two stages:

- *Sparse coding*: In this step, $\mathbf{D}$ is fixed and the following optimization problem is solved to compute the representation vector γ_i for each example $\mathbf{x}_i$

$$i = 1, \cdots, m, \quad \min_{\gamma_i} \|\mathbf{x}_i - \mathbf{D}\gamma_i\|_2^2 \text{ s. t. } \|\gamma_i\|_0 \leq T_0.$$

Since the above problem is NP-hard, approximate solutions are usually sought. Any standard technique [27] can be used but a greedy pursuit algorithm such as orthogonal matching pursuit [44],[45] is often employed due to its efficiency.

- *Dictionary update*: This is the place where both MOD and K-SVD algorithms differ. The MOD algorithm updates all the atoms simultaneously by solving a quadratic problem whose solution is given by $\mathbf{D} = \mathbf{B}\Gamma^\dagger$, where $\Gamma^\dagger$ denotes the Moore-Penrose pseudo-inverse. Even though the MOD algorithm is very effective and usually converges in a few iterations, it suffers from the high complexity of the matrix inversion.

In the case of K-SVD, the dictionary update is performed atom-by-atom in an efficient way rather than using a matrix inversion. For a given atom l, the quadratic term in (6.3) can be rewritten as

$$\|\mathbf{B} - \sum_{i \neq l} \mathbf{d}_i \gamma_i^T - \mathbf{d}_l \gamma_l^T\|_F^2 = \|\mathbf{E}_l - \mathbf{d}_l \gamma_l^T\|_F^2, \tag{6.4}$$

where $\mathbf{E}_l$ is the residual matrix, $\mathbf{d}_l$ is the lth column (atom) of the dictionary $\mathbf{D}$ and γ_i^T are the rows of Γ. The atom update is obtained by minimizing (6.4) for $\mathbf{d}_l$ and γ_l^T through a simple rank-1 approximation of $\mathbf{E}_l$ [191]. It has been observed that the K-SVD algorithm requires less iterations to converge than the MOD method. Hence, our proposed framework is based on the K-SVD algorithm. The K-SVD algorithm is summarized in Fig. 6.3.1.

Objective: Find the best dictionary to represent the samples $\mathbf{B} = [\mathbf{x}_1, \cdots, \mathbf{x}_m]$ as sparse compositions, by solving the following optimization problem:

$$\arg\min_{\mathbf{D},\Gamma} \|\mathbf{B} - \mathbf{D}\Gamma\|_F^2 \text{ subject to } \forall i \ \|\gamma_i\|_0 \leq T_0.$$

Input: Initial dictionary $\mathbf{D}^{(0)} \in \mathbb{R}^{N \times P}$, with normalized columns, signal matrix $\mathbf{B} = [\mathbf{x}_1, \cdots, \mathbf{x}_m]$ and sparsity level T_0.

Output: Trained dictionary $\mathbf{D}$ and sparse representation matrix Γ.

Procedure: Set $J = 1$. Repeat until convergence:

- *Sparse coding stage*: Use any pursuit algorithm to compute the sparse representation vectors γ_i for each signal $[\mathbf{x}_1, \cdots, \mathbf{x}_m]$.

- *Dictionary update stage*: For each column $k = 1, \cdots, P$ in $\mathbf{D}^{(J-1)}$ update by

  - Define the group of examples that use this atom, $\omega_k = \{i | 1 \leq i \leq P, \gamma_T^k(i) \neq 0\}$.

  - Compute the overall representation error matrix, $\mathbf{E}_k$, by

  $$\mathbf{E}_k = \mathbf{B} - \sum_{j \neq k} \mathbf{d}_j \gamma_T^j.$$

  - Restrict $\mathbf{E}_k$ by choosing only the columns corresponding to ω_k and obtain $\mathbf{E}_k^R$.

  - Apply SVD decomposition $\mathbf{E}_k^R = \mathbf{U}\Delta\mathbf{V}^T$. Select the updated dictionary column $\hat{\mathbf{d}}_k$ to be the first column of $\mathbf{U}$. Update the coefficient vector γ_R^k to be the first column of $\mathbf{V}$ multiplied by $\Delta(1,1)$.

- Set $J = J + 1$.

Figure 6.1: The K-SVD algorithm.

6.3.2 Classification based on Learned Dictionaries

Given C distinct classes and m_i training images per class, let $\mathbf{B}_i$ be as defined in equation (6.1) for $i = 1, \cdots, C$. For training, we first learn C class specific dictionaries, $\mathbf{D}_i$, to represent the training samples in each $\mathbf{B}_i$, with some sparsity level T_0, using the K-SVD algorithm. Once the dictionaries have been learned for each class, given a test sample $\mathbf{y}$, we project it onto the span of the atoms in each $\mathbf{D}_i$ using the orthogonal projector $\mathbf{P}_i = \mathbf{D}_i(\mathbf{D}_i^T \mathbf{D}_i)^{-1} \mathbf{D}_i^T$. The approximation and residual vectors can then be calculated as

$$\hat{\mathbf{y}}^i = \mathbf{P}_i \mathbf{y} = \mathbf{D}_i \alpha^i \tag{6.5}$$

and

$$\mathbf{r}^i(\mathbf{y}) = \mathbf{y} - \hat{\mathbf{y}}^i = (\mathbf{I} - \mathbf{P}_i)\mathbf{y}, \tag{6.6}$$

respectively, where $\mathbf{I}$ is the identity matrix and

$$\alpha^i = (\mathbf{D}_i^T \mathbf{D}_i)^{-1} \mathbf{D}_i^T \mathbf{y} \tag{6.7}$$

are the coefficients. Since the K-SVD algorithm finds the dictionary, $\mathbf{D}_i$, that leads to the best representation for each examples in $\mathbf{B}_i$, we suspect $\|\mathbf{r}^i(\mathbf{y})\|_2$ to be small if $\mathbf{y}$ were to belong to the ith class and large for the other classes. Based on this, we can classify $\mathbf{y}$ by assigning it to the class, $d \in \{1, \cdots, C\}$, that gives the lowest reconstruction error, $\|\mathbf{r}^i(\mathbf{y})\|_2$:

$$d = \text{identity}(\mathbf{y}) = \arg\min_i \|\mathbf{r}^i(\mathbf{y})\|_2. \tag{6.8}$$

Similar approaches for texture classification using image patches have also been proposed in [193] and [194].

We now summarize our dictionary-based classification algorithm as follows: Given a test sample $\mathbf{y}$ and C training matrices $\mathbf{B}_1, \cdots, \mathbf{B}_C$ where each $\mathbf{B}_i \in \mathbb{R}^{N \times m_i}$ contains m_i training samples.

1. Learn the best dictionaries $\mathbf{D}_i$, to represent the face images in $\mathbf{B}_i$, using the K-SVD algorithm.

2. Compute the approximation vectors, $\hat{\mathbf{y}}^i$, and the residual vectors, $\mathbf{r}^i(\mathbf{y})$, using (6.5) and (6.6), respectively for $i = 1, \cdots, C$.

3. Identify $\mathbf{y}$ using (6.8).

An example of how our algorithm works is illustrated in Fig. 6.2.
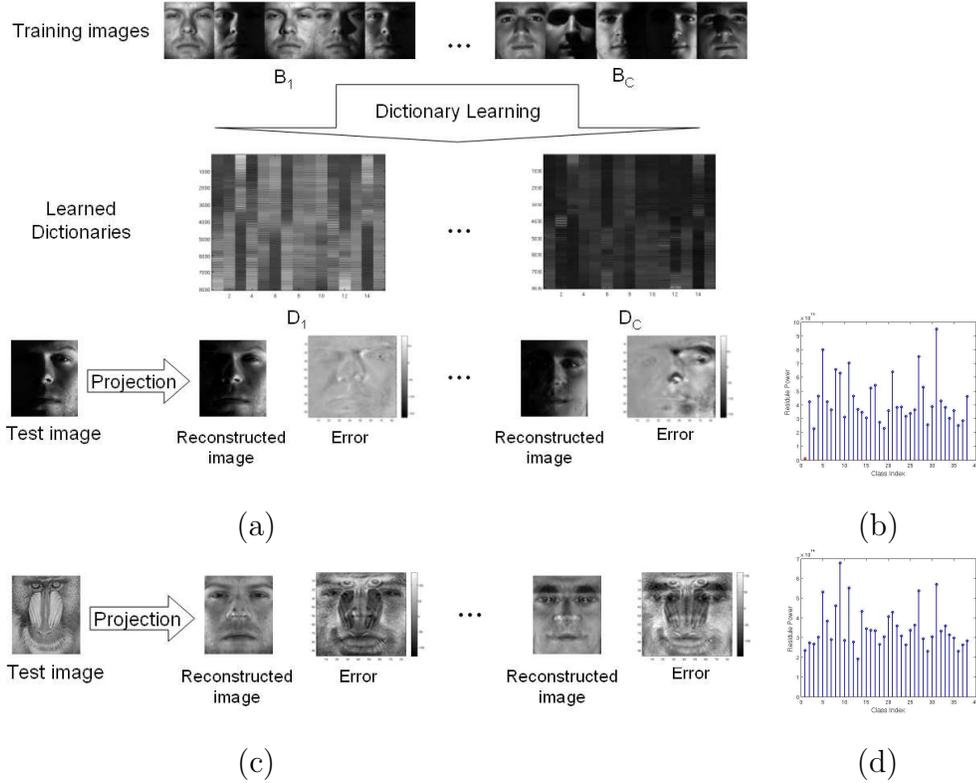
Figure 6.2: Overview of our approach. (a) Given C sets of training images corresponding to C different classes, the K-SVD algorithm is used to learn class specific dictionaries. Then, a novel test image is projected onto the span of the atoms in each of the learned dictionaries and the approximation errors are computed. (b) The class that is associated to a test vector is then declared as the one that produces the smallest approximation error. In this example, class 1 is declared as the true class. (c) and (d) illustrate an example of an invalid test image and the resulting residuals, respectively.

92

6.3.3 Dealing with Small Arbitrary Noise

An assumption underlying the treatment given above is that the test vector $\mathbf{y}$ is free of error. In practice, $\mathbf{y}$ will often be contaminated by some small noise perturbations. Hence, we consider the following more general model for $\mathbf{y}$:

$$\mathbf{y} = \tilde{\mathbf{y}} + \mathbf{z} = \mathbf{D}_i \alpha^i + \mathbf{z}, \tag{6.9}$$

where $\tilde{\mathbf{y}}$ and $\mathbf{z}$ are the underlying noise free image and random noise term, respectively. Recall that constructing an approximation $\hat{\tilde{\mathbf{y}}}$ to $\mathbf{y}$ as

$$\hat{\tilde{\mathbf{y}}}^i = \mathbf{D}_i \alpha^i$$

requires an estimation of α^i. In the case of least-squares approximation, α^i are those that give a reconstructed image that minimizes the squared error between $\mathbf{y}$ and $\hat{\tilde{\mathbf{y}}}$:

$$\hat{\alpha}^i = \min_{\alpha^i} \|\tilde{\mathbf{y}} - \mathbf{D}\alpha^i\|_2^2.$$

In this case, α^i are given by (6.7). However, it is commonly known that least-squares method is sensitive to gross errors or outliers. Hence, we need a formulation that recovers α^i from the noisy observation (6.9) in a robust way. To robustly estimate the coefficients α^i we replace the quadratic error norm with a more robust error norm. We minimize the following problem

$$\hat{\alpha}^i = \min_{\alpha^i} \|\mathbf{y} - \mathbf{D}\alpha^i\|_1,$$

where $\| \mathbf{x} \|_1 = \sum_i |(x_i)|$. The resulting estimate is known as least absolute deviation (LAD) [195] and can be solved by linear programming methods.

6.3.4 Rejection Rule

For classification, it is important to be able to detect and then reject invalid test samples. To decide whether a given test sample is valid or not, we define the following rejection rule.

Given a test image $\mathbf{y}$, for all classes in the training set, the score s_{yi} of the test image $\mathbf{y}$ to the ith class is computed as

$$s_{yi} = \frac{1}{\|\mathbf{r}^i(\mathbf{y})\|_2^2},$$

where $\mathbf{r}^i(\mathbf{y})$ is the residual vector as defined in (6.6). Then, for each test image $\mathbf{y}$, the score values are sorted in the decreasing order such that $s'_{y1} \geq s'_{y2} \geq \cdots \geq s'_{yC}$. The corresponding sorted classes are the candidate classes for each test image. The first candidate class is the most likely class that the test image belongs to. We define the ratio between the score of the first candidate class to the score of the second candidate class:

$$\lambda_y = \frac{s'_{y1}}{s'_{y2}}$$

as a measure of the reliability of the recognition rate. Based on this, a threshold τ can be chosen such that, $\mathbf{y}$ is accepted as a good image if $\lambda_y \geq \tau$, otherwise rejected as a bad image. Since the score values to all the candidate classes are sorted, the score values of the third and the higher order candidates are less than or equal to the score of the second candidate class. Hence, a high ratio λ_y for the test image $\mathbf{y}$ would show that the score of the first candidate class is significantly greater than all the other scores. Therefore, the identification result can be claimed to be reliable.

(a)　　　　　　　　　　　　　　　(b)
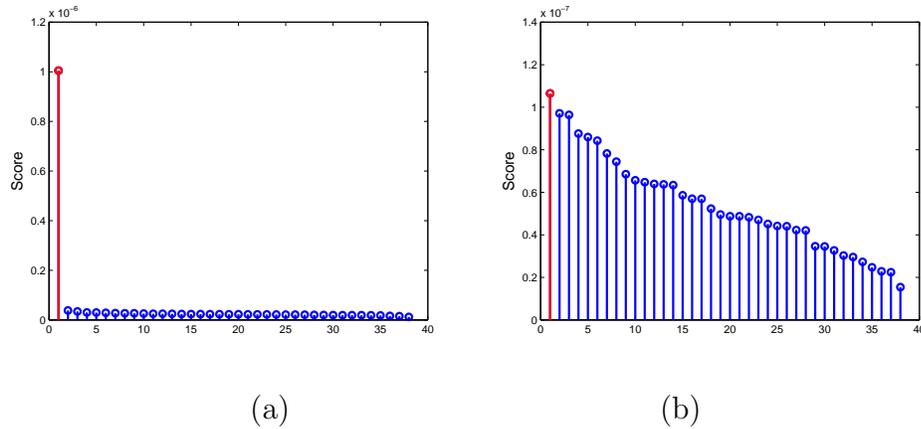
Figure 6.3: (a) and (b) are the score values corresponding to the test samples shown in Fig. 6.2(a) and Fig. 6.2(c), respectively.

To illustrate an example of how this rejection rule works, consider the test images shown in Fig. 6.2(a) and Fig. 6.2(c) and their corresponding score values in 6.3(a) and 6.3(b), respectively. Since, the image shown in Fig. 6.2(a) belongs to class 1, the corresponding ratio comes out to be $\lambda_y = 26.29$, whereas the ratio corresponding to an invalid test image shown in Fig. 6.2(c) comes out to be $\lambda_y = 1.17$. Hence, setting a threshold, τ, high enough this invalid test image can be rejected.

6.4   Face Recognition with a Single Training Image

Recognizing human faces under varying illumination given a single training image is a difficult problem. In this section, we study a method to deal with this illumination problem. The idea is to capture the illumination conditions that might occur in the test sample in the training examples.

We assume Lambertian reflectance model for the facial surface. The surface normals, albedo and the intensity image are related by an image formation model. For Lambertian objects, the diffused component of the surface reflection is modeled using the Lambert's Cosine Law given by

$$I = \rho \max(\mathbf{n}^T \mathbf{s}, 0), \tag{6.10}$$

where I is the pixel intensity, $\mathbf{s}$ is the light source direction, ρ is the surface albedo and $\mathbf{n}$ is the surface normal of the corresponding surface point. Using this model, a nonstationary stochastic filtering framework was recently proposed in [185] to estimate the albedo from a single image. We adapt this method to first estimate the albedo map from a given face image. Then, using the estimated albedo map, we generate new images under any illumination condition using the image formation model (6.10). This is done by combining the estimated albedo maps with the average facial information [196] (see [185] for more details).

It was shown in [188] that an image of an arbitrarily illuminated object can be approximated by a linear combination of the image of the same object in the same

pose, illuminated by nine different light sources placed at preselected positions. The nine pre-specified light source directions are given by [188]

$$\phi = \{0, 49, -68, 73, 77, -84, -84, 82, -50\}°$$
$$\theta = \{0, 17, 0, -18, 37, 47, -47, -56, -84\}°.$$

Hence, the image formation equation can be rewritten as

$$I = \sum_{i=1}^{9} a_i I_i, \tag{6.11}$$

where $I_i = \rho \max(\mathbf{n}^T \mathbf{s}_i, 0)$, and $\{\mathbf{s}_1, \cdots, \mathbf{s}_9\}$ are the pre-specified illumination directions. Since, we want to generate gallery images which will be sufficient to account for any illumination in the probe image, we generate images under the nine pre-specified illumination conditions and use them in the gallery. We can also re-generate the illumination conditions that might occur in the test image using nine images from pre-specified illumination directions and (6.11). This illuminated image can also be included in the gallery. Fig. 6.4 shows some relighted images and the corresponding input images.
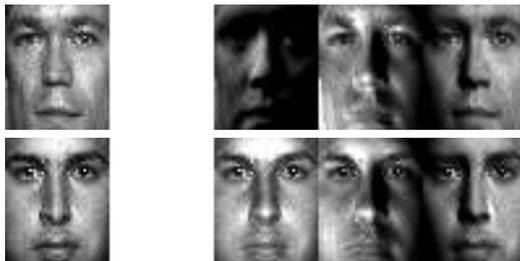
Figure 6.4: Examples of the original images (first column) and the corresponding relighted images with different light source directions from the PIE data set.

6.5   Experimental Results

In this section, we present experimental results on some of the publicly available databases for face recognition such as Extended Yale B data set [176], AR

database [197] and PIE data set [198]. The comparison with other existing face recognition methods in [156] suggests that the SRC algorithm is among the best. Hence, we treat it as state-of-the-art and use it as a bench mark for comparisons in this study.

In all of our experiments, the K-SVD [191] algorithm is used to train the dictionaries with 15 atoms unless otherwise stated. The performance of our algorithm, which we will refer to as the dictionary-based face recognition (DFR) algorithm, is compared with that of five different methods: SRC, nearest neighbor (NN), nearest subspace (NS), support vector machines (SVM) and class dependent principal component analysis (CDPCA) [199]. Our algorithm is also tested using several features, namely, Eigenfaces, Fisherfaces, Randomfaces, and downsampled images.
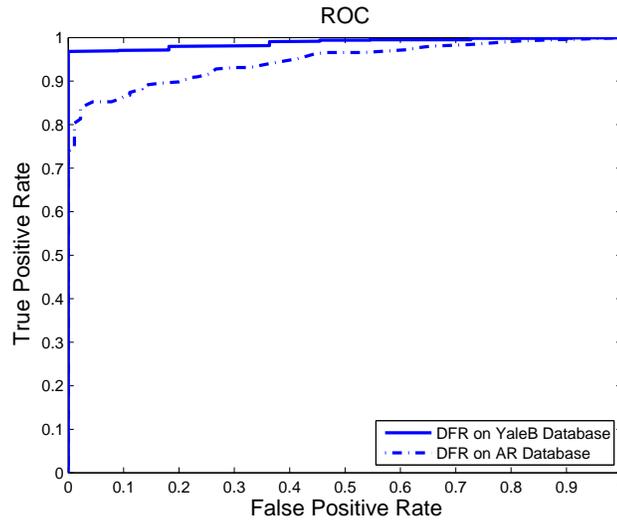
Figure 6.5: ROC curves for the proposed algorithm (DFR) corresponding to the Extended Yale B and AR face databases.

Table 6.1: Recognition result on the frontal images in the PIE data set. f_i denotes the images with the ith flash on as labeled in the PIE data set. Each $(i, j)^{th}$ entry is the rank-1 recognition rate obtained with the images in f_i as gallery and f_j as probe sets.

	f_8	f_9	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{20}	f_{21}	f_{22}	Avg	Avg [200]	Avg [201]	Avg [185]
f_8	-	100	100	100	99	99	97	94	79	99	97	97	96	89	92	87
f_9	100	-	100	100	100	99	99	99	97	99	97	97	99	93	97	95
f_{11}	100	100	-	100	100	100	99	99	96	100	99	99	99	92	95	92
f_{12}	100	100	100	-	100	100	100	99	99	100	100	100	100	96	98	98
f_{13}	100	100	100	100	-	100	100	100	100	100	100	99	100	98	100	99
f_{14}	100	99	100	100	100	-	100	100	100	100	100	100	100	99	99	98
f_{15}	96	97	99	100	100	100	-	100	100	100	100	99	99	96	97	96
f_{16}	96	97	100	100	96	99	99	-	100	97	100	100	98	91	94	93
f_{17}	78	85	94	99	97	96	99	99	-	90	96	96	93	80	87	83
f_{20}	100	100	100	100	99	99	99	99	97	-	100	100	99	91	95	93
f_{21}	97	100	100	100	100	100	100	100	99	100	-	100	100	96	99	97
f_{22}	97	97	100	99	100	100	100	100	100	100	100	-	99	98	98	97
Avg	97	98	99	100	99	99	99	99	97	99	99	99	**99**	-	-	-
Avg [200]	88	94	93	97	99	99	96	89	75	93	98	98	-	93	-	-
Avg [201]	90	97	94	99	99	99	98	93	87	95	99	99	-	-	96	-
Avg [185]	91	97	93	99	99	98	94	91	80	93	99	96	-	-	-	94

Table 6.2: Recognition result on the PIE data set. f_i denotes the images taken with the ith flash on in the PIE data set with different poses. Each $(i, j)^{th}$ entry is the rank-1 recognition rate obtained with the images in f_i as gallery and f_j as probes.

	f_8	f_9	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{20}	f_{21}	f_{22}	Avg
f_8	-	100	100	100	99	99	94	89	70	99	99	97	95
f_9	100	-	100	100	99	99	98	97	88	99	99	99	98
f_{11}	100	100	-	100	100	99	98	93	86	100	100	99	98
f_{12}	100	100	100	-	100	100	100	98	92	100	100	100	99
f_{13}	98	98	99	99	-	99	99	98	92	98	98	99	98
f_{14}	96	98	99	100	100	-	100	99	95	99	99	100	98
f_{15}	79	88	89	95	96	97	-	98	95	91	96	97	93
f_{16}	73	85	81	92	94	96	98	-	99	83	92	96	90
f_{17}	54	71	69	82	85	91	97	97	-	73	88	94	82
f_{20}	99	99	99	100	99	99	98	91	87	-	100	100	97
f_{21}	98	99	100	100	99	100	100	99	95	100	-	100	99
f_{22}	87	91	94	97	98	100	99	98	95	97	99	-	96
Avg	89	94	94	97	97	98	98	96	90	94	97	98	**95**

6.5.1   Results on Extended Yale B Database

There are a total of 2,414 frontal face images of 38 individuals in the Extended Yale B database. These images were captured under various controlled indoor lighting conditions. They were manually cropped and normalized to the size of 192×168 [188].

Our first set of experiments on the Extended Yale B data set consist of testing the performance of our algorithm on face images with different features and dimensions. We follow the experimental setup as considered in [156]. The feature space dimensions of 30, 56, 120, and 504 corresponding to the downsampling ratios of, 1/32, 1/24, 1/16, and 1/8, respectively are computed. We randomly select 32 images per subject (i.e. half of the images) for training and the other half for testing. Recognition rates of different methods with different dimensions and features are compared in Fig. 6.6. The results reported here on NN, NS, SVM and SRC are taken from [156]. Note that in our method, we also flip the gallery images horizontally which gives additional 32 more images for training.
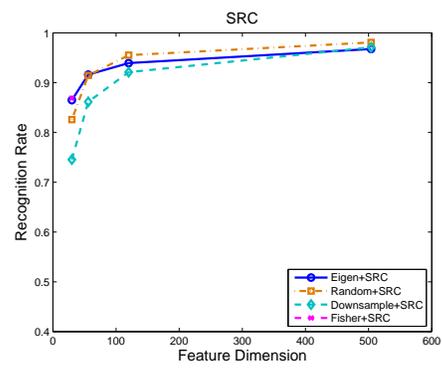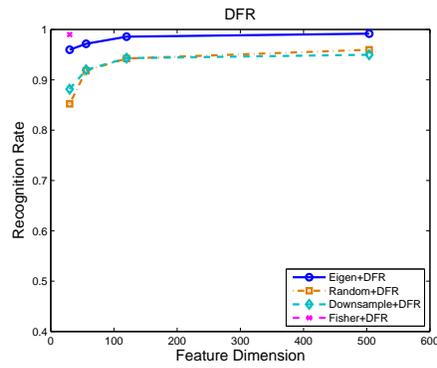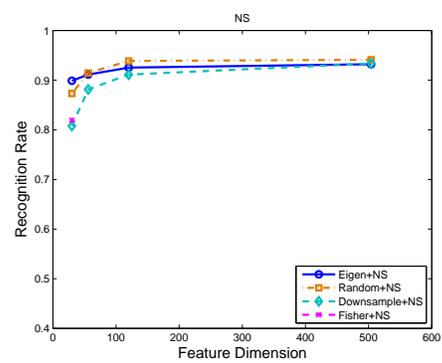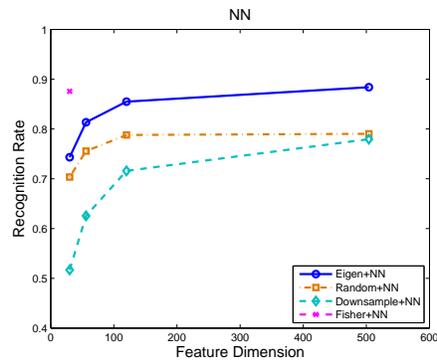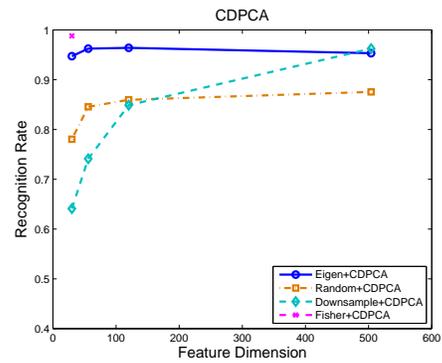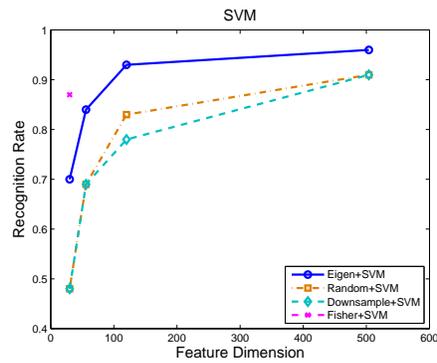
Figure 6.6: Performance comparison on the Extended Yale B database with various features, feature dimensions and methods. (a) Our method (DFR). (b) SRC. (c) NN. (d) NS. (e) SVM. (f) CDPCA.

The maximum recognition rates achieved by DFR are 95.99%, 97.16%, 98.58% and 99.17% for all $30, 56, 120$ and 504 dimensional feature spaces, respectively. The maximum recognition rate achieved by SRC is 98.1% with 504D randomfaces [156]. Also, NN, NS, SVM, and CDPCA achieve the maximum recognition rates of 90.7%, 94.1%, 97.7% and 98.83%, respectively [156]. As can be seen from Fig. 6.6 that DFR performs favorably over some of the competitive methods for face recognition on the Extended Yale B database.

6.5.2 Results on PIE Database

The PIE database consists of $41, 368$ images of 68 subjects. The images were captured under 13 different poses and 21 flashes under pose, illumination and expression variations. The face images were cropped with the size 48×40. In the first set of experiments on the PIE data set, our objective is to perform recognition across illumination with images from one illumination condition forming the gallery while images from another illumination condition forming the test set. In this setting, there is just one image per subject in each gallery and probe set. Given a single training image, we use the method described in Section 6.4 to generate 7 more training images with different lighting conditions and also flip them horizontally. Hence, a total of 16 images are included in the training set. The rank-1 results obtained using our method are reported in Table 6.1. As can be seen from Table 6.1, that our method achieves recognition rate over 99% in most of the experiments and on average it achieves the recognition rate of 99%. For comparison, we have also included the average recognition rates from [200], [201] and [185].

In the second set of experiments, we repeat the same set of experiments as done above but in the presence of different poses. Table 6.2 shows the recognition rates corresponding to these experiments. In Table 6.2, f_i denotes the images taken with the ith flash on in the PIE data set with 13 different poses. Each $(i, j)^{th}$ entry is the rank-1 recognition rate obtained with 13 images per subject for training in f_i and f_j as the probe sets. Even in the presence of different poses, our method performs reasonably well with an average recognition rate of 95%.

6.5.3 Results on AR Database

The AR database consists of over 4,000 frontal face images of 126 subjects (70 men and 56 women). The images feature frontal view faces with different facial expression, illumination variation and occlusion. This database is substantially more challenging than the Yale B database. In this experiment, we choose a subset of the images consisting of 50 male subjects and 50 female subjects. We use the images captured in Session 1 with only illumination variations as the training set. Hence, there are 7 images per subject for training [1]. The images in Session 2 with only the illumination variations are used as the test set. All the images were converted to gray scale and cropped with the size of 165×120.

The best recognition rate achieved by our algorithm is 93.7% which is a little lower than that of SRC and SVM whose reported best recognition rates are 94.7% and 95.7%, respectively [156]. NN and NS achieve the recognition rates of 89.7% and 90.3%, respectively [156] whereas CDPCA achieves the recognition rate of 59.00%. The ROC curves for DFR corresponding to the experiments described in Section 6.5.1 and this section are shown in Fig. 6.5.

6.5.4 Recognition with Partial Face Features

In this section, we report the ability of our algorithm in recognizing human faces from the partial face features. We use the images in the Extended Yale B database for the experiments on the partial face features. We replicate the experimental setup of [156]. For each subject, 32 images are randomly selected for training, and the remaining images are used for testing. The region of eye, nose and mouth are selected as partial face features. Examples of these features are shown in Fig. 6.7. Table 6.3 compares the results obtained by using our method with the other methods presented in [156]. As can be seen from the table, our method achieves recognition rates of 99.3%, 98.8% and 99.8% on eye, nose and mouth region, respectively and it

---

[1]In our approach, we use the relighting method described in Section 6.4 to generate 3 more relighted images per training image and also flip them horizontally.

outperforms the other methods such as SRC, NN, NS and SVM [156].

(a)   (b)   (c)
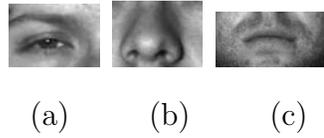
Figure 6.7: Examples of partial facial features. (a) Eye (b) Nose (c) Mouth.

Table 6.3: Recognition results with partial facial features.

	Right Eye	Nose	Mouth
Dimension	5,040	4,270	12,936
DFR	**99.3**%	**98.8**%	**99.8**%
SRC	93.7%	87.3%	98.3%
NN	68.8%	49.2%	72.7%
NS	78.6%	83.7%	94.4%
SVM	85.8%	70.8%	95.3%

6.5.5   Recognition rate vs. Number of dictionary atoms

In this section we repeat the experiment described in Section 6.5.1 on DFR using 504 dimensional eigenfaces with different number of dictionary atoms. Fig. 6.8 shows the recognition rate vs. number of atoms curve for this experiment. It can be observed that even selecting only 5 atoms per each class dictionary provides a reasonable recognition performance on the Extended Yale B database. Experiments have shown that increasing the number of atoms more than 23 usually degrades the performance of our algorithm.

6.5.6   Recognition rate vs. Number of training images

In this section, we study how the performance of our algorithm changes as we vary the number of training images in each class. Again, we use the Extended Yale B database for the experiments in this section. All the images are scaled to
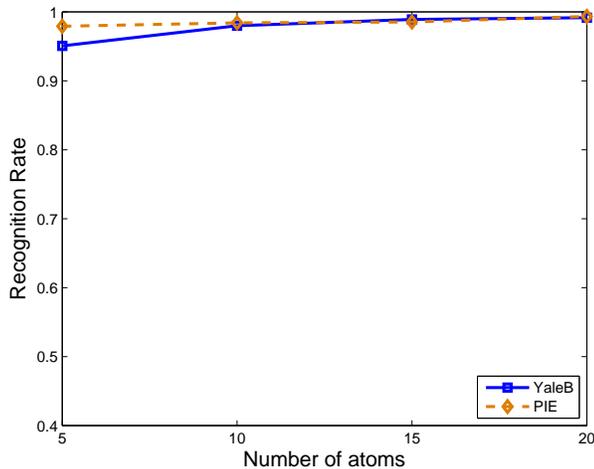
Figure 6.8: Recognition rate vs. Number of dictionary atoms on Extended Yale B database.

the size of 64×64. We randomly select 1, 2, and 3 images per subject for training and the others for testing. The recognition rates of DFR along with SRC, NS and CDPCA are compared in Table 6.4. Note that the gallery in DFR consisted of more 7 relighted images along with their flipped images. This experiment shows that even in the presence of a few training images, our method can provide reasonable recognition of human faces.

Table 6.4: Performances with different number of training samples per subject.

♯ for training images	DFR	SRC	NS	CDPCA
1	**75.89**%	42.37%	36.13%	5.52%
2	**84.71**%	37.45%	46.36%	26.22%
3	**85.18**%	37.20%	52.40%	30.25%

6.5.7 Improving the complexity of SRC

Let $\mathbf{A} = [\mathbf{B}_1, \cdots, \mathbf{B}_C] \in \mathbb{R}^{N \times M}$ be the matrix of training images as defined in (6.2). Consider a novel test image $\mathbf{y}$ of unknown class as a linear combination of the training vectors as $\mathbf{y} = \mathbf{A}\beta$, where $\beta \in \mathbb{R}^{M \times 1}$ is a vector of coefficients. We

make an assumption that given sufficient samples of the ith class, $\mathbf{B}_i$, any test image $\mathbf{y}$ that belongs to the same class will approximately lie in the linear span of the training samples from the class i. Hence, the coefficients not associated with class i will be close to zero. Then, the approach taken in [156] to recognize a face image in $\mathbf{y}$ consists of two steps. In the first step, the following problem is solved:

$$\hat{\beta} = \min_{\beta} \|\beta\|_1 \ \text{ subject to } \ \mathbf{A}\beta = \mathbf{y}. \tag{6.12}$$

In the second step, residuals from each class are computed using only the coefficients associated with each class. Then, the face in image $\mathbf{y}$ is identified as the person with the lowest residual [156].

This method is very effective, however, it suffers from very high computational complexity. For instance, if there are 100 classes and 30 samples are included for training in each class, the resulting training matrix $\mathbf{A}$ will be of size $N \times 3000$. Thus, the size of $\mathbf{A}$ will be very large. One can reduce the size of $\mathbf{A}$ without losing the performance of the algorithm by first learning class specific dictionaries where fewer atoms are selected than originally included in $\mathbf{A}$. For example, we can define a new training matrix $\tilde{\mathbf{A}}$ as a concatenation of learned dictionaries from all classes as :

$$\tilde{\mathbf{A}} = [\mathbf{D}_1, \cdots, \mathbf{D}_C]$$

where $\mathbf{D}_i$ is the learned dictionary corresponding to the class matrix $\mathbf{B}_i$. Then, solving the ℓ_1 minimization problem, (6.12), with $\tilde{\mathbf{A}}$ will reduce the complexity of SRC without decreasing the performance of the algorithm.

Table 6.5: Recognition Rates of SRC on the Extended Yale B Database with Eigenfaces as features.

Dimension	30	56	120	504
SRC (32 atoms)	86.5%	91.63%	93.95%	96.77%
SRC (15 atoms)	**75.71%**	**85.89%**	**92.10%**	**96.01%**

To verify our claim, we do an experiment with the Extended Yale B database where, for each class, we randomly select 32 images for training and the remaining images for testing. We choose Eigenfaces as features and learn class specific

dictionaries with only 15 atoms with the K-SVD algorithm. We then concatenate these learned dictionaries and solve (6.12) with this matrix. The resulting recognition rates are reported in Table 6.5 along with the recognition rates achieved with the traditional SRC with 32 images for training per each class [156]. It is clear from Table 6.5 that reducing the size of the training matrix $\mathbf{A}$ by learning class specific dictionaries can reduce the complexity of SRC without loosing much of its performance at the same time being robust to outliers, occlusion and corruptions.

6.6 Chapter Summary

In this chapter, we studied a face recognition algorithm based on learning dictionaries for simultaneous sparse signal approximations and feature extraction. Experiments indicate that our algorithm can perform better than many competitive face recognition methods harnessing sparsity. One of the main advantages of our algorithm is that it is computationally efficient. Using a unix system with Intel Xeon E5506/2.13 GHz processor, on average our algorithm takes about 0.3 seconds to train a dictionary of 15 atoms for a gallery matrix containing 32 images of size 24×21. Recognizing a test sample from 38 classes takes on average about 0.01 seconds. Also, our experiments indicate that DFR has the ability to recognize human faces even in the case when a few number of examples are given for training. Furthermore, we incorporated a relighting approach into the proposed method to deal with the illumination conditions.

Note that in this chapter, we studied a reconstructive approach to face recognition. It is also possible to learn discriminative dictionaries for face recognition. However, the resulting algorithm is very slow and it does not improve the recognition performance over that of the reconstructive dictionaries.

Chapter 7

Sparsity inspired ATR

In this chapter, an Automatic Target Recognition (ATR) algorithm is presented based on a framework for learning dictionaries for simultaneous sparse signal representation and feature extraction [12]. In the previous chapter, we took a reconstructive approach to dictionary learning, whereas, in this chapter, we learn discriminative dictionaries. The dictionary learning algorithm is based on class supervised simultaneous orthogonal matching pursuit while a matching pursuit-based similarity measure is used for classification. We show how the proposed framework can be helpful for efficient utilization of data, with the possibility of developing real-time, robust target classification. We verify the efficacy of the proposed algorithm using confusion matrices on the well known Comanche forward-looking infrared data set consisting of ten different military targets at different orientations.

7.1 Introduction

In ATR, the objective is to classify each target image into one of a number of classes. However, the presence of high clutter background, sensor noise, the large number of target classes, and the computational load involved in processing all the sensor data has often hampered the development of real-time robust ATR algorithms. The recognition algorithm usually consists of several stages such as target detection, background noise removal, feature extraction and classification. In this work, we mainly focus on the last two stages. Target recognition using forward-looking infrared (FLIR) imagery of different targets in natural scenes is difficult due to large variations in the thermal signatures of targets. Many ATR algorithms have been proposed for FLIR imagery. In [202], an ATR algorithm for FLIR imagery based on modular neural network was proposed. Wavelet based vector quantization was used for FLIR ATR in [203]. See [204] for an excellent survey of papers and

experimental evaluation of FLIR ATR.

Recently, a sparse representation-based face recognition algorithm was proposed in [156], which outperformed many state of the art algorithms. Extensions based on [156] for FLIR ATR were recently presented in [205]. As we discussed in the previous section, one of the main limitations of this approach is that for good recognition performance, the training images are required to be extensive enough to span the conditions that might occur in the test set. This may not be the case in many practical scenarios. Another limitation of this approach is that the large size of the matrix due to the inclusion of large number of gallery images can tremendously increase the computational complexity which can make the real-time processing very difficult.

To overcome the aforementioned limitations, in this chapter, we present an ATR algorithm based on learning class supervised dictionaries for simultaneous sparse signal representation and classification.

7.2 Simultaneous Signal Representation

In this section, we show how simultaneous orthogonal matching pursuit can be used for ATR.

7.2.1 Simultaneous Orthogonal Matching Pursuit (SOMP)

Let D be a redundant dictionary with K atoms in $\mathbb{R}^n$. The elements of the dictionary are indexed by $\gamma \in \Gamma$, i.e

$$D = \{\phi_\gamma : \gamma \in \Gamma\} \subset \mathbb{R}^n.$$

The atoms have unit Euclidean norm i.e., $\|\phi_\gamma\|_2 = 1, \forall \gamma \in \Gamma$. Let $X = [x_1, ..., x_s]$ be a set of training signals, where $x_i \in \mathbb{R}^n$ denotes the ith signal of X. Given D and X, SOMP attempts to approximate these signals at once as a linear combination of a common subset of atoms of cardinality much smaller than n [189]. Under the assumption that these signals belong to a certain class, SOMP extracts their

Input: Dictionary D, signal matrix X, sparsity level T (i.e. number of atoms).

Output: A set Λ_T containing T indices, approximation A and residual matrix R.

Procedure:

1. Initialize the residual $R_0 = X$, $\Lambda = \emptyset$, and $t = 1$.

2. Find index γ_t, which solves the optimization problem

$$\arg \max_{\gamma \in \Gamma} \|R_t^T \phi_\gamma\|_1.$$

3. Set $\Lambda_t = \Lambda_{t-1} \cup \{\gamma_t\}$.

4. Determine the orthogonal projector P_t onto the span of the atoms indexed in Λ_t.

5. Compute the new approximation and residual:

$$A_t = P_t X,$$
$$R_t = (I - P_t)X.$$

6. If $t = T$, then stop. Otherwise, increment $t = t + 1$, and go to step 2.

Figure 7.1: SOMP algorithm.

common internal structure [189]. In fact, by keeping the sparsity low enough, one can eliminate the internal variation of the class which can lead to more accurate recognition while being robust to noise [189],[166],[167]. The SOMP algorithm is summarized in Fig. 7.1. In what follows, we show that after adding a discriminative term into SOMP, how we can use the coefficients of sparse representation together with the residual, over a class specific learned dictionary for recognition.

7.2.2 Separability based SOMP (SSOMP)

To further increase the discriminative power of SOMP, we adapt a supervised learning algorithm based on linear discriminant analysis (LDA). Note that

LDA-based basis selection and feature extraction algorithm for classification using wavelet packets was proposed by Etemand and Chellappa in [165]. Recently, similar algorithms for simultaneous representation and discrimination have also been proposed in [166], [167], and [168].

Let us denote the number of classes by c and assume that

$$X = [X^{(1)}, \cdots, X^{(c)}] \in \mathbb{R}^{n \times m},$$

where $X^{(j)} = [x_1^{(j)}, \cdots, x_{n_i}^{(j)}] \in \mathbb{R}^{n \times n_i}$ denotes the samples that belong to the jth class that has n_i samples and $m = cn_i$. To obtain a supervised atom selection algorithm, we modify the SOMP algorithm by adding a separability constraint that captures within-class and between-class variations. Define the within-class scatter matrix S_w as

$$S_w = \sum_{i=1}^{c} S_i, \tag{7.1}$$

where

$$S_i = \sum_{k=1}^{n_i} (x_k^{(i)} - \mu^{(i)})(x_k^{(i)} - \mu^{(i)})^T, \tag{7.2}$$

and $\mu^{(i)} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_k^{(i)}$. One can also define the between-class scatter matrix S_b as

$$S_b = \sum_{i=1}^{c} n_i (\mu^{(i)} - \mu)(\mu^{(i)} - \mu)^T, \tag{7.3}$$

where $\mu = \frac{1}{cn_i} \sum_{i=1}^{c} n_i \mu^{(i)}$ is the total mean vector. In order to achieve good separability for classification, one needs to have large between-class scatter and small within-class scatter simultaneously. This can be achieved by introducing various cost functions [166],[167],[165]. In this chapter, we use the following cost function

$$J(X) = Tr(S_w^{-1} S_b) \tag{7.4}$$

but similar results can be obtained by using any of the other cost functions defined in [166],[167],[165].

For a dictionary D and a set of indices Λ, let $\Phi_\Lambda \in \mathbb{R}^{n \times |\Lambda|}$ be the matrix induced by the restriction of the dictionary elements whose indices are the elements

of Λ. Then, the sparsity coefficients are given by $\alpha_k^{(j)} = (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T x_k^{(j)}$. From this observation, one can show that

$$Sb(\alpha) = (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T Sb(X) \Phi_\Lambda (\Phi_\Lambda^T \Phi_\Lambda)^{-1}$$

$$Sw(\alpha) = (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T Sw(X) \Phi_\Lambda (\Phi_\Lambda^T \Phi_\Lambda)^{-1}.$$

Hence, we can write the optimization problem that we want to solve in step 2 of the SOMP algorithm (to get the supervised SOMP) as follows

$$\arg \max_{\gamma \in \Gamma} \left(\|R_t^T \phi_\gamma\|_1 + \lambda J(\alpha) \right), \tag{7.5}$$

where $\lambda \geq 0$ controls the trade-off between discrimination and reconstruction. We call the resulting algorithm supervised SOMP (SSOMP).

7.2.3 Classification Using SOMP and SSOMP

Once the dictionaries are learned for each class, one can design a classifier based on either residuals (i.e. approximation error) or coefficients. For instance, SOMP (or SSOMP) approximations of the test sample g can be found using the learned dictionaries. The test sample can then be assigned the label of the class whose dictionary gives the best approximation of g (i.e. the smallest residual). However, a test signal may find an economic representation in many dictionaries. Hence, the approximation error by itself may not be the most reliable measure for classification.

The approach of comparing coefficient vectors of projected and original objects have also been proposed for classification [155]. Also, in [167] and [168, 166], nearest neighbor (NN) and SVM classifiers are used on the coefficient vectors for classification, respectively.

Since, the matching pursuit approximation defines a signal s in terms of its projection, the coefficient vector and the residual, we propose to use these for classification. Let P_s be the projection operator defined by the dictionary learned for the class containing s. Let $\alpha(s, P_s)$ be the coefficient vector and $R(s, P_s)$ be the residual.
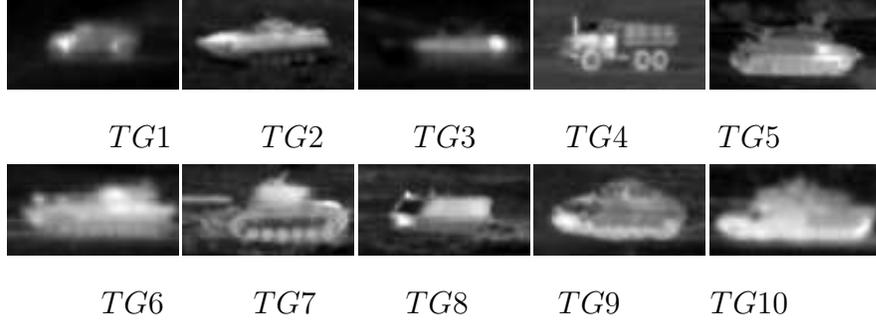
Figure 7.2: Side view of all 10 targets present in the SIG data set.

Then, in order to compare two signals g and s, we project g onto the projection P_s of s and noting the coefficient vector $\alpha(g, P_s)$ and residual $R(g, P_s)$. Based on these, the matching pursuit dissimilarity measure (MPDM)[206] has been defined as

$$\delta(g, s) = \sqrt{\theta F_R(g, s) + (1 - \theta) F_\alpha(g, s)}, \qquad (7.6)$$

where $\theta \in [0, 1]$ determines the importance of the residuals and coefficients in δ, $F_R(g, s)$ is the difference between the residuals of g and s when both samples are projected onto the projection P_s of s

$$F_R(g, s) = \|R(g, P_s) - R(s, P_s)\|^2 \qquad (7.7)$$

and $F_\alpha(g, s)$ compares their corresponding coefficient vectors

$$F_\alpha(g, s) = \|\alpha(g, P_s) - \alpha(s, P_s)\|^2. \qquad (7.8)$$

Note that MPDM is a dissimilarity measure as small values indicate similar signals, while large values indicate dissimilar signals [206]. Once the class specific dictionaries are learned, the classification is accomplished using the NN classification rule in the MPDM sense. To further increase the recognition performance, one can also perform the k-NN in terms of MPDM.

It is also simple to introduce a reject threshold using the MPDM; if the value is too big, then the sample is considered not to belong to any class and should be rejected.

112

7.3 Experimental Results

In this section, we present some preliminary results of our proposed algorithm on the Comanche FLIR data set consisting of different military targets at different orientations.

7.3.1 Dataset

The data set contains 10 different vehicle targets. We will denote these targets as $TG1, TG2, \cdots, TG10$. For each target, there are 72 orientations, corresponding to aspect angles of $0°, 5°, \cdots, 355°$ in azimuth. The data consists of a training set and a test set. We will refer to the training set as the SIG set and the test set as the ROI set. The SIG data set has about 13,816 image chips, while there are 3,353 images in the ROI data set. The SIG data set consists of the images that were collected under very favorable conditions. The SIG data set contains 874 to 1468 images per target class. The ROI set consists of only five targets, namely $TG1, TG2, TG3, TG4$ and $TG7$. The target images for the ROI set were taken under less favorable conditions, such as targets with different weather conditions, in different background, in and around clutter; hence, these data are very challenging. There are 577 to 798 images for each of these five target classes. The images are of size 40×75 pixels. All the images in the SIG and ROI sets were normalized to a fixed range with the target put approximately in the center. The orientation in the ROI set was given very coarsely; every $45°$. In Fig. 7.2 we show side view of all the 10 targets present in the SIG set.

7.3.2 Dictionary

In our experiments, the dictionary, D, contained about 1500 elements. It consisted of 2-D DCT atoms, 2-D Daubechies (4-taps) wavelet atoms, Gabor atoms and a few target chips. Fig. 7.3 shows some of the atoms from our dictionary.
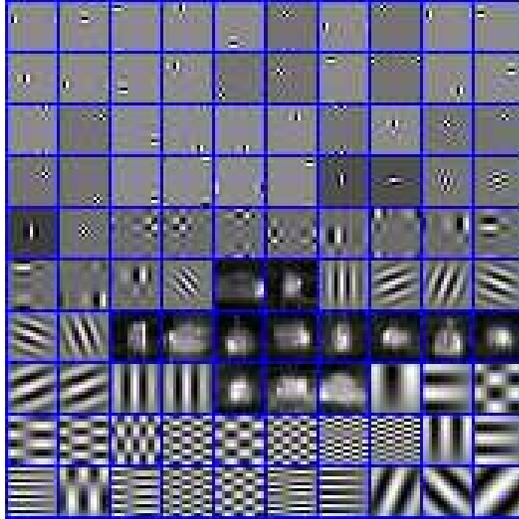
Figure 7.3: A few 16×16 atoms from the dictionary, D.

7.3.3 Results

In the first set of experiments, we randomly selected 11 targets per aspect angle from the SIG data set for training, called TRAIN-SIG, and another set of 1000 targets disjoint from the training data for testing, called TEST-SIG. We used SOMP and SSOMP for training class specific dictionaries with 10 atoms. The value for λ in (7.5) was chosen to be 0.2 and the θ value in (7.6) was fixed to 0.5. In all the experiments, the target chips size was reduced from 40×75 to 16×16. Given c target classes, $\omega^1, \cdots, \omega^c$, each represented by its own separate dictionary, the classification rule we use is the following

$$\text{if } \delta(x, x_k^{(j)}) < \delta(x, x_k^{(l)}), \forall j \neq l, \ \forall k = 1, \cdots, n_i$$

then classify x into ω^j. The probabilities of correct classification for this experiment are 93.60 and 94.80 percent for the SOMP and SSOMP, respectively. The confusion matrices for this experiment are shown in Fig. 7.4 (a) and (c) for SOMP and SSOMP, respectively.

In the second set of experiments, we again randomly selected 11 targets per aspect angle from the SIG data set for training. We randomly chose a set of 1000 targets from the ROI data set for testing, called TEST-ROI. Again, we used SOMP
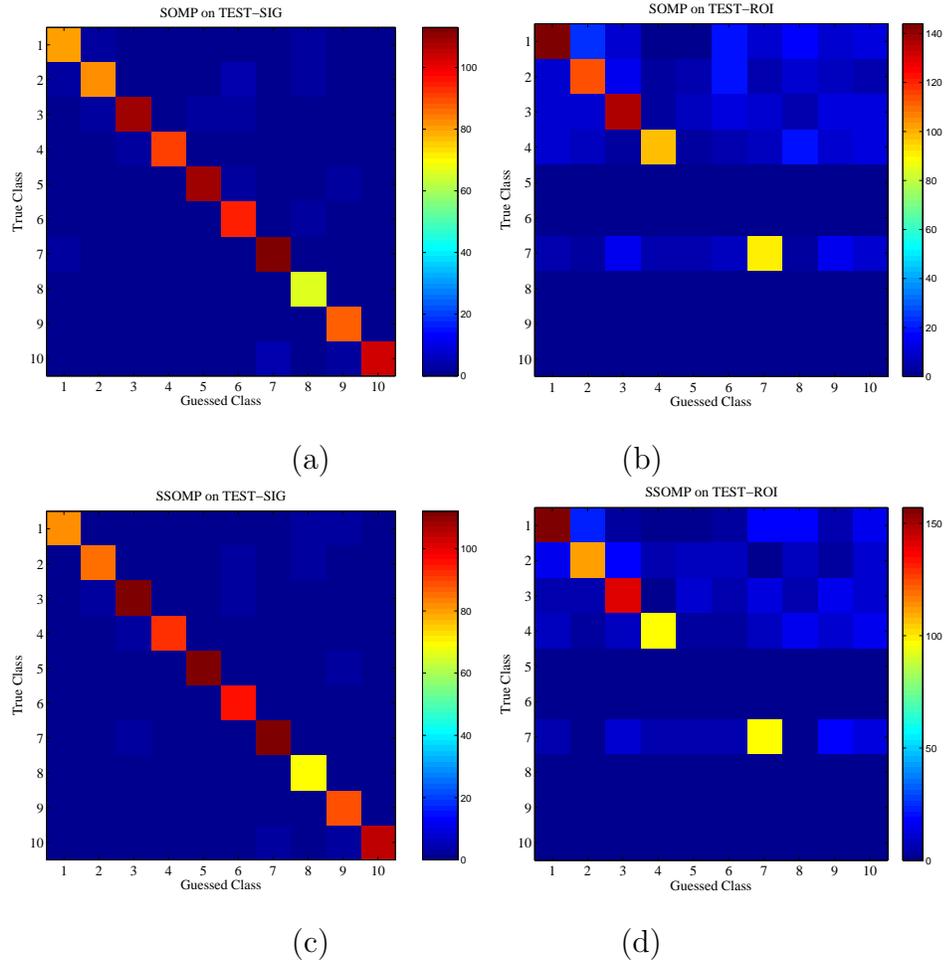
Figure 7.4: Confusion matrices. (a) SOMP on TEST-SIG. (b) SOMP on TEST-ROI. (c) SSOMP on TEST-SIG. (d) SSOMP on TEST-ROI.

and SSOMP for training class specific dictionaries with 10 atoms. The same values for λ and θ were used as before. The probabilities of correct classification for this experiment are 71.89 and 76.19 percent for the SOMP and SSOMP, respectively. The confusion matrices for this experiment are shown in Fig. 7.4 (b) and (d) for SOMP and SSOMP, respectively.

Table 7.1: Recognition rates (in %) for different methods.

Methods	CNN4	MNN	LVQ	SOMP	SSOMP
TRAIN-SIG	95.16	95.49	99.72	100	100
TEST-SIG	-	90.53	-	93.60	94.80
TEST-ROI	59.25	75.58	75.12	71.89	76.19

From the above experiments, it is clear that introducing a discriminative term into SOMP generally improves the classification performance over SOMP. Also, note that our method is more general than the methods presented in [202] and [203]. In their methods, to deal with the background artifacts, they use several rectangular windows of different size based on the ground truth silhouette computer-aided design models. As a result, their performance significantly depends on the choice of windows [202],[203],[204]. In contrast, the method presented here does not require any windowing. Results obtained using different techniques are compared in Table. 1, where CNN4, MNN and LVQ stand for 4 layered convolutional neural network [204], modular neural network [202] and learning vector quantization [203], respectively.

7.4 Chapter Summary

In this chapter, we studied a framework for simultaneous sparse signal representation for robust ATR. Supervised SOMP was proposed to learn discriminative class specific dictionaries. The classification rule was based on a dissimilarity measure that combined both the coefficient vector and the residuals. Promising results

were obtained on a difficult FLIR target data set.

Chapter 8

Directions for Future Work

Several future directions of inquiry are possible considering the problems addressed in this dissertation and the methods proposed to solve them. We discuss a few below.

8.1 Dictionary-based Deconvolution

Many complicated techniques that go beyond trimming wavelet coefficients by a threshold parameter have been proposed, such as methods based on the hidden Markov models [207]. It is very likely that utilization of such techniques adapted to the shearlet domain will provide an even greater performance when combined to a deconvolution routine. It would also be of interest to apply the methods developed here to blind deconvolution when the knowledge of the point spread function is not assumed.

It has been observed that for image restoration, learning dictionaries from the data instead of using pre-specified ones, usually leads to state-of-the-art results [162], [163], [164]. One can develop a deconvolution method based on learning dictionaries.

8.2 Compressive Synthetic Aperture Sonar Imaging

Synthetic Aperture Sonar (SAS) is an imaging technology that forms an image of the complex reflectivity of targets or an area by coherently combining successive returns that are transmitted, scattered by various targets and received by the system [208], [209]. In SAS, it is possible to achieve an along-track resolution that is independent of range and signal frequency. It is determined only by the width of the real aperture. The stripmap mode SAS imaging geometry is shown in Fig. 8.1.

One can extend the CS techniques considered for SAR to sonar. It may offer
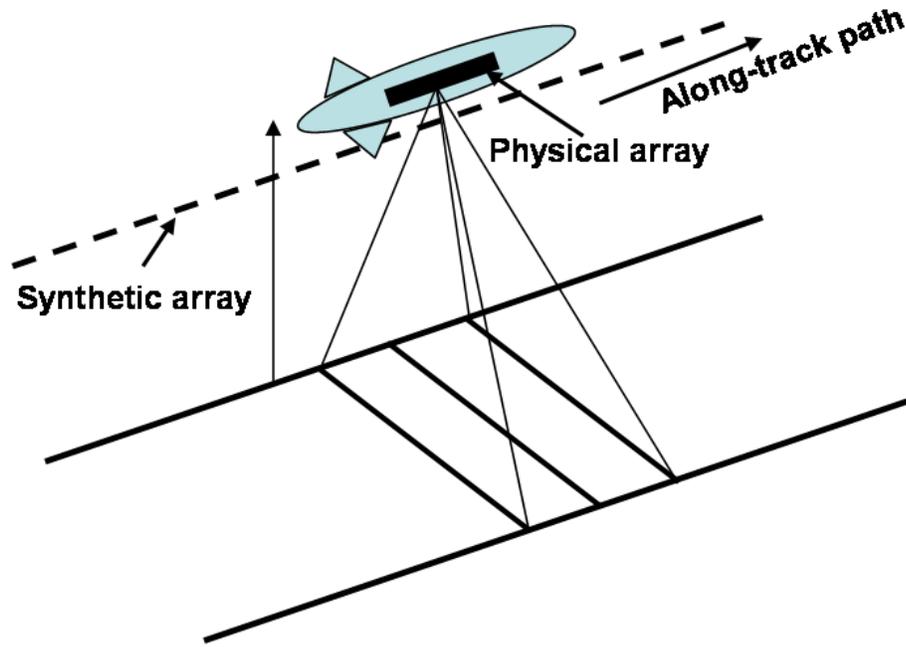
Figure 8.1: The stripmap mode SAS imaging geometry.

many advantages such as reduced sampling requirements, higher area coverage rates and relaxed sensing geometry requirements. One can also devise a method for recognizing underwater objects directly from the compressive measurements without explicitly reconstructing the image.

8.3 Reduction of motion artifacts using CS techniques

In [210], an algorithm for imaging a time varying object from its projection at different fixed times was introduced. One could extend the ideas presented in [210] using the theory of CS to get better reconstruction. Also, these techniques can be used for imaging moving targets in SAR and sonar [57], [211].

8.4 Robust albedo estimation

One of the main disadvantages of the method presented in [185] for albedo estimation is that it requires the images to be aligned and it can not handle cast

shadows. An interesting avenue for research would be to come up with an albedo estimation technique that can handle cast shadows and is robust to registration errors.

Bibliography

[1] R. Neelamani, H. Choi, and R. G. Baraniuk, "Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Transactions on Image Processing*, vol. 52, no. 2, pp. 418–433, Feb. 2004.

[2] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic, New York, 1998.

[3] G. R. Easley, D. Labate, and W. Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Journ. Amer. Statist. Assn.*, vol. 25, no. 1, pp. 25–46, July 2008.

[4] V. M. Patel, G. R. Easley, and D. M. Healy, "Shearlet-based Deconvolution," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2673–2685, Dec. 2009.

[5] V. M. Patel, G. R. Easley, R. Chellappa, and D. M. Healy, "Enhancing sparsity using gradients for compressive sensing," *Proceedings of IEEE International Conference on Image Processing*, Cairo, Egypt, Oct. 2009.

[6] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[8] V. M. Patel, G. R. Easley, D. M. Healy, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 244–254, April 2010.

[9] ——, "Compressed sensing for synthetic aperture radar imaging," *International Conference on Image Processing*, Cairo, Egypt, Oct. 2009.

[10] V. M. Patel, G. R. Easley, and D. M. Healy, "A new multiresolution generalized directional filter bank design and application in image enhancement," *International Conference on Image Processing*, pp. 2816–2819, 2008.

[11] V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Sparsity inspired automatic target recognition," *Defense, Security and Sensing, Proceedings of SPIE*, Orlando, April 2010.

[12] ——, "Automatic target recognition based on simultaneous sparse representation," *IEEE International Conference on Image Processing*, Hong Kong, Sep. 2010.

[13] K. Guo, W. Lim, D. Labate, G. Weiss, and E. Wilson, "Wavelets with composite dilations and their mra properties," *Appl. Computat. Harmon. Anal.*, vol. 20m, pp. 231–249, 2006.

[14] G. R. Easley, V. Patel, and D. M. J. Healy, "A m-channel directional filter bank compatible with the contourlet and shearlet frequency tiling," *in Wavelets XII, Proceedings of SPIE*, San Diego, CA, Aug. 2007.

[15] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM J. Math. Anal.*, vol. 39, no. 1, pp. 298–318, 2007.

[16] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, June 2002.

[17] E. J. Candes, "Compressive sampling," *International Congress of Mathematics, Madrid, Spain*, vol. 3, pp. 1433–1452, 2006.

[18] E. J. Candes and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[19] G. R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, July 2007.

[20] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, March 2008.

[21] L. Donoho, D. and Y. Tsaig, "Recent advances in sparsity-driven signal recovery," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. v713–v716, Philadelphia, PA, March 2005.

[22] O. Holtz, "Compressive sensing: a paradigm shift in signal processing," *2009 Joint mathematics meeting of AMS and MAA, Washington, DC*, Dec 2008.

[23] S. Mallat, *A wavelet tour of signal processing, second edition.* Academic Press, San Diego, CA, 1999.

[24] E. Candes and J. Romberg, "Practical signal recovery from random projections," *Proceedings of the SPIE*, vol. 5674, pp. 76–86, 2005.

[25] E. Candes and T. Tao, "Decoding by linear programing," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[26] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 533–548, March 2006.

[27] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[28] E. Candes and J. Romberg, "Quantitatively robust uncertainty principles and optimally sparse decompositions," *Foundations of Comput. Math.*, vol. 6, no. 2, pp. 227–254, April 2006.

[29] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[30] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, August 2006.

[31] E. J. Candes and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[32] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, New York, NY, 2004.

[34] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematics Programming*, vol. 95, no. 1, pp. 3–51, 2003.

[35] T. Blumensath and E. Davies, "Iterative thresholding for sparse approximations," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, Dec. 2008.

[36] I. Daubechies, M. Defries, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, 2004.

[37] T. Blumensath and E. Davies, "Iterative hard thresholding for compressed sensing," *preprint.*

[38] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/ Thresholding Algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[39] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–598, 2007.

[40] E. v. Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," Department of Computer Science, University of British Columbia, Tech. Rep.

[41] W. Yin, S. Osher, J. Darbon, and D. Goldfarb, "Bregman iterative algorithms for compressed sensing and related problems," UCLA CAM, Tech. Rep.

[42] T. Goldstein and S. Osher, "The split Bregman algorithm for L1 regularized problems," UCLA CAM, Tech. Rep.

[43] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse Reconstruction by Separable Approximation," *preprint.*

[44] S. Mallat and Z. Zhang, "Matching Pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, p. 33973415, 1993.

[45] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive function approximation with applications to wavelet decomposition," *In Proc. 27th Ann. Asilomar Conf. Signals, Systems, and Computers*, Nov. 1993.

[46] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2006.

[47] T. Blumensath and M. E. Davies, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370–2382, June 2008.

[48] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by Stagewise Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, March 2006, preprint.

[49] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, vol. 26, pp. 301–321, 2008.

[50] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, pp. 707–710, Oct. 2007.

[51] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, March 1997.

[52] E. J. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.

[53] M. Lustig, D. Donoho, and M. J. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[54] D. C. J. Munson, J. D. O'Brian, and W. K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proceedings of the IEEE*, vol. PROC-71, pp. 917–925, Aug. 1983.

[55] C. V. J. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-mode synthetic aperture radar: a signal processing approach.* Springer, New York, NY, 1996.

[56] M. Çetin, *Feature-Enhanced Synthetic Aperture Radar Imaging.* Ph.D. Thesis, Boston University, College of Engineering, 2001.

[57] C. L. Logan, *An Estimation-Theoretic Technique for Motion-Compensated Synthetic-Aperture Array Imaging.* Ph.D. Thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, Feb. 2000.

[58] W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms.* Artech House, Norwood, MA, 1995.

[59] K. Tomiyasu, "Tutorial review of Synthetic-Aperture Radar (SAR) with applications to imaging of the ocean surface," *Proceedings of the IEEE*, vol. 66, no. 5, pp. 563–583, May 1978.

[60] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing.* Wiley, New York, NY, 1991.

[61] M. A. Richards, *Fundamentals of Radar Signal Processing.* McGraw-Hill, New York, NY, 2005.

[62] V. C. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis.* Artech House, Norwood, MA, 2002.

[63] G. E. Stimson, *Introduction to airborne radar.* SciTech Publishing Inc., Mendham, NJ, 1998.

[64] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[65] M. D. Desai and W. K. Jenkins, "Convolution backprojection image reconstruction for spotlight mode synthetic aperture radar," *IEEE Transactions on Image Processing*, vol. 1, no. 4, pp. 505–517, Oct. 1992.

[66] A. E. Brito, *Iterative adaptive extrapolation applied to SAR image formation and sinusoidal recovery.* Ph.D. Thesis, The University of Texas at El Paso, Dept. of Electrical and Computer Engineering, July 2001.

[67] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, July 1995.

[68] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, March 1992.

[69] S. R. Degraaf, "SAR imaging via modern 2-D spectral estimation methods," *IEEE Transactions on Image Processing*, vol. 7, no. 5, pp. 729–761, May 1998.

[70] D. L. Donoho and M. Johnstone I., "Adapting to unknown smoothness via wavelet shrinkage," *Journ. Amer. Statist. Assn.*, vol. 90, no. 432, pp. 1200–1224, 1995.

[71] ——, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, December 1994.

[72] D. L. Donoho, M. Vetterli, D. R. A., and I. Daubechies, "Data compression and harmonic analysis," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.

[73] D. Labate, W. Q. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," *Proc. SPIE, Wavelet Applications in Signal and Image Processing XI*, vol. 5914, pp. 254–262, San Diego, CA, Sep. 2005.

[74] D. L. Donoho, "Sparse components of images and optimal atomic decomposition," *Constr. Approx.*, vol. 17, pp. 353–382, 2001.

[75] G. R. Easley, D. Labate, and F. Colonna, "Shearlet-based total variation diffusion for denoising," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 260–268, Feb. 2009.

[76] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[77] A. L. Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.

[78] E. J. Candès and D. L. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges," *in Curves and Surface Fitting, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press*, 1999.

[79] E. J. Candès, D. L., D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, 2006.

[80] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Appl. Comput. Harmon. Anal.*, vol. 2, pp. 101–126, April 1995.

[81] E. J. Candés and D. L. Donoho, "Recovering edges in ill-posed inverse problems: optimality of curvelet frames," *Ann. Statist.*, vol. 30, no. 3, pp. 784–842, 2002.

[82] F. Abramovich and B. W. Silverman, "Wavelet decomposition approaches to statistical inverse problems," *Biometrika*, vol. 85, no. 1, pp. 115–129, Oct. 1998.

[83] J. Starck, M. K. Nguyen, and F. Murtagh, "Wavelets and curvelets for image deconvolution: a combined approach," *Signal Processing*, vol. 83, no. 10, pp. 2279–2283, 2003.

[84] R. R. Coifman and D. L. Donoho, "Translation invariant de-noising," *Wavelets and Statistics, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag*, no. 10, pp. 125–150, 1995.

[85] M. K. Ng, R. H. Chan, and W. Tang, "A fast algorithm for deblurring models with Neumann boundary conditions," *SIAM J. Sci. Comput.*, vol. 21, no. 3, pp. 851–866, 1999.

[86] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion.* SIAM, Philadelphia, 1998.

[87] J. Kalifa, M. S., and R. B., "Deconvolution by thresholding in mirror wavelet bases," *IEEE Transactions on Image Processing*, vol. 12, no. 4, pp. 446–457, 2003.

[88] I. M. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo, "Wavelet deconvolution in a periodic setting," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, vol. 66, no. 3, pp. 547–573, 2004.

[89] M. Pensky and B. Vidakovic, "Adaptive wavelet estimator for nonparametric density deconvolution," *Ann. Statist.*, vol. 27, no. 6, pp. 2033–2053, 1999.

[90] F. Jianqing and J. Koo, "Wavelet deconvolution," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 734–747, Mar. 2002.

[91] R. N. Neelamani, M. Deffenbaugh, and R. G. Baraniuk, "Texas two-step: A framework for optimal multi-input single-output deconvolution," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2752–2765, Nov. 2007.

[92] W. H. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, pp. 55–59, 1972.

[93] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical Journal*, vol. 79, pp. 745–754, 1974.

[94] L. Landweber, "An iterative formula for fredholm integral equations of the first kind," *American Journal of Mathematics*, vol. 73, no. 3, pp. 615–624, July 1951.

[95] J. L. Starck, A. Bijaoui, and F. Murtagh, "Multiresolution support applied to image filtering and deconvolution," *CVIP: Graphical Models and Image Processing*, vol. 57, pp. 420–431, 1995.

[96] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a gem algorithm exploiting a class of heavy-tailed priors," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 937–951, Apr. 2006.

[97] M. Elad, B. Matalon, and M. Zibulevsky, "Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization," *Appl. Comput. Harmon. Anal.*, vol. 23, pp. 346–367, Nov. 2007.

[98] G. Teschke, "Multi-frame representations in linear inverse problems with mixed multi-contraints," *Appl. Comput. Harmon. Anal.*, vol. 22, pp. 43–60, 2007.

[99] I. Daubechies, G. Teschke, and L. Vese, "Iteratively solving linear inverse problems," *Inverse Problems and Imaging*, vol. 1, no. 1, pp. 29–46, 2007.

[100] M. J. Fadili and J. L. Starck, "Sparse representations-based image deconvolution by iterative thresholding," *Preprint*, 2007.

[101] H. Takeda, S. Farsiu, and P. Milanfar, "Deblurring using regularized locally adaptive kernel regression," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 550–563, Apr. 2008.

[102] I. Daubechies, M. Fornasier, and I. Loris, "Accelerated projected gradient method for linear inverse problems with sparsity constraints," *Preprint*, 2008.

[103] C. Vonesch and M. Unser, "A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 539–549, Apr. 2008.

[104] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[105] P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," *Proceedings of IEEE International Conference on Image Processing*, Thessaloniki, Greece, Oct. 2001.

[106] V. Katkovnik, K. Egiazarian, and J. Astola, "A spatially adaptive nonparametric regression image deblurring," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1469–1478, Oct. 2005.

[107] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[108] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[109] N. Weyrich and G. T. Warhola, "Wavelet shrinkage and generalized cross validation for image denoising," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 82–90, Jan. 1998.

[110] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Processing*, vol. 56, no. 1, pp. 33–44, Jan. 1997.

[111] M. Jansen and A. Bultheel, "Multiple wavelet threshold estimation by generalized cross validation for images with correlated noise," *IEEE Transactions on Image Processing*, vol. 8, no. 7, pp. 947–953, July 1999.

[112] F. Luisier, T. Blu, and M. Unser, "A new sure approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, March 2007.

[113] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill 3rd edition, New York, 2002.

[114] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*. Wiley, New York, 1977.

[115] A. D. Hillery and R. T. Chin, "Iterative wiener filters for image restoration," *IEEE Transactions on Signal Processing*, vol. 39, no. 8, pp. 1892–1899, Aug. 1991.

[116] S. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical wiener filtering," *Proc. SPIE, Wavelet Applications in Signal and Image Processing V*, vol. 3169, pp. 389–399, San Diego, CA, Oct. 1997.

[117] R. Maleh, A. C. Gilbert, and M. J. Strauss, "Sparse gradient image reconstruction done faster," *International Conference on Image Processing*, vol. 2, pp. 77–80, Oct. 2007.

[118] A. Agrawal, R. Raskar, and R. Chellappa, "What is the range of surface reconstruction from a gradient field?" *Proceedings of IEEE European Conference on Computer Vision*, Graz, Austria, May 2006.

[119] A. Agrawal, R. Chellappa, and R. Raskar, "An algebraic approach to surface reconstruction from gradient fields," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, pp. 174–181, Oct. 2005.

[120] R. Chartrand, "Nonconvex compressive sensing and reconstruction of gradient-sparse images: random vs. tomographic fourier sampling," *International Conference on Image Processing*, San Diego, California, Oct. 2008.

[121] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-ray Sci. Tech.*, vol. 14, pp. 119–139, 2006.

[122] T. Simchony, R. Chellappa, and M. Shao, "Direct analytical methods for solving poisson equations in computer vision problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 435–446, May 1990.

[123] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, 1990, NUMBER = 7, PAGES = 629-639.

[124] J. Weickert, "A review of nonliner diffusion filtering," *In scale-space theory in computer vision, Lecture Notes in Comp. Science. (Springer, Berlin)*, 1997.

[125] J. Romberg, "L1-Magic," *http://www.acm.caltech.edu/l1magic/*.

[126] S. G. Mobley and M. W. Maier, "Synthetic aperture radar with a non-uniform pulse repetition interval," *27th Southeastern Symposium on System Theory, SSST*, pp. 498–502, 1995.

[127] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and Implementation of an Analog-to-Information Converter using Random Demodulation," *IEEE International Symposiaum on Circuits and Systems*, pp. 1959–1962, New Orleans, LA, May 2007.

[128] J. Laska, S. Kirolos, Y. Massoud, R. G. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss, "Random Sampling for Analog-to-Information Conversion of Wideband Signals," *IEEE CAS Workshop*, pp. 119–122, Dallas, Tx, Oct. 2006.

[129] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: efficient sampling of sparse bandlimited signals," *IEEE Trans. Info. Theory*, submitted, Jan. 2009.

[130] R. Baraniuk and P. Steeghs, "Compressive radar imaging," *Proc. IEEE Radar Conf.*, p. 128133, Waltham, MA, Apr. 2007.

[131] A. C. Gurbuz, J. H. McClellan, and W. R. Scott, "Compressive sensing for GPR imaging," *41st Asilomar Conference on Signals, Systems and Computers*, pp. 2223–2227, Pacific Grove, CA, Nov. 2007.

[132] ——, "A compressive sensing data acquisition and imaging method for stepped frequency GPRs," *IEEE Transactions on Signal Processing, To appear*, vol. 57, no. 7, pp. 2640–2650, July 2009.

[133] M. Herman and T. Strohmer, "High resolution radar via compressed sensing," *IEEE Trans. Signal Processing*, vol. 57, no. 6, pp. 2275–2284, June 2009.

[134] J. Romberg, "Compressive sampling by random convolution," *SIAM Journal on Imaging Science*, vol. 2, no. 4, pp. 1098–1128, Nov. 2009.

[135] S. Bhattacharya, T. Blumensath, and M. Davis, "Fast encoding of synthetic aperture radar raw data using compressed sensing," *IEEE wrokshop on SSP*, Aug. 26–29 2007.

[136] A. C. Fannjiang, P. Yan, and T. Strohmer, "Compressed remote sensing of sparse objects," *submitted*, April 2009.

[137] R. L. Cook, "Stochastic sampling in computer graphics," *ACM Transactions on Graphics*, vol. 5, no. 1, pp. 51–72, Apr. 1986.

[138] M. Dippe and E. Wold, "Antialiasing through stochastic sampling," *SIGGRAPH*, vol. 19, no. 3, pp. 69–78, Nov. 1985.

[139] A. V. Balakrishnan, "On the problem of time jitter in sampling," *IRE Trans. Inf. Theory*, vol. 8, no. 3, pp. 226–236, Apr. 1962.

[140] G. Hennenfent and F. J. Herrmann, "Simply denoise: wavefield reconstruction via jittered undersampling," *Geophysics*, vol. 73, no. 3, pp. v19–v28, June 2008.

[141] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of nois," *IEEE Trans. Info. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[142] J. Provost and F. Lesage, "The application of compressed sensing for photo-acoustic tomography," *IEEE Trans. Medical Imaging*, vol. 28, no. 4, pp. 585–594, April 2009.

[143] J. Goodman, "Some fundamental properties of speckle," *Jour. Opt. Soc. Amer.*, vol. 66, pp. 1145–1150, Nov. 1976.

[144] S. Shi, J. Osher, "A nonlinear inverse scale space method for a convex multiplicative noise model," *Tech. Rep. 07-10, Computational and Applied Math., Univ. of California, Los Angeles*, 2007.

[145] M. Çetin, W. C. Karl, and A. S. Willsky, "Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging," *Optical Engineering*, vol. 54, no. 1, p. 017003, January 2006.

[146] R. J. L. Montero, "Pulse-echo image formation using nonquadratic regularization with speckle-based images," *M.S. thesis, Univ. of Illinois Urbana-Champaign*, 2005.

[147] E. Candes and J. Romberg, "Signal recovery from random projections," *in Proc. of SPIE Computational Imaging III*, vol. 5674, 2005.

[148] W. W. Goj, *Synthetic Aperture Radar and Electronic Warfare.* Artech House, Norwood, MA, 1992.

[149] I. G. Cumming and F. H. Wong, *Digital processing of synthetic aperture radar data*. Artech House, Norwood, MA, 2005.

[150] P. N. Stoyle, "Synthetic aperture radar assemply and a method of creating a radar image of a planet surface using such an assembly," *United States patent number 5051749*, Sep. 24, 1991.

[151] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete and Computational Geometry*, vol. 35, no. 4, p. 617652, 2006.

[152] J. D. Blanchard, C. Cartis, and J. Tanner, "The restricted isometry property and ℓ_q-regularization: Phase transition for sparse approximation," *submitted*, 2009.

[153] "Center for imaging science image database, web page, http://cis.jhu.edu/data.sets/mstar/."

[154] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[155] P. J. Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 150–164, 1998.

[156] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[157] J. Huang, X. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, Ancorage, Alaska, June, 2008.

[158] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Towards a practical face recognition system: Robust registration and illumination by sparse representation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 597–604, Miami, FL, 2009.

[159] J. Pillai, V. M. Patel, and R. Chellappa, "Sparsity inspired selection and recognition of iris images," *Third IEEE International Conference on Biometrics : Theory, Applications and Systems*, Washington D.C., Sept. 2009.

[160] P. Nagesh and B. Li, "A compressive sensing approach for expression-invariant face recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1518–1525, Miami, FL, June 2009.

[161] X. Li, T. Jia, and H. Zhang, "Expression-insensitive 3d face recognition using sparse representation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2575–2582, Miami, FL, June 2009.

[162] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[163] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, Jan. 2008.

[164] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, submitted 2009.

[165] K. Etemand and R. Chellappa, "Separability-based multiscale basis selection and feature extraction for signal and image classification," *IEEE Transactions on Image Processing*, vol. 7, no. 10, pp. 1453–1465, Oct. 1998.

[166] F. Rodriguez and G. Sapiro, "Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries," *Tech. Report, University of Minnesota*, Dec. 2007.

[167] E. Kokiopoulou and P. Frossard, "Semantic coding by supervised dimensionality reduction," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 806–818, Aug. 2008.

[168] K. Huang and S. Aviyente, "Sparse representation for signal classification," *NIPS*, vol. 19, pp. 609–616, 2007.

[169] M. Ranzato, F. Haung, Y. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[170] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," *Advances in Neural Information Processing Systems*, Vancouver, B.C., Canada 2006.

[171] J. Mairal, F. Bach, J. Pnce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," *Proc. of the Conference on Computer Vision and Pattern Recognition*, Anchorage, AL, June 2008.

[172] J. Mairal, M. Leordeanu, F. Bach, M. Herbert, and J. Ponce, "Discriminative sparse image models for class-specific edge detection and image interpretation," *Proc. of the European Conference on Computer Vision*, 2008.

[173] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," *International Conference on Machine Learning*, Montreal, Canada, June 2009.

[174] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," *Advances in Neural Information Processing Systems*, Vancouver, B.C., Canada, Dec. 2008.

[175] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, pp. 399–458, Dec. 2003.

[176] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Ilumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, June 2001.

[177] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?" *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1996.

[178] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[179] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object," *J. Optical Soc. Am.*, vol. 18, no. 10, pp. 2448–2459, 2001.

[180] L. Zhang and D. Samaras, "Face recognition under variable lighting using harmonic imae exemplars," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, Nice,France, Oct. 2003.

[181] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying iluminations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129–139, Feb. 2001.

[182] H. Wang, S. Z. Li, and Y. Wang, "Generalized quotient image," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, 2004.

[183] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Trans. Imag. Proc.*, vol. 18, no. 11, pp. 2599–2606, Nov. 2009.

[184] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1519–1524, Sep. 2006.

[185] S. Biswas, G. Aggarwal, and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 884–899, Mar. 2009.

[186] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[187] P. Belhumeur, J. Hespanda, and D. Kriegman, "Eigenfaces versus fisher-faces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, July 1997.

[188] K. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, May 2005.

[189] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, March 2006.

[190] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, pp. 2443–2446, 1999.

[191] M. Aharon, M. Elad, and A. M. Bruckstein, "The k-svd: an algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[192] O. Bryt and M. Elad, "Compression of facial images using the K-SVD algorithm," *J. Vis. Commun. Image R.*, vol. 19, pp. 270–282, 2008.

[193] G. Peyre, "Sparse modeling of textures," *Journal of Mathematical Imaging and Vision*, vol. 34, no. 1, pp. 17–31, May 2009.

[194] K. Skretting and J. H. Husoy, "Texture classification using sparse frame-based representations," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–11, 2006.

[195] p. J. Huber and E. M. Ronchetti, *Robust Statistics, second ed.* Wiley, New Jersey, 2009.

[196] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, Sept. 2003.

[197] A. Martinez and R. Benavente, "The ar face database," *CVC Technical Report 24*, 1998.

[198] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[199] M. Deriche, "A simple face recognition algorithm using eigeneyes and a class-dependent pca implementation," *International Journal of Soft Computing*, vol. 3, pp. 438–442, 2008.

[200] S. Zhou, R. Chellappa, and D. Jacobs, "Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints," *Proc. of European Conference on Computer Vision*, pp. 588–601, 2004.

[201] G. Aggarwal and R. Chellappa, "Face recognition in the presence of multiple light sources," *Proc. of International Conference on Computer Vision*, pp. 1169–1176, 2005.

[202] L. Wang, S. Z. Der, and N. M. Nasrabadi, "Automatic target recognition using a feature-decomposition and data-decomposition modular neural network," *IEEE Transactions on Image Processing*, vol. 7, no. 8, pp. 1113–1121, Aug. 1998.

[203] L. A. Chan and N. M. Nasrabadi, "An application of wavelet-based vector quantization in target recognition," *International Journal on Artificial Intelligence Tools*, vol. 6, no. 2, pp. 165–178, 1997.

[204] B. Li, R. Chellappa, Q. Zheng, S. Der, N. M. Nasrabadi, L. Chan, and L. Wang, "Experimental evaluation of flir ATR approaches - a comparative study," *Computer Vision and image understanding*, vol. 84, pp. 5–24, 2001.

[205] V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Sparsity inspired automatic target recognition," *SPIE Defense and Security Symposium, Automatic Target Recognition XX*, April 2010.

[206] R. Mazhar, P. D. Gader, and J. N. Wilson, "Matching-pursuits dissimilarity measure for shape-based comparison and classification of high-dimensional data," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 5, pp. 1175–1188, Oct. 2009.

[207] M. Crouse, R. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[208] W. W. J. Bonifant, *Interferometric synthetic aperture sonar processing.* M.S. Thesis, Georgia Institute of Technology, Atlanta, Georgia, USA, Department of Electrical Engineering, July 1999.

[209] H. J. Callow, *Signal processing for synthetic aperture sonar image enhancement.* Ph.D. Thesis, University of Canterbury, Christchurch, New Zealand, Department of Electrical and Electronic Engineering, April 2003.

[210] D. Healy, T. Olson, and J. Weaver, "Reduced motion artifacts in medical imaging by adaptive spatio-temporal reconstruction," *Numerical Algorithms*, vol. 9, no. 1, pp. 55–84, 1995.

[211] I. Stojanovic and W. C. Karl, "Imaging of moving targets with multi-static SAR using an overcomplete dictionary," *preprint*.