Distance-weighted regularization for compressed-sensing video recovery and supervised hyperspectral classification

By

Eric W. Tramel

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Computer Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2012

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Distance-weighted regularization for compressed-sensing video recovery

and supervised hyperspectral classification

By

Eric W. Tramel

Approved:

James E. Fowler Professor and Graduate Coordinator of Electrical and Computer Engineering (Major Professor)	Robert Moorhead Professor of Electrical and Computer Engineering (Committee Member)
Qian Du Associate Professor of Electrical and Computer Engineering (Committee Member)	Song Zhang Associate Professor of Computer Science and Engineering (Committee Member)
Sarah A. Rajala Dean of the James Worth Bagley College of Engineering	

Name: Eric W. Tramel

Date of Degree: December 15, 2012

Institution: Mississippi State University

Major Field: Computer Engineering

Major Professor: Dr. James E. Fowler

Title of Study: Distance-weighted regularization for compressed-sensing video recovery and supervised hyperspectral classification

Pages of Study: 117

Candidate for Degree of Doctor of Philosophy

The compressed sensing (CS) model of signal processing, while offering many unique advantages in terms of low-cost sensor design, poses interesting challenges for both signal acquisition and recovery, especially for signals of large size. In this work, we investigate how CS might be applied practically and efficiently in the context of natural video. We make use of a CS video acquisition approach in line with the popular single-pixel camera framework of blind, non-adaptive, random sampling while proposing new approaches for the subsequent recovery of the video signal which leverage inter-frame redundancy to minimize recovery error. We introduce a method of approximation, which we term multi-hypothesis (MH) frame prediction, to create accurate frame predictions by comparing hypotheses drawn from the spatial domain of chosen reference frames to the non-overlapping, block-by-block CS measurements of subsequent frames. We accomplish this frame prediction via a novel distance-weighted Tikhonov regularization technique. We verify through our experiments that MH frame prediction via distance-weighted regularization provides

state-of-the-art performance for the recovery of natural video sequences from blind CS measurements.

The distance-weighted regularization we propose need not be limited to just frame prediction for CS video recovery, but may also be used in a variety of contexts where approximations must be generated from a set of hypotheses or training data. To show this, we apply our technique to supervised hyperspectral image (HSI) classification via a novel classifier we term the nearest regularized subspace (NRS) classifier. We show that the distance-weighted regularization used in the NRS method provides greater classification accuracy than state-of-the-art classifiers for supervised HSI classification tasks. We also propose two modifications to the core NRS classifier to improve its robustness to variation of input parameters and and to further increase its classification accuracy.

Key words: Compressed Sensing, Video Compression, Block Compressed Sensing, Supervised Classification, Hyperpsectral Imagery

DEDICATION

To Laura et al.

ACKNOWLEDGEMENTS

It takes a village to raise a child.

– Someone (presumably wise)

Perhaps a politician invented it, but I prefer the romantic idea that this saying came up from the bosom of the savanna. If there had been such a thing as a Doctorate of Philosophy at that time, out in the bush among the lions and elephants, I think that the (perhaps imaginary) wizened elder would have said something different. Looking out across the crackling fire and into the eyes of the prospective student, shadows darting along the packed dirt and landing on gently swaying grasses, they would have said "*Don't you dare drag us into this.*" Following which, spitting and sauntering off into the distance, their disappearance hastened as a cloud passes over the moon, the fire inexplicably goes out.

Unlike this tale, however, I have been surrounded by a great train of supporters who have lent me their time, their minds, their strength, and, most of all, their patience. I cannot presume to have accomplished anything by my own hand alone. Even this dissertation is built upon the contributions of many tireless investigators who themselves were lifted up by their many supporters. I imagine you have your own train, urging you on as you trudge through this very tome. Never let these people forget how much they have done for you, lest you find yourself alone, under the stars, moon hidden, chill overtaking, and hyenas cackling.

To this end, I have many thanks to give. My wife, first among my cohort, you knew exactly what you were getting into and you jumped in anyway. Any amount of Scottish snow-driving PTSD is worth it for you. I cannot thank you enough for everything you've sacrificed. My parents, you brought me into this world and have been beside me ever since. You have been the fire, the balm, and the strong girders underneath it all. Thank you. Pops & Ladybug, I cannot express how much you have done for Laura and I in love and prayer. Forrest, Ginny, Stirling, Sarah, you always bring joy into my heart and encourage me to continue on, thank you for your warm embraces. Ruby, Bud, you two can't read yet, but someday you will, and I can only hope that you'll cite my papers. My grandparents, how can we thank you for your labors? Some things we can never repay, some boots we long to fill. I hope this is close enough to mathematics.

Dr. Fowler, thank you for giving me this opportunity, your endless advice, your patience at my many questions, your help, your hard work, your example, and for convincing me to ski the second day. My committee members, Dr. Moorhead, Dr. Du, and Dr. Zhang, thank you for your help, apt insights, and keen eyes in reviewing the work presented here. Sungkwang, you've been excellent over the years and your joy and generosity is infectious. Wei, Chen, Nam, Vineetha, Azam, Zheng, thank you for your friendship and help throughout the many problems we've faced together. Axel, thank you for your encouragement on this work and for a great summer. And of course, many thanks to all the investigators, researchers, graduate students, and professors whose correspondence on these topics has been so invaluable to me.

Chris, our countless discussions and adventures are absolutely without price. Anna, you are strong like coffee and tall as a giraffe, thank you for being both laughter and sandpaper. Paul, Donna, the things you've left behind can fill books. I've profited so much from your sacrifices, I hope that I can be brave enough to do the same for others. Josiah, Kimberly, Wesley, Laura, Jeremy, Joy, Daniel, Josie, Carroll, Alon, Steven, Jeff, Jason, thank you.

Anything I accomplish belongs to you.

– Eric

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Compressed sensing (CS) [14, 17], a new signal-sampling and recovery model, has emerged in recent years with much excitement and many papers detailing its impact on many different areas of research such as natural-image acquisition, remote sensing, cognitive radio, and medical imaging, just to name a few. CS combines both signal acquisition and dimensionality reduction into a single step, potentially reducing computational and hardware burdens on sensing devices. The promise of CS to lower sampling costs or to inform more effective signal-recovery strategies has even helped to open new avenues, such as terahertz imaging [20] and single-pixel cameras [35].

For these applications, research has largely focused on the blind recovery of CS-acquired signals. Much work has been done in this area with many different varieties of solvers proposed over the past several years in an effort to decrease recovery computation time without sacrificing distortion performance. For the most part, these reconstruction strategies are oblivious to the structure of the signal being recovered beyond a general assumption of sparsity, or compressibility, in some transform basis. Recently, however, several reconstruction techniques have focused on situations in which side information about signal content is available to aid signal recovery. Some proposed methods, such as Bayesian CS

[54] as well as model CS [37, 4], exploit certain *a priori* knowledge of signal structure, or the probability thereof, to guide recovery. These methods, however, do not directly address the situation in which one or more predictions of the signal to be recovered are available to the CS reconstruction process. Signal prediction techniques are especially prevalent in video-processing applications such as source coding. Typically in video coding, one or more reference frames are used to make predictions of some target frame such that the resulting residual frame representing the difference between the two has dramatically lowered signal energy leading to more efficient representation and processing.

In this dissertation, we consider the CS recovery of video sequences in which frame-to-frame predictions are used to aid the CS-recovery process. In effect, we perform CS recovery on the prediction residual which is, in most cases, significantly more compressible than the original frame, resulting in a higher-quality CS recovery. Key to our approach is the use of motion estimation (ME) and motion compensation (MC) such that the frame-to-frame predictions compensate for object motions between frames. Such use of ME/MC derives from traditional video-compression algorithms which make extensive use of sophisticated MC strategies.

One form of MC widely employed in traditional video compression is that of multihypothesis (MH) prediction in which multiple, distinct predictions are created and then combined to yield a composite prediction superior to any of the constituent single-hypothesis (SH) predictions [98]. As a primary contribution of this dissertation, we show how such MH prediction can be incorporated into the CS recovery of video sequences so as to increase reconstruction quality over equivalent SH-driven recovery. Central to this discussion

is a formulation of the MH prediction process in the domain of the random CS projections; as this formulation results in a ill-posed optimization, we resort to Tikhonov regularization [103] which is widely used to yield tractable solutions to such ill-posed problems. In experimental results, we compare our proposed Tikhonov-based regularization against an alternative strategy enforcing an ℓ_1-based sparsity constraint on the MH predictions [30]. We find that our proposed approach usually yields significantly superior reconstruction, particularly when the video frames are acquired at very low subsampling rate.

We have published a number of papers which make use of side information to aid CS recovery of images. In one set of papers, we considered the application of prediction-aided residual recovery for single views from a multiview image set [105] as well as for an entire set of images using successive stages of recovery [106]. We combined the ideas presented in these two conference papers in the journal article [107], which is currently in review. We have also written specifically on the topic of video recovery in two conference papers [104, 21]. We have also written a comprehensive long-form work on block-based CS of still images and video using the techniques proposed here in [42]. We additionally presented an extension of the block-based compressed sensing with smoothed projected Landweber (BCS-SPL) [76, 77] acquisition and recovery framework, multiscale BCS-SPL (MS-BCS-SPL), which performs acquisition within the wavelet domain [41].

Additionally, we discuss an extension of the MH technique for supervised classification tasks, which we term the nearest regularized subspace (NRS) classifier. Specifically, we investigate the effectiveness of the NRS classifier for hyperspectral classification tasks when operating on hyperspectral-image (HSI) data acquired from aerial platforms. In our

experiments we show how this technique achieves better average classification accuracy performance compared to other state-of-the-art HSI classification techniques. The NRS classifier was initially presented in a journal article [66] which is still under review.

The remainder of this dissertation is organized as follows. A brief background on source coding for images and video as well as compressed sensing is given in Chapter 2. In Chapter 3, we cover common methods of acquiring CS measurements for images and video as well as practical methods for their recovery. After this, in Chapter 4, we propose a method whereby a prediction of a given frame is made using block-based CS measurements of the frame in addition to previously recovered reference frames. The video-recovery performance results are also given in this chapter for the proposed methods. In Chapter 5, we cover the NRS classifier and its applications to supervised HSI-classification tasks. Chapter 6 covers the conclusions of our findings in this dissertation as well as suggesting an outline for future work in the area CS video recovery.

CHAPTER 2

BACKGROUND

2.1 Natural-Image and Video Coding

In traditional source coding, a signal is known by an encoder, which then chooses the most effective strategy for representing the given signal in a compressible manner in order to decrease the description length of the signal. This has the effect of reducing the dimensionality of a given signal by only storing or transmitting the most important pieces of information about the signal in question to a decoder.

By compressible, we mean that only a small number of values can be used to either perfectly represent a given signal, as in lossless coding, or approximately representing a signal by incurring error, as in lossy coding. We call a signal compressible if there is some method, such as a decorrelating transform like the Fourier or wavelet transforms, by which we can represent the signal with coefficients obeying a power-law decay in magnitude [5, 36].

This property of compressible signals means that most of the coefficients within the transform basis are near zero in magnitude, and can therefore be thought of as insignificant in terms of representing the signal. Often, insignificant coefficients are dropped during lossy encoding in order to decrease the number of bits needed to store or transmit a signal.

If this is done properly, the loss of these small-magnitude coefficients when the other coefficients are received and inverted back to the original signal domain by the decoder should make only a very small perceptual difference.

The manner in which the most important pieces of information are chosen or represented is highly dependent upon the type of signal being encoded. For natural images, it is well known that the blocked discrete cosine transform (DCT) and the two-dimensional discrete wavelet transform (2D-DWT) perform the task of decorrelation and energy compaction very well. JPEG, one of the most commonly used image-encoding formats, performs a block DCT followed by a quantization step and run-length encoding. JPEG2000, a more modern image-encoding standard, uses the 2D-DWT, quantization, embedded block coding with optimal truncation (EBCOT), and an arithmetic coder to provide better compression-to-quality performance than JPEG.

Traditional video coding operates in somewhat the same manner as image coding, except with the addition of another dimension to the signal, a temporal axis. Trivial implementations of video encoding simply apply an image encoder, such as JPEG, to each frame in the video sequence. However, this does not employ any kind of temporal decorrelation, effectively throwing out an entire dimension of the signal in terms of compression. In video-coding standards such as H.264, temporal decorrelation is implemented by creating predictions of frames in the sequence from previously encoded frames. By subtracting a prediction from each inter-predicted frame in the sequence, a residual frame is left behind which is much more compressible than the original frame. The effectively removes redundancy across the temporal axis.

In order to create frame predictions, the content of a reference frame is rearranged to represent a target frame. The most common method for doing this is through motion estimation and motion compensation (MEMC)). In MEMC prediction, the target frame is split into blocks. A sliding-block search is then conducted on the reference frame to find a block which best matches the target block according to some distance metric such as mean squared error (MSE) or mean absolute difference (MAD). A set of displacement, or motion, vectors are then transmitted along with the encoded residual frame in order to describe how to permute the decoded reference frame to recreate the prediction at the decoder side so that it can be added back to the decoded residual frame.

It is our intention to use knowledge of traditional video coding in order to better inform the recovery of video signals which have been acquired using a CS device. In the next section, we will overview CS theory.

2.2 Compressed Sensing (CS)

In essence, CS combines signal acquisition and compression by measuring the linear projection of a given signal, $x \in \mathbb{R}^N$, using some projection operator, $\Phi \in \mathbb{R}^{M \times N}$, where $M \ll N$; i.e.,

$$y = \Phi x, \tag{2.1}$$

where $y \in \mathbb{R}^M$ [1]. The "compression" in CS is the dimensionality reduction of the original signal via linear projection with Φ. We measure the degree of dimensionality reduction via the subsampling rate, or *subrate*, $S = M/N$. In this framework, x remains unknown to the

[1] While CS can indeed be adapted to operate on complex valued or continuous-time signals, our work is limited in scope to discrete-time signals consisting of real-valued entries

receiver, and so the task of the receiver is then to recover an approximation of this original signal from y.

From [14, 17], we know that, if x is sufficiently sparse, or compressible [49], in some transform basis Ψ, then x is recoverable from y by via Basis Pursuit (BP),

$$\hat{x} = \arg\min_{\theta} \quad \|\theta\|_1 \quad \text{s.t.} \quad y = A\theta, \tag{2.2}$$

where $A = \Phi\Psi^{-1}$. The BP is an example of a linear program (LP), and as such, is solvable using well known approaches such as interior-point and simplex methods, or even iteratively reweighted least squares [28]. The exact recovery of x is conditioned upon M sufficiently large and A obeying the *Restricted Isometry Property* (RIP) for small δ_s. According to [15], the RIP is defined as:

For each integer $s = 1, 2, \ldots$, define the isometry constant δ_s of a matrix Φ as the smallest number such that

$$(1 - \delta_s)\,||x||_2^2 \leq ||Ax||_2^2 \leq (1 + \delta_s)\,||x||_2^2 \tag{2.3}$$

holds for all s-sparse vectors. A vector is said to be s-sparse if it has at most s nonzero entries.

The magnitude of δ_s for which the RIP holds denotes the degree to which the projected space of A preserves distance relationships from the original high-dimensional space. It was shown in [16] that if A obeys

$$\delta_s + \delta_{2s} + \delta_{3s} < 1, \tag{2.4}$$

then any s-sparse vector can be recovered from its projection by A by solving (2.2) for M sufficiently large.

Since the sensor will only observe the projection of x by Φ and is, in general, agnostic to any model prior for x outside of sparsity, it is important choose Φ which minimizes the δ_s of A for the widest range of possible Ψ. In [11], random matrices with i.i.d entries, such as those drawn from $\mathcal{N}(0, \frac{1}{N})$, or from $\pm\frac{1}{\sqrt{N}}$ uniformly, are shown to serve well in this capacity. Random matrices maximize the *mutual incoherence* between Φ and any structured Ψ [17]. Mutual incoherence between any two matrices is defined as

$$\mu\left(\Phi, \Psi\right) = \sqrt{N} \max_{1 \leq k,j \leq N} \quad |\left\langle \phi_k, \psi_j \right\rangle| \quad \in \left[1, \sqrt{N}\right]. \tag{2.5}$$

Using the mutual incoherence, we can also determine how many measurements are required to guarantee recovery for the s-sparse approximation of x in the sparse-representation domain of Ψ. Specifically, if

$$M \geq C \cdot \mu^2\left(\Phi, \Psi\right) \cdot s \cdot \log(N), \tag{2.6}$$

where C is some positive constant, then the exact recovery is guaranteed with overwhelming probability using (2.2) [17].

However, in practical applications where the nature of x is not exactly known *a priori*, there are a few barriers which impede exact signal recovery [10]. First, most natural signals are not truly sparse in any transform basis Ψ. Second, noise may be present during acquisition, corrupting the measurements. Because of these issues, the equality constraint of (2.2) is commonly relaxed to a bound on the residual, giving some slack to the minimization while enforcing consistency between the minimizer and the measurements,

$$\hat{x} = \arg\min_{\theta} \quad \|\theta\|_1 \quad \text{s.t.} \quad \|y - A\theta\|_2 \leq \epsilon, \tag{2.7}$$

thus rephrasing the recovery problem as a quadratically constrained linear program (QCLP). In practice, an unconstrained version of (2.7), known as basis pursuit denoising (BPDN), is commonly used [23],

$$\hat{x} = \arg\min_{\theta} \quad \frac{1}{2} \|y - A\theta\|_2^2 + \lambda \|\theta\|_1 . \tag{2.8}$$

Many different approaches to solving (2.8) have been proposed, such as gradient projection [40], iterative thresholding [27], Bregman iterations [121], and homotopy continuation [71] to name a few. The BPDN is closely related to another approach to sparse regularization, the least absolute shrinkage and selection operator (LASSO) [102],

$$\hat{x} = \arg\min_{\theta} \quad ||y - A\theta||_2^2 \quad \text{s.t.} \quad ||\theta||_1 \leq K, \tag{2.9}$$

which maybe be solved through least-angle regression (LARS) [39], or through other, more general, solutions to convex optimization.

Another approach to CS recovery focuses on the sparse "mother" problem, the ℓ_0-regularized minimization,

$$\hat{x} = \arg\min_{\theta} \quad \|\theta\|_0 \quad \text{s.t.} \quad y = A\theta, \tag{2.10}$$

and, specifically, its related relaxed forms,

$$\hat{x} = \arg\min_{\theta} \quad ||y - A\theta||_2^2 \quad \text{s.t.} \quad ||\theta||_0 \leq K, \tag{2.11}$$

and

$$\hat{x} = \arg\min_{\theta} \quad ||y - A\theta||_2^2 + \lambda||\theta||_0. \tag{2.12}$$

While the ℓ_0 semi-norm is optimal for enforcing sparse signal support, its non–differentiability means that solving (2.10)-(2.12) is a combinatorially hard problem. For

this reason, [16] originally proposed the ℓ_1 convex relaxation of (2.10) to BP. This relaxation comes with the penalty of stricter requirements on the size of δ_s [15]. However, (2.11)-(2.12) may be solved in a greedy fashion using orthogonal matching pursuit (OMP) [108], and its related family of methods including stagewise OMP (StOMP) [33], regularized OMP (ROMP) [79], sparsity adaptive matching pursuit (SAMP) [31], and compressive sampling matching pursuit (CoSaMP) [80]. Additionally, it was shown in [8] that certain forms of iterative hard thresholding (IHT) may also solve (2.11)-(2.12).

Central to all of these recovery methods is a tradeoff between recovery time, computational complexity, and the quality of the recovery for a given subrate S. All of these different approaches to recovering x from y, however, require a significant amount of computation for N large. The significant resources required for CS reconstruction has been a significant impediment to the widespread implementation of CS sampling for high-dimensional signals such as natural images and video. Next, we look at the CS framework and how it can be modified to better suit such signals.

2.3 Compressed Sensing of Images

2.3.1 Acquisition

From the discussion of the framework given in the previous section alone, it is not readily apparent how a CS imaging device might be constructed. While it would be possible to sample an entire image with a high-resolution grid of CMOS sensors and then, in post-processing, convert the image into a series of CS measurements on the sampling device, this approach would offer no advantage when compared to the traditional image source-

coding techniques already employed in conventional imaging systems. Instead, correctly implemented CS acquisition can simultaneously sample and "compress" an image [2]. Since each measurement taken is a linear projection of the entire image, a single scalar value, it is possible to build an imaging sensor of arbitrary resolution which utilizes only a single sensor rather than a dense grid of senors. To achieve this, the linear projection of the image, or field of light intensities, against a given measurement vector drawn from Φ must be calculated optically *prior to sampling*.

Devising systems to achieve this linear projection optically is a non-trivial undertaking. However, one such device was famously constructed at Rice University, the single-pixel camera (SPC) [35, 100, 113]. The SPC is a CS imaging device which utilizes a digital micro mirror device (DMD) to achieve an spatial coding of the field of light intensities. DMDs are a specialized set of devices first used for digital projection systems, consisting of an array of very small mirrors which can pivot in one of two directions. In the SPC, the DMD is used to direct light onto, or away from, a lens which focuses onto a single photosensor. In this manner, the DMD can be used to effectuate a linear projection of the image against the rows of Φ with binary entries. For each measurement taken, the DMD is adjusted to match a different row of Φ. After all the measurements are taken, the imaged scene may be recovered at a spatial resolution which matches the resolution of the DMD used, one mirror corresponding to one pixel in the recovered image. A diagram of the SPC imaging system in given in Fig. 2.1.

[2]Note here that when we say "compress," we refer to the dimensionality-reducing characteristic of measuring the linear projection of a signal. This is in contrast with what is usually meant by compression in the source-coding community, which includes the quantization and entropy coding of the image.

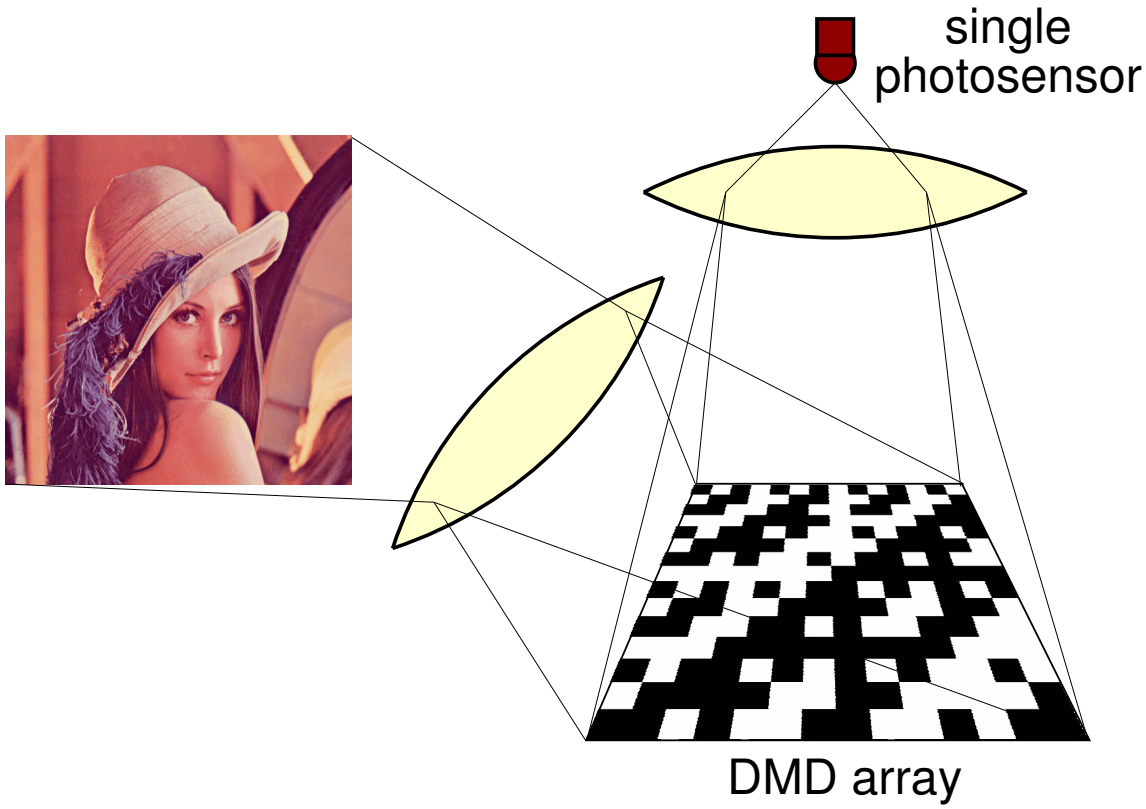


Figure 2.1

Diagram of the SPC CS imaging system (from [35, 100, 113]).

While the SPC proved that CS imaging is indeed possible, it does not necessarily show that such imaging is practical for imaging within the spectra of visible wavelengths. Visible imaging systems have been continually refined for academic, industrial, and consumer settings for decades. Even very high-resolution CMOS sensors are in no way cost prohibitive for most applications. However, in applications which require exotic sensor materials or sensors too large to achieve high-resolution imaging in a cost-effective manner, CS can offer significant cost reductions in sensor design by requiring only a single, well engineered sensor. Imaging systems operating within the infrared or other extra-visible spectra stand to benefit the most from CS imaging [6]. The techniques developed and discussed subsequently focus on imaging within the visible spectra, as this data is the most readily available for testing. However, all of these approaches can be used on extra-visible spectra with little to no modification.

2.3.2 Recovery

Proposed methods of image recovery from CS measurements attempt to address two major points: reducing the required computation time for this class of high-dimensionality signals, and improving quality of the recovered signal from a given number of measurements. Both of these points can be addressed by making the best use of signal priors and natural-image models.

The most direct approach to CS image recovery is simply to solve one of the CS recovery problems covered previously in Sec. 2.2 using a transform basis, Ψ, which best suits natural images in terms of compressible representation via energy compaction. For

example, the discrete cosine transform (DCT) and discrete wavelet transform (DWT) are two very well developed methods for compact image representation. Since images are 2D signals, and the equations of Sec. 2.2 are only defined for the recovery of vectors, a rasteriszation approach can be used to enforce agreement between transforms meant for images and CS recovery meant for vectors. For ease and clarity, in the remainder of this work we assume that any such rasterizations are implicitly defined within Ψ.

Except for sparsity, the forms of recovery in Sec. 2.2 are largely agnostic to the inherent structure of the signal. Consequently, such techniques treat the support of the signal democratically. However, the information content of the transform coefficients for natural images is not uniformly distributed over the transform basis, instead, coefficients of significant magnitude tend to cluster in the low-pass regions of the transform basis. This low-pass, or piecewise-smooth, prior can be utilized to construct CS reconstruction methods specific to natural images which require far fewer measurements for equal recovery performance when compared to their model-agnostic counterparts.

One of the earliest proposed approaches, Total Variation (TV) minimization [90, 18, 19], uses the piecewise-smooth characteristic of natural images to great effect. Instead of finding the sparsest solution within a transform domain, such as the DCT or DWT, TV minimization finds the "smoothest" solution within the space of possible solutions. Anisotropic TV-minimization makes use of the ℓ_1 norm to enforce sparsity upon the gradient of the solution, creating a penalty function of the form

$$TV(x) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}| \,. \tag{2.13}$$

Using the penalty above, the CS recovery problem can be stated as

$$\hat{x} = \arg\min_{x} ||y - \Phi x||_2 + \lambda \, TV(x). \tag{2.14}$$

TV-minimization has been widely used in CS recovery, and still represents the state-of-the-art in terms of CS image recovery [42, 76] for images sampled within the spatial domain. However, to date, many of the methods used to solve (2.14), such as second-order cone program formulations using interior-point [48] or log-barrier methods [13, 11], leave much to be desired in terms of computation time. Indeed, the cost of reconstruction using such approaches has prevented the use of TV-minimization for CS reconstruction when the image resolution is high (in excess of 512×512, for example). Other, more computationally efficient, approaches to solving (2.14) have been proposed, such as iterative soft thresholding [7] and alternating minimization [116].

In [63] an augmented Lagrangian formulation coupled with an alternating direction algorithm (TV-AL3) was proposed for solving (2.14) for both its anisotropic and isotropic forms. The TV-AL3 method retains the same reconstruction accuracy afforded by TV-minimization for CS image recovery while decreasing the computation time by orders of magnitude over other TV-minimization techniques.

Another method for gradient minimization includes the piecewise autoregression approach of model-based adaptive recovery for compressive sensing (MARX) [120]. As an extension of TV minimization, this technique makes use of smoothness priors for certain regions, while also allowing for the preservation of edge or texture content which matches a given prior model.

Other techniques seek to impose a tree-structured prior on the magnitude of wavelet coefficients during image recovery. This prior comes from the parent-child relationship of coefficients across different scales of DWT decomposition [96], coefficients at a given scale are likely to have large magnitude if the coefficient covering the same spatial location at a coarser scale has large magnitude. Tree-structured wavelet compressed sensing (TSW-CS) [50, 51] is one such approach which utilizes a hierarchical Bayesian model to enforce predicted wavelet coefficient relationships during image recovery. Another similar technique uses Gaussian scale mixtures to similar effect [59].

While many different techniques for CS image recovery have been proposed, few address the practical concerns associated with CS image acquisition and recovery, namely, tractable recovery time for high-resolution images and feasible sensor implementations. We address these points in the next section.

2.3.3 Block-based CS

The CS image recovery techniques covered in the previous section are oriented towards a dense, global sampling of the entire image during acquisition. However, this approach, while being desirable in theory, has a number of distinct disadvantages in practice.

First, the sensing device must store the set of projections Φ in order to orient its spatial scattering element, such as an array of DMD mirrors. Each measurement requires a linear projection with a vector of dimensionality equal to the desired resolution of the final image. Since the number of measurements required also scales with the resolution of the desired image, the total number of elements within Φ can become too large to reasonably store for

strict sensor-hardware constraints. For example, for a desired image resolution of 512×512 pixels, and at a subrate of 0.2, Φ contains $512^2 * (0.2 * 512^2) \approx 1.37 \times 10^{10}$ elements which must be quantized to a certain number of bits. In the best-case scenario, when each element is a binary value requiring one bit to store, Φ would require ≈ 1.68 gigabytes to store explicitly. Second, the majority of CS reconstruction techniques have a computational complexity which scales polynomially with the dimensionality of the problem, causing the reconstruction of high-resolution images to become impractical for most cases.

Gan [45] suggests that, in the case of natural images, the computational complexity of CS reconstruction and the memory burden for the sensing device can be assuaged by breaking up the signal into distinct blocks during acquisition. Block-based CS (BCS) removes the global sampling of x by a dense Φ and replaces it with a repeating block-diagonal measurement matrix by which local sampling of x within distinct blocks of size $B \times B$ is accomplished. When the same Φ_B is used for every block, Φ takes on a block-diagonal form,

$$\Phi = \begin{bmatrix} \Phi_B & 0 & \cdots & 0 \\ 0 & \Phi_B & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \Phi_B \end{bmatrix}, \tag{2.15}$$

such that (2.1) can be effectuated in a block-by-block fashion; i.e.,

$$y_i = \Phi_B x_i, \tag{2.16}$$

where x_i is block i of the image. The size of Φ_B is $M_B \times B^2$ such that the subrate of BCS is $S = M_B/B^2$.

The blocked sampling of the image effectively breaks up a large ℓ_1-minimization problem into many sub-problems, which can be solved more quickly by virtue of their smaller dimensionality, and in parallel. It has been shown in [115] that localizing measure in this way does not significantly degrade the achievable performance of the reconstruction.

For recovery, [45] suggests a procedure that couples projected Landweber iterations similar to IHT with smoothing in the form of Wiener filtering. This smoothed projected Landweber (SPL) procedure thus combines a fast, iterative solution to (2.7) with the imposition of a smoothness constraint designed to eliminate blocking artifacts. In [76], the overall process of BCS sampling and SPL reconstruction was called BCS-SPL. BCS-SPL was extended [45] by use of bivariate shrinkage for thresholding and directional transforms such as a dual-tree discrete wavelet transform (DDWT) and a contourlet transform (CT). These modifications provide significant recovery-quality improvement while maintaining a reasonable reconstruction time. The results in [76] suggest that BCS-SPL augmented with such directional transforms is competitive with the state of the art for CS recovery of a single still image.

Additionally, a variation of BCS-SPL was implemented within the wavelet domain, multiscale BCS-SPL (MS-BCS-SPL) [41], which assumes a wavelet-domain sensing device. The MS-BCS-SPL approach samples wavelet coefficients in independent blocks, similar to spatial domain BCS-SPL. However, the block sizes in MS-BCS-SPL scale depending on the scale of the wavelet coefficients being sampled. Additionally, the number

of samples taken for each block in a given scale varies, concentrating samples to low-pass bands in order to preserve the most significant information about the image.

Structured random matrices (SRMs) [46, 32] are another approach to reducing both the memory and computational burden of CS for images. A SRM-based acquisition does not require the explicit storage of a projection matrix, Φ. Instead, SRMs induce a random sampling by randomly permuting the elements of a simple image transform such as the block cosine or Hadamard transform, and then taking a random subsampling of the resulting transform. Also, SRM sampling speeds up recovery, as its forward and back projection can be quickly calculated, requiring fewer costly matrix to matrix operations.

For the remainder of this work, we focus on the BCS paradigm, and specifically, the BCS-SPL method of image recovery. While the SRM approach offers similar advantages in terms of memory and computational requirements, we will see in the next section how spatial block sampling specifically can be leveraged to facilitate CS video sampling and recovery. For more discussion and results on the performance of different CS image recovery techniques, we refer the reader to [42].

CHAPTER 3

COMPRESSED SENSING OF VIDEO

CS promises to be a viable paradigm for signal acquisition for many different forms of signals, such as one-dimensional waveforms in seismology, two-dimensional natural imaging, three-dimensional magnetic resonance imaging (MRI) and tomography, as well as many-dimensional hyperspectral data sets such as those from multi and hyperspectral imagers. These applications have fit well in the CS framework because of the many data-compression techniques unique to each which allow for their compact descriptions and representations needed for CS recovery. All of these applications also possess signal structures which allow for practical CS hardware implementations for the acquisition of CS measurements.

Video, too, can potentially be another viable area for the application of CS. However, the very large dimensionality of video signals as well as their time-varying nature pose barriers to practical CS acquisition and recovery that must be addressed. In the following sections, we will review these limitations as well as propose a variety of methods for the recovery of CS-acquired video.

3.1 CS Video Acquisition

CS acquisition of video would ideally be global in the sense that CS measurements would span the entire spatial and temporal extent of a video sequence; however, such global CS acquisition of video is largely considered impractical to implement in a real device [34]. As a consequence, we focus on the case in which each video frame is acquired independently with still-image-based CS measurement, for instance, with successive applications of a single-pixel camera as was done in [114]. While other approaches to acquisition might eventually be possible, we consider the single-pixel camera to be a straightforward and realizable framework for capturing images and video via linear projection.

More specifically, CS theory dictates that it is possible to recover a signal of dimension N from a set of measurements of dimension M where $M \ll N$. In the canonical CS acquisition or measurement process,

$$\mathbf{y} = \mathbf{\Phi x}, \tag{3.1}$$

we see that these measurements are calculated as a projection of the entire N-dimensional signal by an $M \times N$ random projection matrix, $\mathbf{\Phi}$. Complications in designing CS hardware arise due to the global nature of this dimensionality-reduction step. That is, a CS device must be able to simultaneously view the entirety of a signal (for video, this means in space as well as in time) and calculate its projection by $\mathbf{\Phi}$ non-computationally in the ambient signal domain. Because of this requirement, simultaneous spatial and temporal measurement of video appears impractical [34], and thus one opts for frame-by-frame measurement.

3.1.1 Hardware Limitations

In the case of natural-image signals, we have already employed the single-pixel camera [100, 113, 114, 35] for static-scene measurement. However, the single-pixel camera entails multiple measurements conducted sequentially in time such that the total time to acquire a given signal is increased by a factor of M. Thus, for dynamic scenes, there is a potential for disagreement between successive measurements, as each measurement of the scene is at a different point in time. If object motion between measurements is significant, there could be blurring or other reconstruction errors when recovering the signal from the measurements. For this reason, a viable CS sensing device for dynamic scenes (i.e., video) which uses single-pixel acquisition must have a very short exposure time and delay between measurements.

More specifically, for the CS measurement of video at a target frame rate, R_f, using M measurements for each frame, each measurement must be captured within $1/(R_f M)$ seconds. For high-resolution video, M can be somewhat large (though, of course, still much less than the number of pixels, N), and this puts a tight restriction on the sensing device when the measurements are captured sequentially. However, the latest micro-mirror arrays have attained very fast switching speeds, and these speeds are increasing each year as research into microelectromechanical systems (MEMS) continues. For example, [81] reports a MEMS device capable of switching mirror states in 222 ns; such a DMD would permit CS measurement of video with 720×480 frames at a subrate of $S = 0.3$ with a frame rate of

$$R_f = \frac{1}{720 \cdot 480 \cdot 0.3 \cdot 225 \times 10^{-9}} \approx 42 \text{ frames/sec.} \tag{3.2}$$

While devices with such switching speeds are not yet in commercial production, even devices with much lower switching rates (e.g., on the order of 0.1–1 MHz) can accomplish video acquisition by trading-off spatial resolution, frame rate, or reconstruction quality. For example, at a switching rate of 0.5 MHz and a subrate $S = 0.2$, a video sequence could be captured at standard CIF resolution (352×288) at $R_f \approx 24$ frames/sec.

If the measurement matrix $\mathbf{\Phi}$ has real-valued, rather than binary, entries (as is the case with a dense Gaussian measurement operator), it is possible to use pulse-width modulation or dithering to simulate fractional transmittance from the mirror to the sensor. However, such approaches are problematic. Firstly, the fractional values are subject to quantization error induced by the accuracy of the pulse-width modulation. Secondly, for a fixed measurement subrate, the mirror-switching speed must be increased by a factor dependent on the quantization precision used.

To reduce some of the necessary tradeoffs caused by sequential measurement, multiple sensors operating in parallel may be used to increase the effective measurement subrate. A multiple-pixel device would operate in much the same manner as the single-pixel camera and would still maintain a low sensor density as compared to a full-resolution sensor.

3.1.2 Physical Limitations

The exposure time necessary to accurately measure the amount of radiation incident on the photosensor is another—and perhaps more significant—component to the time required for each measurement. Exposure time presents a challenge: if the exposure time is too short, then noise from dark current within the system or Poisson noise induced by the pho-

ton arrival rate can overwhelm the actual measurements, requiring more sophisticated—and costly—photosensors. However, the increased cost of such photosensors is offset by the fact that there would be need for only one photosensor rather than an entire array. Also, using the single-pixel framework, photons from all the mirrors that face the sensor are concentrated onto a single sensor during measurement. For example, consider BCS using blocks of size 64×64. If we assume that, on average, half the mirrors point toward the sensor for any given measurement, the sensor is exposed to $64^2/2 \approx 2000$ times more light than a single sensing element would be in a traditional dense-array sensor. This focusing of energy increases the signal-to-noise performance of the single sensor as compared to the limited spatial binning capabilities of sensors in a dense-grid configuration.

These considerations make high spatial- and time-resolution video difficult but arguably not impossible. In a general sense, the design of CS acquisition devices would necessitate some tradeoff between the number of measurements acquired for each frame, the desired frame rate (temporal resolution), and the exposure time for each measurement. For example, applications such as distributed video networks, or other *ad hoc* distributed sensor networks, could make use of cheaper CS video-sensor systems in surveillance capacities wherein spatial resolution is not the top priority. With smaller-resolution frames, the number of measurements for each frame decreases for a given target subrate, thereby allowing greater exposure time for each measurement within a frame-rate constraint.

Many of these design constraints might be justified in settings wherein every measurement is costly or the sensors themselves are costly. Sensing signals in exotic spectra—such as in thermal, terahertz, and medical imaging—represent areas wherein CS can potentially

reduce either device or acquisition cost. Except for perhaps niche applications, imaging within the visible domain using CS is not likely to be competitive with existing low-cost CCD or CMOS imagers. However, in the remainder of our discussion, we use visible-domain imagery in our experiments to explore potential recovery techniques for signals. Thermal, infrared, and medical images exhibit characteristics similar to natural, visible domain, images—most importantly, piecewise smoothness. Because of these similarities, we anticipate that the methods we demonstrate could also be applied to non-visible spectra with similar effect.

3.1.3 Block-based Acquisition of Video

So far, this discussion has considered only the case of a globalized, and therefore dense, structure of Φ. From prior discussion, we know there are some inherent drawbacks to such a dense measurement process, such as reconstruction time and the memory requirements of storing Φ. However, the hardware of the single-pixel camera can accommodate a CS measurement procedure, and the practical considerations discussed here still apply in the BCS context. BCS also has the added advantage of decreasing the bandwidth required to transmit measurement vectors between system memory and the DMD array since a small measurement vector can be transmitted once and subsequently translated across the DMD array for each block.

For static images, BCS measurement is straightforward, requiring only a block-diagonal Φ be employed. In the case of BCS of video, however, since each block is measured independently, each block represents a different point in the time of the scene, rather than each

frame. On the one hand, because fewer measurements are required for a single block than for an entire image, the time duration between the first measurement and the last measurement in a given block is a multiplicative factor less than if we take global measurements of the entire frame. This decreases the possibility for blurring within a given block in a frame. On the other hand, since each block represents a different point in time, there could be some content drift between the blocks of a given frame if the dynamic content being represented is changing sufficiently fast. If it is more desirable to have blurring rather than drift, then the CS device could scan through the blocks repeatedly, taking a single measurement at a time. This would simulate the measurement timing of a global CS measurement of the frame and could be accomplished by simply reordering the rows of the block-diagonal $\mathbf{\Phi}$.

3.2 CS Video Recovery

In this section, we detail a variety of recovery procedures that could be employed to recover a frame-by-frame CS-acquired video sequence. In order to have a successful recovery technique, it is imperative that a method be found to adequately represent the video sequence in a compressible fashion through some form of multidimensional decorrelation and energy compaction.

3.2.1 Frame by Frame

The first approach we investigate for the recovery of CS acquired video is the naïve one. The frame-by-frame recovery procedure serves as our baseline CS video reconstruction technique. It does not incorporate any information about the time-varying nature of the video signal into the compressible representation of the video signal. Instead, each frame

is treated as an independent image and recovered using only a spatial, two-dimensional, sparse-representation basis. Assuming that the recovery is given by a set of measurements, y_t, and the projection used, Φ_t, for each frame in the sequence, the recovery is simply

$$\hat{x}_t = \mathrm{CSRecover}_{\Psi}\left(y_t, \Phi_t\right). \tag{3.3}$$

Here, $\mathrm{CSRecover}_{\Psi}$ is any CS image-recovery technique, such as GPSR or BCS-SPL, which employs the basis Ψ as the sparse representation basis for the frame recovery. Ψ could be any two-dimensional transform, including the DCT, DWT, DDWT, CT, or RDWT.

Because this method does not take into account any temporal redundancy or motion model, we could say that, in the traditional sense, each frame is reconstructed in an Intra-mode. The distortion performance of the recovery of each frame, $||\hat{x}_t - x_t||_2$, is dependent upon only its own information content and the number of measurements taken. As the number of measurements decreases for a given level of information content in the signal, the recovery performance will decrease. In the case of CS-acquired video, because there are many frames which must be sampled, the number of measurements allocated to each individual frame must be fairly small. The lack of an adequate number of measurements can cause severe distortions at each frame, lowering the overall quality of the video recovery significantly.

We anticipate that any method which takes into account the highly correlated nature of frames within any natural video sequence in a meaningful way will see a significant performance improvement when compared to frame-by-frame recovery. The problem from

here is to find the best way to incorporate this prior knowledge of the structure of video sequences into the recovery procedure.

3.2.2 Volumetric Recovery

One of the most direct ways to employ decorrelation along all three axes (two spatial and one temporal) of a video sequence is through a three-dimensional transform such as the three-dimensional DWT (3D-DWT). The 3D-DWT is simply a multidimensional extension of the DWT, much like the 2D-DWT, where the wavelet filters are applied along each dimension of the signal.

With this technique, a group of frames, or the entire video sequence, is reconstructed simultaneously by a CS recovery algorithm while employing the 3D-DWT as the sparse representation basis, Ψ. This methodology can be easily incorporated into iterative CS solvers such as the IHT/IST or even BCS-SPL. The thresholding step used in these solvers can be accomplished in the three-dimensional case in much the same way as the two-dimensional case, since the parent-child wavelet-coefficient-magnitude correlation observations used in techniques such as bivariate shrinkage [25] still apply in the multidimensional-wavelet case.

However, the volumetric approach to video recovery does not provide adequate distortion performance in many circumstances, especially in video sequences with a large amount of complex motion, fast transitions, strobing patterns, or panning and zooming. We can explain this intuitively by remembering that each frame in a video sequence represents a sample taken at a discrete point in time, separated from its temporally neighboring

frames by gap of time in which no measurements are taken. In the physical scene being captured, all motion and change is continuous in nature; however, video sequences acquired at moderate frame rates (30-40 frames per second) do a poor job of capturing this continuity and only represent a discretized version of the motion and change playing out in the scene. This discretization in time can cause sharp transitions in a pixel-luminosity level across time. These discontinuities require large-magnitude wavelet coefficients in high-pass bands, degrading the overall performance of the temporal decorrelation and compressible representation.

As an example, the first 32 frames of the common video test sequence *Foreman* were sampled using subrates $S = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ using BCS, as in Eq. 2.16, and each frame was sampled according to Sec. 3.1. The sequence was recovered using three levels of 3D-DWT decomposition for each of the different subrates used. The performance of this recovery was then compared against the baseline, frame-by-frame, recovery procedure discussed in Section 3.2.1.

In Figure 3.1, the performance of CS recovery appears to perform worse when using the 3D-DWT than if we had just recovered each frame as an independent image. As discussed in this section, we first assume that the 3D-DWT should provide added decorrelation along the temporal axis and therefore aid the CS recovery. This would be true in the case of continuous, or very slow, motion or change in our scene. The foreground of the *Foreman* sequence is very dynamic in nature, and the 3D-DWT does more harm than good, here. The temporal discontinuities between frames prohibit the use of direct volumetric recoveries except perhaps when the video signal is sampled with $S > 0.5$, allowing for more

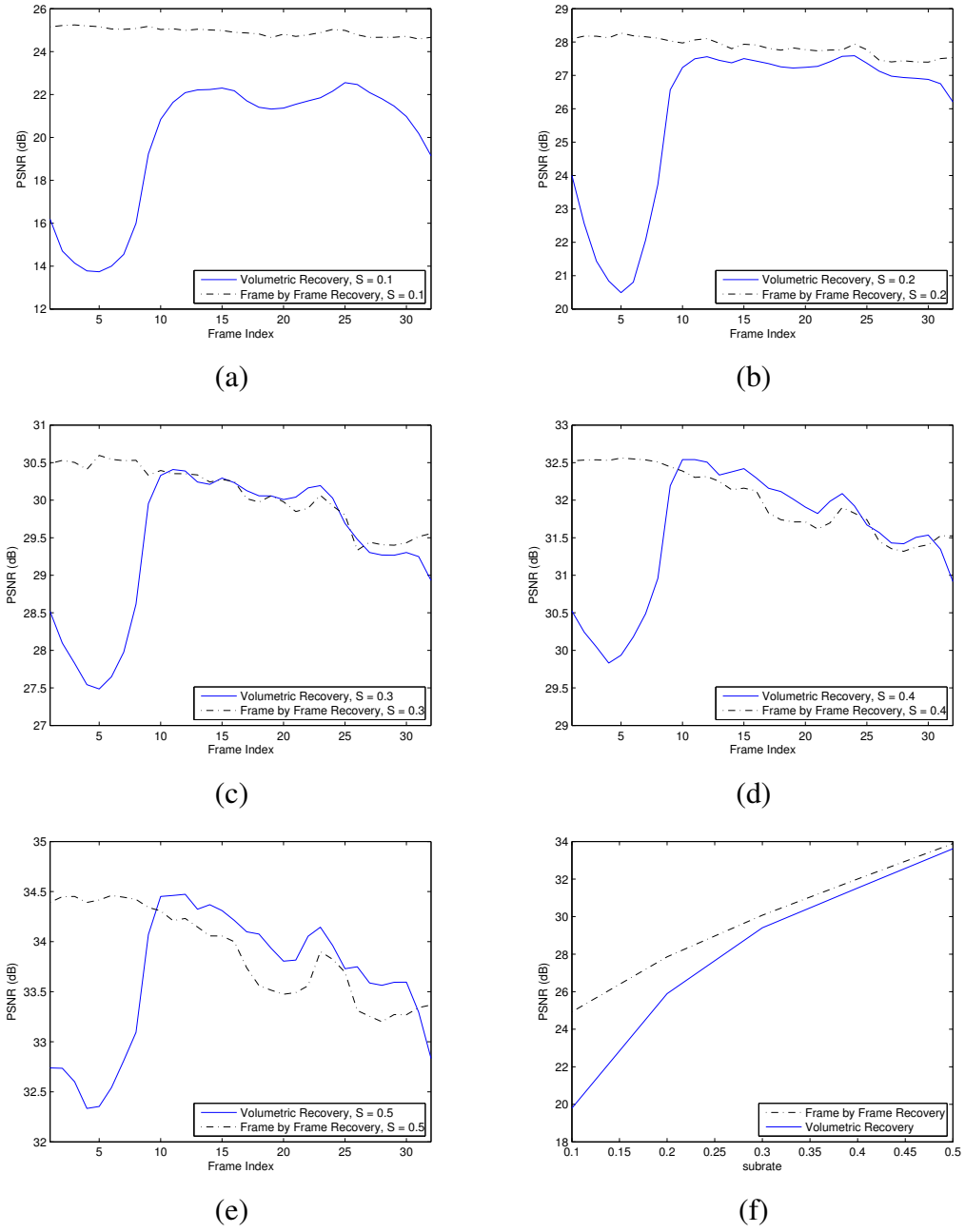


Figure 3.1

Frame-by-frame and 3D-DWT recovery quality of the first 32 frames of the *Foreman* sequence.

(a) S = 0.1, (b) S = 0.2, (c) S = 0.3, (d) S = 0.4, (e) S = 0.5. (f) Recovery quality averaged over entire sequence for S = [0.1,0.5].

accurate recovery even in the presence of low compressibility. A high subrate in this range is undesirable, however, and so we must try to find some other method if we wish to recover video sequences with better distortion performance than the frame-by-frame approach.

3.2.3 Residual Recovery

Residual reconstruction seeks a sparser representation of a given signal by recovering the difference between the signal and some prediction. The philosophy is very similar to that of DPCM in traditional signal coding—if a prediction is similar to the signal it is intended to approximate, then the value of the residual over most of the support should be insignificant in magnitude. In traditional video coding, this technique is used extensively to create highly compressible residual frames which are then compressed with a still-image coder.

Residual reconstruction can be easily integrated into the CS paradigm because it requires no change on the part of the signal acquisition and has a simple implementation on the reconstruction side. Suppose that we sample a given signal x using a measurement basis Φ such that measurements y are calculated via (2.1). If we are given some kind of prediction of x in the ambient domain of x—namely, $\tilde{x}$, for which we hope $\tilde{x} \approx x$—then we can find the residual r between the two signals as $r = x - \tilde{x}$. Because y is acquired simply by taking the inner products of x with the rows of Φ, the projection of r into the measurement basis is

$$q = \Phi r = \Phi\left(x - \tilde{x}\right) = y - \Phi\tilde{x}. \tag{3.4}$$

Because of the linear nature of the signal-sampling process, a simple subtraction of a projection of $\tilde{x}$ provides us with a projected residual signal at the reconstruction side without changing our signal-acquisition procedure. This residual should be more amenable to CS recovery because it is expected to be much more compressible than x itself. We may then calculate the final reconstruction of y,

$$\hat{x} = \tilde{x} + \mathrm{CSRecover}_{\Psi}(q, \Phi). \tag{3.5}$$

The quality of $\hat{x}$ is directly tied to the ability of the reconstruction to recover r from q; i.e.,

$$\begin{aligned} \|x - \hat{x}\|_2 &= \|x - (\tilde{x} + r + e_r)\|_2 \\ &= \|(x - \tilde{x}) - r - e_r\|_2 \\ &= \|e_r\|_2 , \end{aligned} \tag{3.6}$$

where e_r is the error resulting from a non-exact recovery of r.

As in the previous section, an experiment was conducted to compare the performance of the residual reconstruction method to the frame-by-frame approach. The recoveries were calculated using the same measurements as the previous experiment, as well. For residual reconstruction, a prediction for each frame of the sequence is required. These predictions, P_t, were formulated using bidirectional MEMC with integer pixel accuracy,

$$P_t = 0.5 * (\mathrm{MC}(X_t, X_{t-1}, MV_t^f) + \mathrm{MC}(X_t, X_{t+1}, MV_t^b)). \tag{3.7}$$

The sets of motion vectors, MV^f and MV^b, given to the CS recovery procedure are the true motion vectors calculated from the original video sequence. We use these motion vectors for comparison purposes.

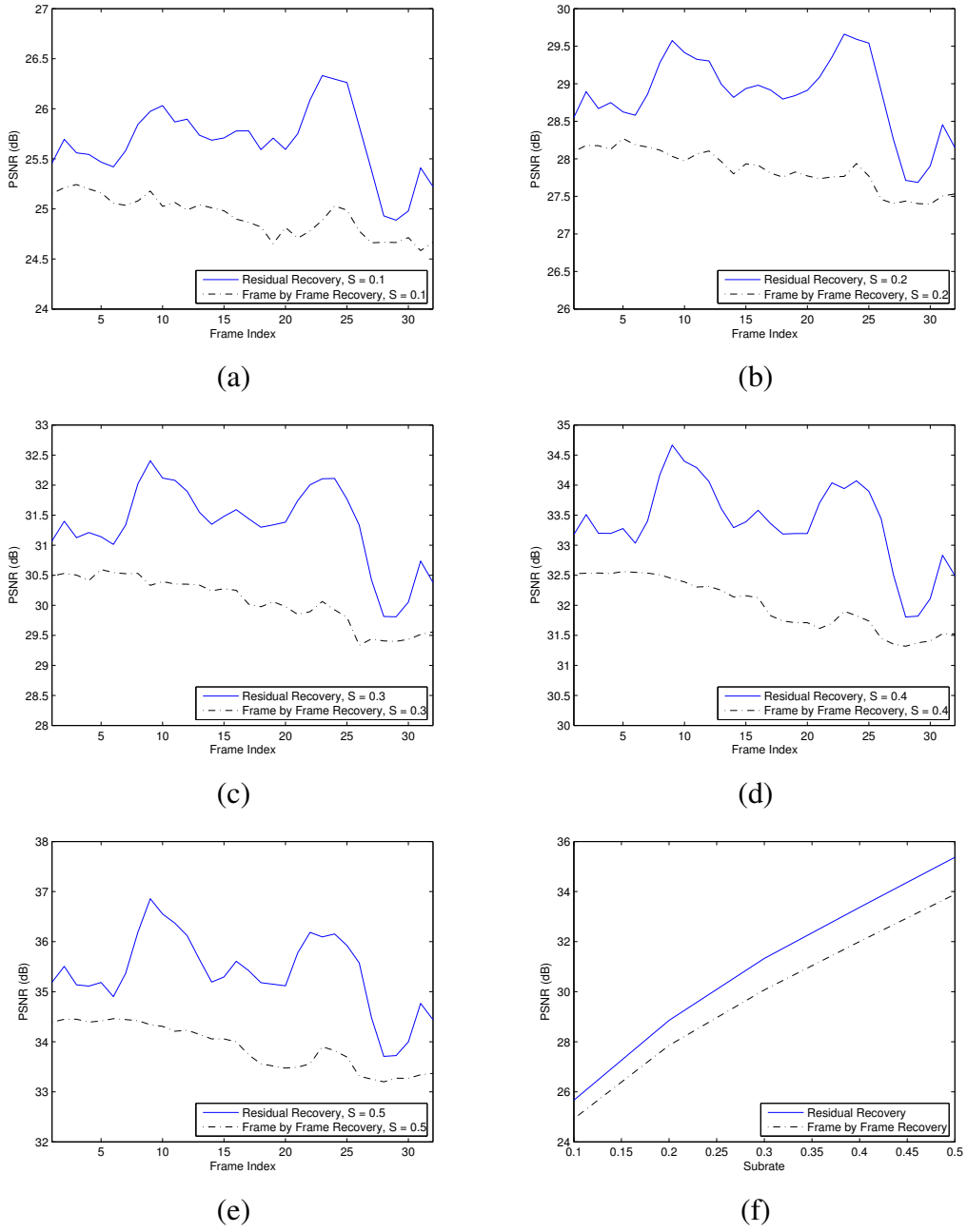


Figure 3.2

Frame-by-frame and residual recovery quality of the first 32 frames of the *Foreman* sequence.

(a) S = 0.1, (b) S = 0.2, (c) S = 0.3, (d) S = 0.4, (e) S = 0.5. (f) Recovery quality averaged over entire sequence for S = [0.1,0.5].

Figure 3.2 shows a marked increase in distortion performance when residual recovery is used instead of the frame-by-frame approach. The performance of the residual reconstruction with such a simple prediction step leads us to believe that this method shows the most promise if adequate predictions can be calculated at the time of recovery without the use of outside information other than the given measurements. In the following chapter, we will investigate different prediction strategies in order to find one that provides the best performance when used in conjunction with residual reconstruction.

3.2.4 Other Approaches to Recovery

A number of approaches to the CS reconstruction of video were developed for the particular case of dynamic magnetic resonance imagery (MRI). This type of image sequence tends to have less motion, and the motion tends to be less of a strictly translational nature, than does video acquired from natural photographic scenes. Initial work adopted the volumetric reconstruction employed originally in [114, 113]—for example, [44] reconstructs a dynamic MRI volume using a temporal Fourier transform coupled optionally with a spatial wavelet transform as a 3D sparsity basis.

Given the computational issues with reconstructing volumes, most CS reconstructions for video have focused on frame-based recovery that exploits the fact that successive frames are strongly correlated. For example, [72] reconstructs multiple frames simultaneously while capitalizing on the fact that frame-to-frame differences are expected to be highly compressible; thus the ℓ_1 norm of the frame difference is incorporated into the CS reconstruction.

Various other strategies have been adopted to handle frame-to-frame correlation. For example, Vaswani *et al*. [111, 112, 68, 88, 110] have proposed a variety of related approaches for the CS reconstruction of dynamic MRI data. Fundamental to several of these techniques [111, 68, 88] is the general strategy of residual reconstruction from a prediction of the current frame as in (3.5); the key difference from the work proposed here is that, rather than using a MEMC-based prediction, Vaswani *et al*. employ a least-squares [111] or Kalman-filtered [88] prediction. These predictions are driven by an explicit sparsity pattern for the current frame; the techniques attempt to track this sparsity pattern as it evolves from frame to frame. It is assumed that the sparsity pattern evolves slowly over time, an assumption that may not hold in general video with arbitrary object motion. However, the "Modified-CS-Residual" algorithm of [68] is a prominent benchmark in the literature for gauging CS-reconstruction performance for not only dynamic MRI but also video as well.

Alternatively, [94] proposes another strategy that also attempts to explicitly track temporal changes in video. In this case, [94] deploys a linear dynamical system (LDS) that models the evolution of a video scene in terms of low-dimensional dynamic parameters and high-dimensional static parameters such that the compressive measurement process is applied to only the dynamic portion of the signal. It is observed in [94], however, that this LDS-based strategy works well for relatively low-motion content of a largely textural nature, such as flames, water, and traffic, whereas the more complex and translational motion often associated with more arbitrary video content is not properly handled by the model.

This observation echoes the primary drawback that applies in general to the techniques for CS reconstruction of video that do not employ MEMC. Methods such as those con-

sidered above are typically best suited to video content that varies only quite slowly over time, such as dynamic MRI. For more complex temporal variation, particularly the non-stationary translational object motions that often occur in video of natural scenes, the use of explicit MEMC is warranted. In the literature, there have been a only handful of approaches that incorporate MEMC for CS reconstruction of video. We have discussed [78] and [30] previously; others in this vein include the following.

In [99], an extension of the dynamic-MRI reconstruction of [112] was proposed. In essence, rather than simply estimate updates to the time-varying sparsity pattern directly from the preceding frame, the preceding frame is first motion-compensated, allowing for more arbitrary object motions to be handled. The technique of [99] inherits, however, the drawback identified previously for [112] in that the temporal evolution of the sparsity pattern is assumed to be slow.

The method of [58] exploits temporal correlation by constructing a motion-compensated interpolation between consecutive key frames. This motion-compensated interpolation is then used as the initialization point of a still-image CS reconstruction. The key frames are reconstructed using an independent still-image CS reconstruction with a subrate higher than that used for the non-key frames.

Another reconstruction algorithm driven by MEMC between high-quality key frames was considered in [56, 57]. This algorithm, called k-t FOCUSS in [56], assumes that there exist one or two key frames obtained through some separate means, and then CS reconstruction is driven by residuals between each intervening non-key frame and a block-based bidirectional motion-compensated prediction from each of the key frames (or a single uni-

directional motion-compensated prediction in the event that only one key frame is available). As in MC-BCS-SPL, full-search block matching is used for the MEMC process. We note that k-t FOCUSS was designed specifically for dynamic MRI; consequently, [56, 57] uses relatively long distances between key frames (e.g., 25 frames for a cardiac cine sequence in [57]) with perfect key frames (i.e., subrate = 1.0) at each end.

A final strategy to incorporating explicit MEMC into CS reconstruction for video is represented by the technique proposed in [85]. In contrast to the MEMC-based reconstructions like MC-BCS-SPL as well as those of [56, 57, 30] which are inspired by the traditional hybrid video-coding architecture, the technique of [85] adopts the motion-compensated temporal filtering (MCTF) of [95]. In essence, MCTF is combined with a spatial DWT to implement a motion-compensated 3D transform. 3D reconstruction simultaneously across all frames similar to [113] is then conducted. A key aspect of the proposed approach is that the 3D reconstruction is applied in each resolution level of the spatial DWT separately, using the reconstruction of the previous spatial resolution for determining the motion vectors between all the frames to drive the MCTF within the current resolution level. As a consequence, the technique of [85] has the advantage of explicit MEMC like those methods surveyed above. However, it also inherits the computational drawback of the cross-frame volumetric reconstruction identified in Sec. 3.2.2 as a significant impediment for 3D techniques such as [114, 113]; specifically, computational issues are compounded in [85] since a separate volumetric reconstruction, as well as MEMC process, is conducted for each spatial resolution level.

CHAPTER 4

FRAME PREDICTION FOR RESIDUAL RECOVERY

The key to the successful use of residual reconstruction is to generate a compressible residual frame, r, through the use of a very accurate prediction, $\tilde{x}$, of the target frame, x. Thus, the goal is to carry out the optimization,

$$\tilde{x} = \arg \min_{\theta \in \mathcal{P}(x_{ref})} \|x - \theta\|_2 , \tag{4.1}$$

where $\mathcal{P}(x_{ref})$ is the set of all possible predictions which can be generated from the reference frame, x_{ref}, using some defined strategy. However, the creation of the prediction $\tilde{x}$ occurs during CS reconstruction; as a consequence, x is unknown, and (4.1) cannot be implemented as written. There are two strategies to approximate (4.1) using only information known to the CS reconstruction. The first would be to approximate x with an initial CS recovery from y and use the resulting approximation to x to drive the prediction process; i.e.,

$$\tilde{x} = \arg \min_{\theta \in \mathcal{P}(x_{ref})} \left\| \mathrm{CSRecover}_{\Psi}(y, \Phi) - \theta \right\|_2, \tag{4.2}$$

The resulting $\tilde{x}$ is then used in (3.5) to form the final reconstruction $\hat{x}$ using a CS reconstruction from the measurement-domain residual, $q = y - \Phi\tilde{x}$.

We propose a different approach, shifting (4.2) from the ambient-signal domain into the measurement domain of Φ,

$$\begin{aligned}\tilde{x} &= \arg \min_{\theta \in \mathcal{P}(x_{ref})} \|\Phi x - \Phi\theta\|_2 \,, \\ &= \arg \min_{\theta \in \mathcal{P}(x_{ref})} \|y - \Phi\theta\|_2 \,. \end{aligned} \tag{4.3}$$

Although (4.3) recasts the search for the prediction into the measurement domain, the Johnson-Lindenstrauss (JL) lemma [55, 26, 1] suggests that the solution of (4.3) will likely match or lie near to that of (4.1), especially in the case that $\mathcal{P}(\cdot)$ is chosen to be based upon single block-matching MEMC. In brief, the JL lemma holds that L points in $\mathbb{R}^N$ can be projected into a K-dimensional subspace and approximately maintain the original pairwise distance relationships between the points so long as $K \geq O(\log L)$. As a consequence, the $\tilde{x}$ closest to x in (4.1) should map to the $\Phi\tilde{x}$ that is closest to y in (4.3), provided that the number of candidates searched in the minimizations is not too large.

Our experimental observations reveal that the measurement-domain prediction of (4.3) provides better predictions in general than the ambient-domain strategy represented by (4.2) (see Fig. 4.1). This is due to the fact that (4.2) uses only a noisy approximation to x, whereas the JL lemma suggests that (4.3) should nearly duplicate the targeted procedure of (4.1). As a consequence, we focus on measurement-domain predictions in the form of (4.3) in the remainder of our development. We originally presented this material in [104] which we further extended in [42, 21].

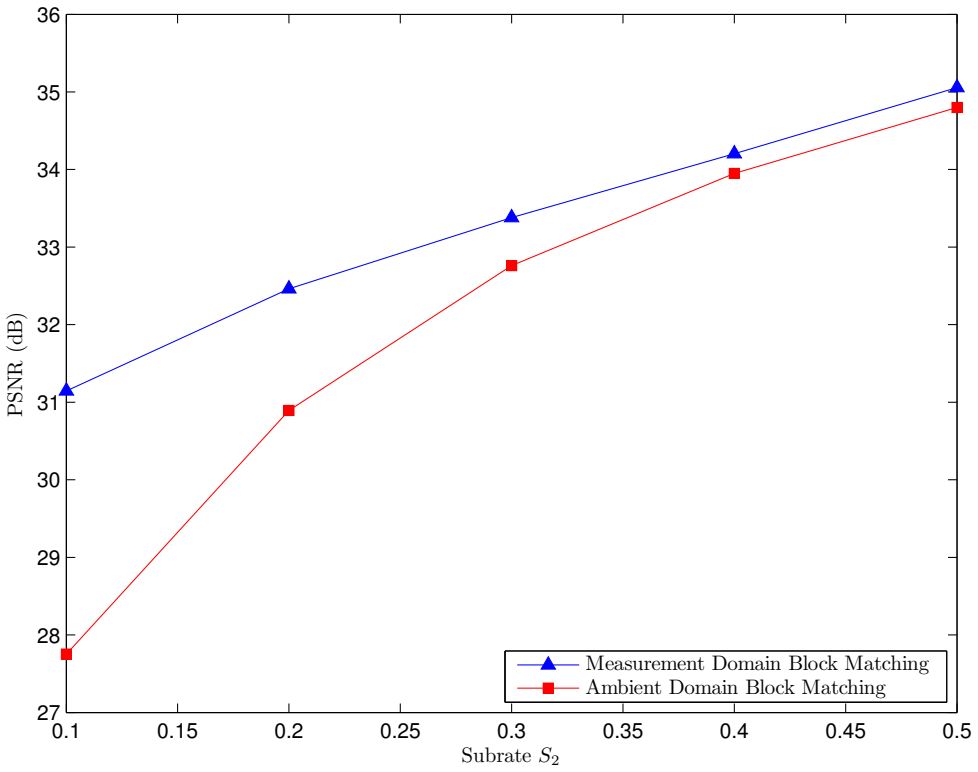


Figure 4.1

Recovery quality over subrate for the second frame of the *Foreman* sequence when using the first as a reference measured at $S_1 = 0.5$.

4.1 Single Hypothesis (SH) Frame Prediction

In traditional video coding, frame predictions are calculated from temporally neighboring frames which are likely to have similar content to the target frame using MEMC. Specifically, the frame at time t to be predicted, x_t, is split into blocks of size $B \times B$. The chosen reference frame or frames are then searched within a spatial region surrounding the location of the target block within x_t. The best-matching block, chosen according to some distortion measurement, in the reference frames then forms the prediction of the target block. This is known as SH prediction in the video-coding community since a single, best-matching hypothesis prediction (a block in one of the reference frames, in this case) is chosen to represent the target block.

In the CS reconstruction of video wherein each frame has been sampled using BCS applied frame by frame, the ensemble of measurements for frame x_t is

$$y_{t,i} = \Phi x_{t,i}, \tag{4.4}$$

where i is a block index. In order to create a prediction of a given block, $x_{t,i}$, we recast (4.3) as

$$\tilde{x}_{t,i} = \arg \min_{\theta \in \mathcal{H}_{t,i}} \|y_{t,i} - \Phi\theta\|_2 , \tag{4.5}$$

where $\mathcal{H}_{t,i}$ is the set of blocks culled from the reference frame or frames within the search space given for block $x_{t,i}$ (typically a rectangular region about the spatial location of $x_{t,i}$ in the frame).

4.2 Multi Hypothesis (MH) Frame Prediction

Video coding has long exploited MH methods to improve video-coding quality [98]; common forms include subpixel-accurate MC [47], overlapped-block MC [82, 84], bidirectional MC (B-frames), and long-term-memory MC [118]. These techniques can be viewed as tradeoffs specific to a rate-limited environment; that is, these techniques impose specific structures on the hypotheses that form the ultimate prediction in order to limit the amount of additional motion-vector rate overhead entailed by multiple predictions of a single block. However, in the context of CS reconstruction, the MH predictions are all calculated at the reconstruction side of the system, there is no associated rate burden, and we are able to consider more intensive forms of MH prediction, essentially combining all the best hypotheses available from the reference frames without the imposition of rate-limiting structure.

For MH CS recovery, we alter the MEMC prediction strategy of (4.1) such that each block prediction, $\tilde{x}_{t,i}$, is calculated as a linear combination of all available hypotheses in $\mathcal{H}_{t,i}$ while maintaing a high degree of accuracy; that is,

$$w_{t,i} = \arg\min_{w} \|x_{t,i} - H_{t,i}w\|_2 \,, \tag{4.6}$$

$$\tilde{x}_{t,i} = H_{t,i}w_{t,i}. \tag{4.7}$$

We note that the above equations are defined at each block where i serves as the block index. Also, $H_{t,i} \in \mathbb{R}^{B^2 \times K}$ consists of rasterized versions of the hypotheses as its K columns and $K = |\mathcal{H}_{t,i}|$.

Of course, in the case of CS reconstruction, (4.7), like (4.1) before it, cannot be implemented—we cannot calculate $w_{t,i}$ directly because we do not have access to $x_{t,i}$; we have only its measurements, $y_{t,i}$. We thus adopt the measurement-domain approach of (4.5), modifying it to the MH case. However, this makes the optimization a much more difficult, ill-posed problem, because we have to calculate the optimal linear combination within the projected space of Φ; i.e., combining (4.5) and (4.6) yields

$$\hat{w}_{t,i} = \arg\min_{w} \left\| y_{t,i} - \Phi H_{t,i} w \right\|_2 . \tag{4.8}$$

In general, $w_{t,i} \neq \hat{w}_{t,i}$ unless Φ is square, which is necessarily not the case for CS. The ill-posed nature of this problem requires some form of regularization of the LSQ optimization.

It would seem natural here to adopt the approach used in traditional CS which would be to impose a sparse prior on $w_{t,i}$ in order make this inverse problem well-posed. However, this is a problematic approach on two counts. First, the fitness of a sparse prior in this context is not necessarily known. The dictionary of hypotheses we use, $H_{t,i}$, has no special construction for which we should assume that $w_{t,i}$ should be sparse. So, while sparsity could enforce a well posed problem, and give us an agreement between $w_{t,i}$ and $\hat{w}_{t,i}$, $\tilde{x}_{t,i}$ might not be an accurate approximation of $x_{t,i}$ for sparse $w_{t,i}$. Second, since the hypotheses used to construct $H_{t,i}$ are spatially collocated, many of the atoms of $H_{t,i}$ are highly correlated. This correlation can cause the matrix $A = \Phi H_{t,i}$ to have less than desirable properties for CS recovery, increasing the number of measurements needed to accurately recover a sparse approximation. Instead, we look for a different approach to regularization which avoids these impediments.

In a normal LSQ problem, the goal is to find a solution which most closely matches a set of observations, namely, the solution which minimizes

$$\mathcal{J}(\theta) = ||\theta - Ay||_2. \tag{4.9}$$

The Moore-Penrose pseudoinverse of A gives the solution which minimizes $\mathcal{J}(\theta)$, namely,

$$\arg\min_{\theta} \mathcal{J}(\theta) = A^{\dagger}y = \left(A^T A\right)^{-1} A^T y. \tag{4.10}$$

Another approach to finding the LSQ solution is via singular value decomposition (SVD). First, decompose the matrix A,

$$A = U\Sigma V^T, \tag{4.11}$$

where Σ is a diagonal matrix of the singular values, σ_i of A. From this, $A^{\dagger}$ can be alternately calculated as

$$A^{\dagger} = VDU^T \tag{4.12}$$

where D is a diagonal matrix with entries $1/\sigma_i$.

In our setting, the matrix A is often has many more rows than columns. The underdetermination of the system $y = Ax$ causes the matrix $A^T A$ to become ill-conditioned (near singular) which prohibits the accurate calculation of the matrix inverse $\left(A^T A\right)^{-1}$. The large condition number of the matrix $A^T A$ resulting from this near singularity also removes any hope for backwards stability, causing small perturbations on y to yield drastic changes to the solution.

One of the most common approaches to address these problems has been Tikhonov regularization (also known as ridge regression) [103, 53] which introduces an ℓ_2 penalty to $\mathcal{J}(\theta)$. The solution of the regularized problem is given as the solution which minimizes

$$\mathcal{J}_T(\theta) = ||\theta - Ay||_2^2 + \lambda||\Gamma\theta||_2^2. \tag{4.13}$$

The addition of the regularization term $\lambda||\Gamma\theta||_2$ to the LSQ cost penalizes solutions which have large ℓ_2 norms. Tikhonov regularization is known as a shrinkage method because of this property.

Tikhonov regularization trades off between accuracy and variance by balancing the minimization of $\mathcal{J}_T(\theta)$ between the LSQ residual, $||\theta - Ay||_2$, and the regularizer, $||\Gamma\theta||_2$. The regularization parameter, also known as the ridge parameter, λ, balances the effect of the regularizer against the residual. The solution of the Tikhonov regularization can be calculated as

$$\arg\min_{\theta} \mathcal{J}_T(\theta) = \left(A^T A + \lambda^2 \Gamma^T \Gamma\right)^{-1} A^T y. \tag{4.14}$$

In the simple case $\Gamma = I$, the singular values of the resulting matrix, $A^T A + \lambda^2 I$, are shrunk according to

$$\tilde{\sigma}_i = \frac{\sigma_i}{\sigma_i^2 + \lambda^2}. \tag{4.15}$$

Since the condition number of a matrix is calculated as the ratio between its largest and smallest singular values,

$$\kappa = \frac{\sigma_{MAX}}{\sigma_{MIN}}, \tag{4.16}$$

the addition of the matrix $\lambda^2 I$ causes the condition number, κ, to decrease,

$$\kappa_T = \kappa \left(\frac{\sigma_{MIN}^2 + \lambda^2}{\sigma_{MAX}^2 + \lambda^2} \right). \tag{4.17}$$

Thus, if constructed properly, the Tikhonov regularization may allow for stable recovery by lowering the condition number of the matrix being inverted in (4.14).

4.2.1 Tikhonov Based MH Regularization

To calculate the MH weightings for a given block within a video frame, we construct the following Tikhonov regularization,

$$\hat{w}_{t,i} = \arg\min_{w} \|y_{t,i} - \Phi H_{t,i} w\|_2 + \lambda \|\Gamma w\|_2 , \tag{4.18}$$

where Γ is known as the Tikhonov matrix. The Γ term allows the use of prior knowledge to bias the solution; in some contexts, it may make sense to use a high-pass or difference operator for Γ to obtain a smooth result, or, in others, to set $\Gamma = I$ to impose an energy constraint on the solution.

In our method, we propose that hypotheses which are most similar to the target block should be allowed a much more significant contribution to the linear combination than those which are not. To accomplish this we construct Γ as a diagonal matrix,

$$\Gamma = \begin{bmatrix} \|y_{t,i} - \Phi h_1\|_2 & & 0 \\ & \ddots & \\ 0 & & \|y_{t,i} - \Phi h_K\|_2 \end{bmatrix}, \tag{4.19}$$

where $h_1, h_2, \ldots, h_K$ are the columns of $H_{t,i}$. With this structure, Γ penalizes weights of large magnitude assigned to hypotheses which have a significant distance from $y_{t,i}$ when projected into the measurement domain. With the structure of Γ chosen, (4.18) may be solved directly by calculating

$$\hat{w}_{t,i} = \left((\Phi H_{t,i})^T (\Phi H_{t,i}) + \lambda^2 \Gamma^T \Gamma \right)^{-1} (\Phi H_{t,i})^T y_{t,i}. \tag{4.20}$$

In this formulation, λ is a scale factor that controls the relative effect of the Tikhonov-regularization term in the optimization of (4.18). The choice of λ can have a large effect on the performance of the regularization, so it is important to find a value which imposes an adequate level of regularization without causing $\|y_{t,i} - \Phi H_{t,i} w\|_2$ to become too large. We found in practice that, over a large set of different frames, a value of $\lambda \in [0.1, 0.3]$ provided the best results; consequently, we use $\lambda = 0.25$ from this point on.

4.2.2 ℓ_1-Based MH Regularization

An alternate to the Tikhonov regularization used in (4.18)–(4.19) was suggested in [30]. Specifically, it was assumed in [30] that the MH weights $w_{t,i}$ in (4.6) are sparse; i.e., only a relative few of the possible hypotheses in $H_{t,i}$ should contribute the prediction in (4.7). As a consequence of this assumption, [30] imposes an ℓ_1-penalty term on $\hat{w}_{t,i}$ in the form of

$$\hat{w}_{t,i} = \arg\min_{w} \|\Phi H_{t,i} w - y_{t,i}\|_2 + \|w\|_1 . \tag{4.21}$$

The intuition here is that only a few blocks within the search space should contribute significantly to the linear combination; this is reflective of the structure often imposed on MH prediction in traditional video coding, structure that is necessary to limit motion-vector rate overhead. However, in the context of CS reconstruction, a regularization enforcing sparsity is needlessly restrictive on the structure of $\hat{w}_{t,i}$, which can potentially result in lower prediction quality. Furthermore, Tikhonov regularization in the form of (4.18)–(4.19) is a much more amenable solution than ℓ_1 regularization in terms of scalability and computation time, as well. That is, with the ℓ_1 penalty, the optimization in (4.21) is approached as a

traditional CS problem using some generic CS solver ([30] uses SAMP [31], for example). Such CS solvers are based on some kind of iterative search to arrive at a final solution and are thus strictly linear in the computation. Yet, the weights $\hat{w}_{t,i}$ must be calculated for every block in x_t, so the computation time can be very significant when using these linear solvers. On the other hand, the Tikhonov regularization we propose can be calculated directly at each block with simple matrix math in the form of (4.20).

A major focus of experimental results which follow is an investigation into the relative performance of the Tikhonov-regularization approach to MH prediction that we propose in Sec. 4.2.1 as opposed to that of the ℓ_1-based approach of [30]. We explore these experimental results next.

4.3 Full-Sequence Video Recovery

In the previous discussion, we examined how one might recover a single frame using MH predictions from reference frames. However, we have not yet addressed how to employ these techniques into a full-sequence video-reconstruction strategy. We do so now, introducing a video reconstruction we call MH-BCS-SPL that couples BCS-SPL still-image reconstruction [76] of MEMC residuals with MH prediction in the measurement domain as described in Sec. 4.2.

Because MH prediction requires that hypotheses be drawn from recovered reference frames, it is important that reference frames have as little distortion as possible. Consequently, "key frames," which are sampled at a relatively high subrate and therefore have relatively high quality, anchor the CS reconstruction of video; such key frames, which are

commonly used in CS reconstruction of video (e.g., [78, 56, 57, 30]) take obvious inspiration from the so-called I-frames in traditional video-compression systems. In CS reconstruction of video, these key frames are interspersed at regular intervals during acquisition, and, in order to ensure that the key frames are suitable references for frame prediction, they are acquired using a subrate typically much higher than that of the non-key frames.

The proposed MH-BCS-SPL reconstruction procedure is depicted in Fig. 4.2. MH-BCS-SPL first recovers each key frame independently using intraframe BCS-SPL. These recovered key frames are then used as references to create MH predictions of the non-key frames temporally adjacent to them; specifically, each non-key frame is predicted bidirectionally from the two nearest key frames using the Tikhonov regularized MH method of (4.18)–(4.19) where $H_{t,i}$ is formed as the union of the search windows in each reference frame. Then, the key frames themselves are reconstructed again using the same bidirectional MH prediction procedure—this time, the two adjacent recovered non-key frames serve as reference frames. Finally, these "enhanced" key frames are then used as references to bidirectionally predict each non-key frame in the video sequence.

4.4 Experimental Results

4.4.1 Single Frame Recovery

We now consider the recovery of a test frame, x_2, using the immediately preceding frame, x_1, as a reference and evaluate the various measurement-domain prediction approaches discussed in this chapter. The PSNR performance of the test-frame recovery as the subrate, S_2, for the test frame varies is presented in Figs. 4.3–4.6. Fig. 4.7 presents

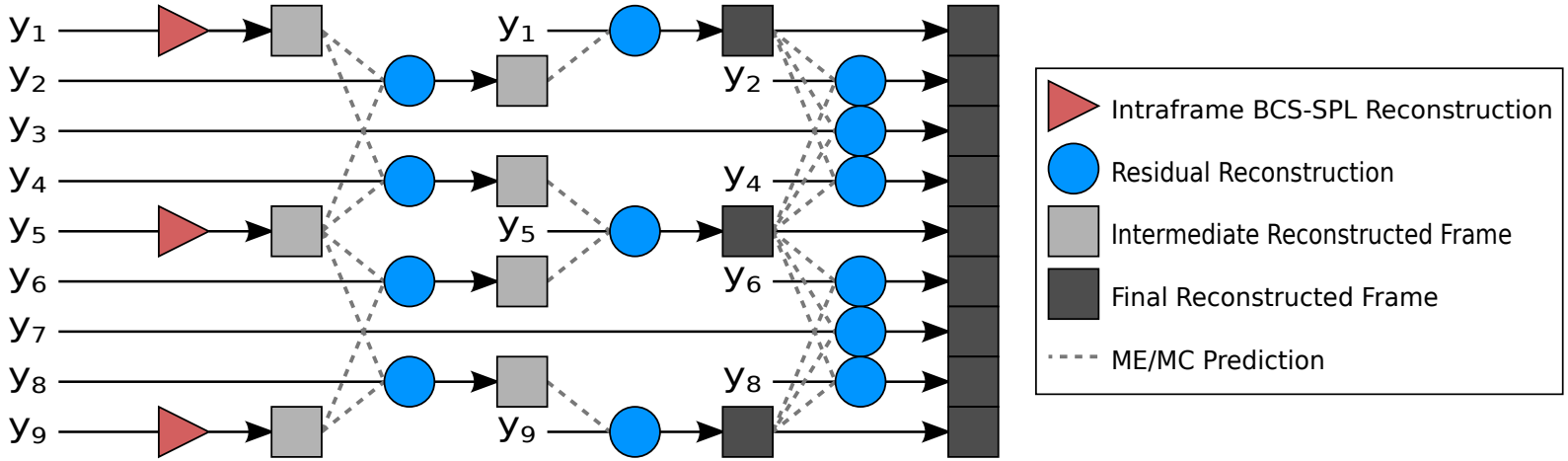


Figure 4.2

The MH-BCS-SPL reconstruction for video

Here, y_i represents a set of CS measurements corresponding to frame i; in this example, key frames occur every 5^{th} frame.

visual results of the reconstructions of a given frame of the *News* sequence sampled at a subrate of $S_2 = 0.1$. For all of these experiments, $B = 16$ pixels is used for the acquired block size and a value of $W = 15$ pixels is used for the search window size. Additionally, x_1 was sampled at a subrate of $S_1 = 0.5$ in the same manner as x_2. In all cases, BCS-SPL was used as the CS recovery procedure utilizing the 2D-DWT for Ψ. As can be seen in Figs. 4.3–4.6, the proposed Tikhonov-regularized MH prediction provides significantly superior recovery for x_2 at low subrates as compared to the ℓ_1-regularized prediction of [30]. For higher subrates near $S_2 \approx 0.5$, the performance of the ℓ_1 regularization is generally more competitive, and even exceeds that of the proposed Tikhonov regularization for the *News* sequence at $S_2 = 0.5$. However, such a high-subrate is of less interest than low-subrate reconstructions due to the necessity of minimizing the number of measurements used for non-key frames so as to maintain a low sampling rate overall. For such low-subrate reconstruction, the proposed Tikhonov-regularized prediction also yields better visual quality, as is evident in Fig. 4.7.

Interestingly, as can be seen in Figs. 4.3–4.6, the ℓ_1-regularized MH prediction does not always outperform SH prediction, though it does show consistently superior performance for the high-motion *Football* sequence (Fig. 4.5). On the other hand, the Tikhonov-regularized always outperforms the SH prediction at all subrates considered. Additionally, the proposed Tikhonov regularization appears suited to both low-motion as well as high-motion sequences (e.g., *Susie* and *Football*, respectively).

In terms of computation, SH prediction performs much more quickly than the other methods, taking just 10 to 20 seconds on our test system, while the ℓ_1 methods can take

exceedingly long to calculate, up to 4 or 5 hours for a single frame. The Tikhonov regularization, which can take just a few minutes to calculate for an entire frame, appears to be a reasonable tradeoff between increased computation time and performance gain.

The proposed method works in general for both low-motion and high-motion sequences (e.g. *Susie* and *Football*, respectively). Also, the system can work for any number of key or reference frames (and of any quality). In these experiments, we used only a single frame to form a prediction, but a bidirectional prediction can be formed by adding a reference frame to the prediction stage. This has the effect of doubling the size of our hypothesis set, and can increase the computation time of the prediction, but can overcome occlusion errors in the prediction. The experiments were repeated for the bidirectional case by using a second temporally neighboring frame which was encoded and decoded in the same manner as described for the single reference frame case.

4.4.2 Full-Sequence Video Recovery

The results of the previous section compared various prediction strategies for the reconstruction of a single video frame; however, the task of primary interest is the CS reconstruction of an entire video sequence. We now present a comprehensive comparison between several CS reconstruction algorithms for video. We use the first 88 frames of the *Foreman*, *Coastguard*, *Hall Monitor*, and *Mother and Daughter* sequences. We define a group of pictures (GOP) to be the distance P between consecutive key frames. In all cases, we use a GOP size of $P = 8$ frames with key frames starting each GOP sampled with a

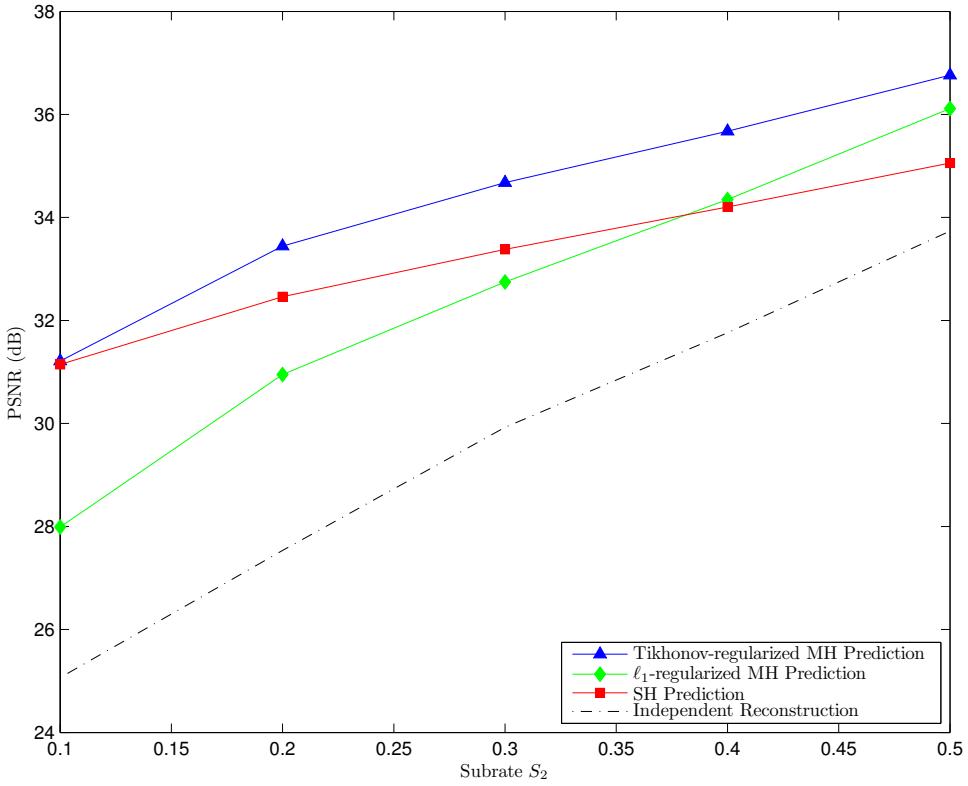


Figure 4.3

Recovery of frame x_2 of *Foreman* using frame x_1 as reference.

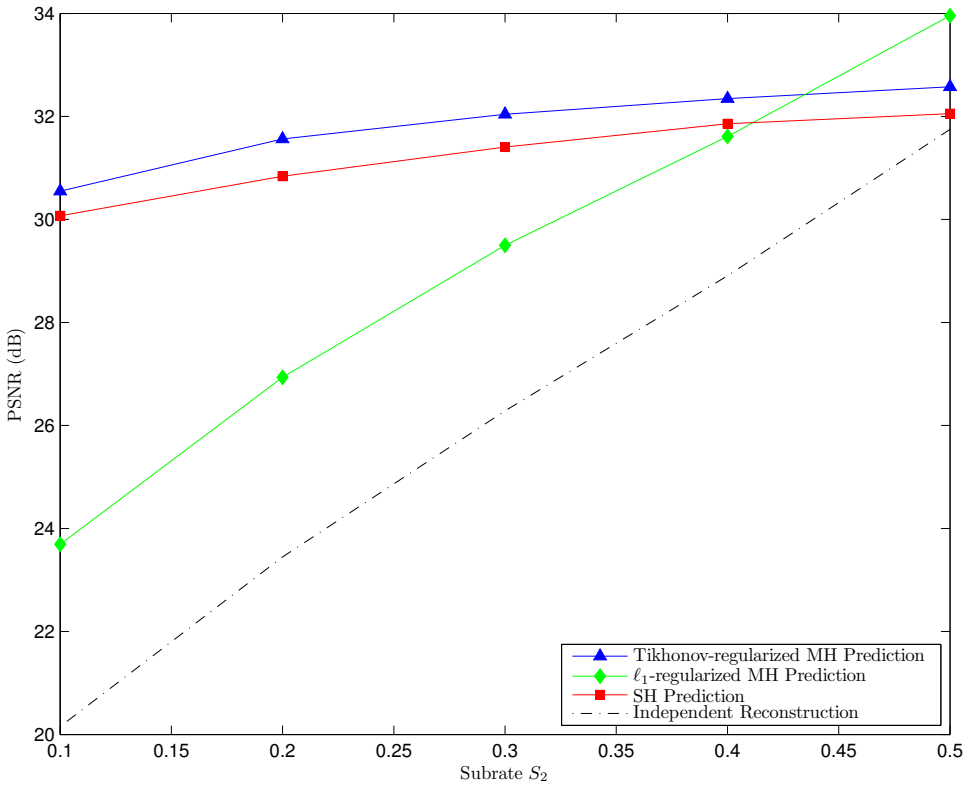


Figure 4.4

Recovery of frame x_2 of *News* using frame x_1 as reference.

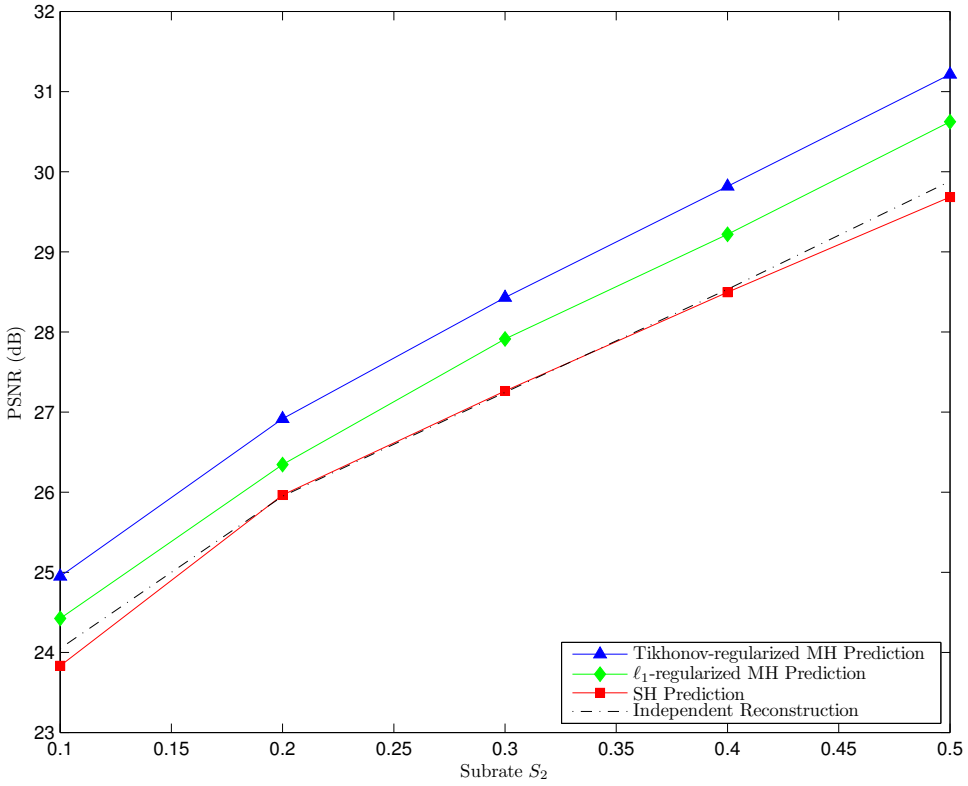


Figure 4.5

Recovery of frame x_2 of *Football* using frame x_1 as reference.

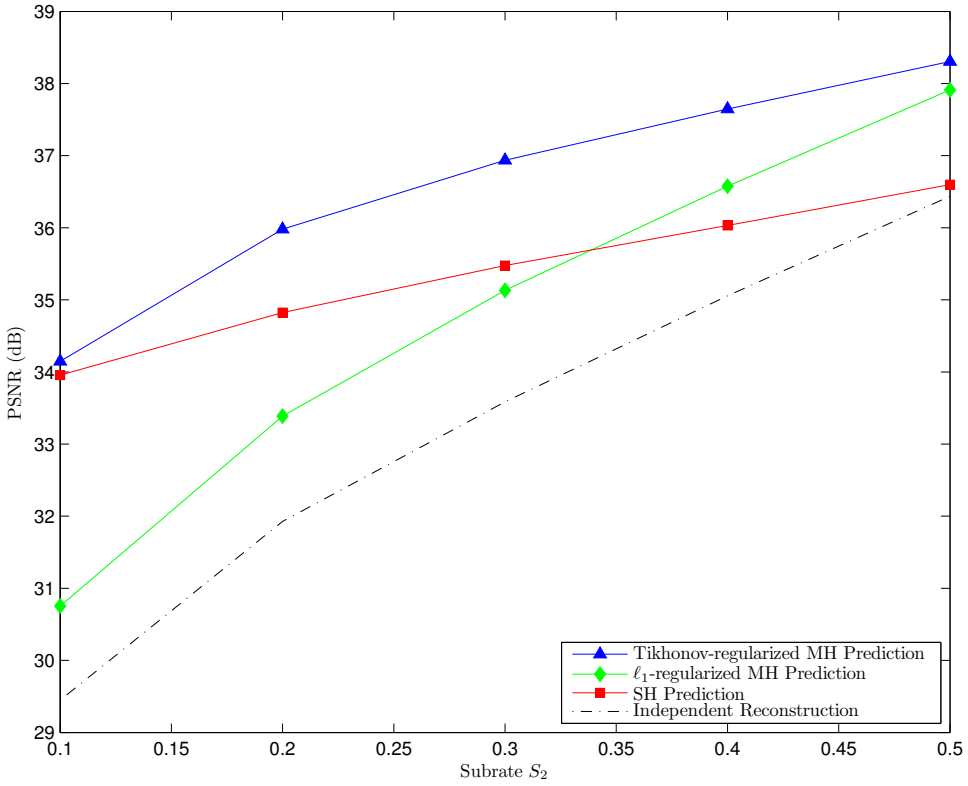


Figure 4.6

Recovery of frame x_2 of *Susie* using frame x_1 as reference.

(a) Original

(b) Independent (PSNR = 20.16 dB)

(c) RR w/ SH (PSNR = 30.07 dB)

(d) RR w/ ℓ_1-MH (PSNR = 23.69 dB)

(e) RR w/ Tikhonov-MH (PSNR = 30.55 dB)

Figure 4.7

Recovery of frame x_2 of the *News* sequence using frame x_1 as reference, $S_2 = 0.1$, $S_1 = 0.5$.

subrate of $S_{\mathrm{K}} = 0.7$. The intervening non-key frames have subrate S_{NK} varying between 0.1 and 0.5.

We compare the MH-BCS-SPL reconstruction discussed in Sec. 4.3 to MC-BCS-SPL [78]; we use the MC-BCS-SPL implementation available at the BCS-SPL website[1]. Being block-based techniques, both MH-BCS-SPL as well as MC-BCS-SPL feature block-based sampling in the spatial domain applied identically to each video frame; the block size for both techniques is 16×16, and a DDWT [60] is used as the sparsity transform.

We also compare to two prominent CS reconstruction algorithms, Modified-CS-Residual [112] and k-t FOCUSS [56, 57], both of which we have described previously in Sec. 3.2. As discussed before, k-t FOCUSS uses iterative recovery with MEMC of non-key frames from the neighboring key frames. On the other hand, Modified-CS-Residual does not employ MEMC but rather attempts to explicitly track the sparsity pattern frame to frame. We use the implementations of k-t FOCUSS[2] and Modified-CS-Residual[3] available from their respective authors. Although both k-t FOCUSS and Modified-CS-Residual were originally designed for the reconstruction of dynamic MRI data, they are both largely considered to be benchmark algorithms in present literature for the reconstruction of video as well as. Both techniques, being oriented toward dynamic MRI, feature frame-by-frame sampling driven by a 2D full-frame Fourier transform applied identically to each frame with low-frequency coefficients benefiting from a higher sampling rate.

[1] `http://www.ece.msstate.edu/~fowler/BCSSPL/`
[2] `http://bisp.kaist.ac.kr/research_02.htm`
[3] `http://home.engineering.iastate.edu/~luwei/modcs/`

Finally, we compare to straightforward, "intraframe" reconstruction of each frame of the sequence independently from the others. We consider the multiscale (MS) variant of BCS-SPL originally proposed in [41]; in the results here, we refer to it as "intraframe MS-BCS-SPL." We also consider an intraframe implementation of TV reconstruction [12] ("intraframe TV"). We note that, in the results of [41], MS-BCS-SPL and TV outperformed other techniques in terms of reconstruction quality for a single still image, with MS-BCS-SPL generally producing higher-quality reconstructions with much less computation, but TV being amenable to fast, spatial-domain sampling using a structurally random matrix (SRM) [46]. In these results, the intraframe MS-BCS-SPL features block-based sampling in the wavelet domain with blocks of size 16×16, while intraframe TV uses a full-frame block-Hadamard SRM sampling [46]. Although Sec. 3.2 surveys a number of reconstruction algorithms for video, none of these other than k-t FOCUSS and Modified-CS-Residual have, to our knowledge, implementations readily available at the time of this writing. As a consequence, we present results for only those algorithms identified above.

Figs. 4.8–4.11 illustrate the performance of the various reconstructions for varying non-key-frame subrate S_{NK}. Visual results for a single frame of the *Foreman* sequence are given in Fig. 4.12. As is apparent, MH-BCS-SPL almost always outperforms the other techniques considered, sometimes by as much as 2–3 dB. The sole exceptions are the lowest subrate for *Hall Monitor*, at which k-t FOCUSS slightly outperforms MH-BCS-SPL, and the highest subrate for *Foreman*, at which intraframe TV outperforms MH-BCS-SPL. The performance of the techniques other than MH-BCS-SPL is rather mixed—sometimes MC-BCS-SPL or intraframe TV will be somewhat competitive with MH-BCS-SPL for cer-

tain subrates and sequences. Additionally, the two techniques designed for dynamic MRI (k-t FOCUSS and Modified-CS-Residual) are typically rather distant in performance from MH-BCS-SPL with the exception of low subrates for the *Hall Monitor* sequence.

Although none of the implementations have been particularly optimized for execution speed, we present reconstruction times for the algorithms in Table 4.1. Here, we measure the average length of time required to reconstruction one frame out of the sequence. We see that, while the intraframe MS-BCS-SPL is reconstruction is the fastest, the intraframe TV reconstruction is the slowest, requiring some 20 minutes per frame.

Table 4.1

Reconstruction time in seconds per frame (spf)

Algorithm	*Time (spf)*
Intraframe MS-BCS-SPL	10
k-t FOCUSS	46
MC-BCS-SPL	159
MH-BCS-SPL	324
Modified-CS-Residual	699
Intraframe TV	1223

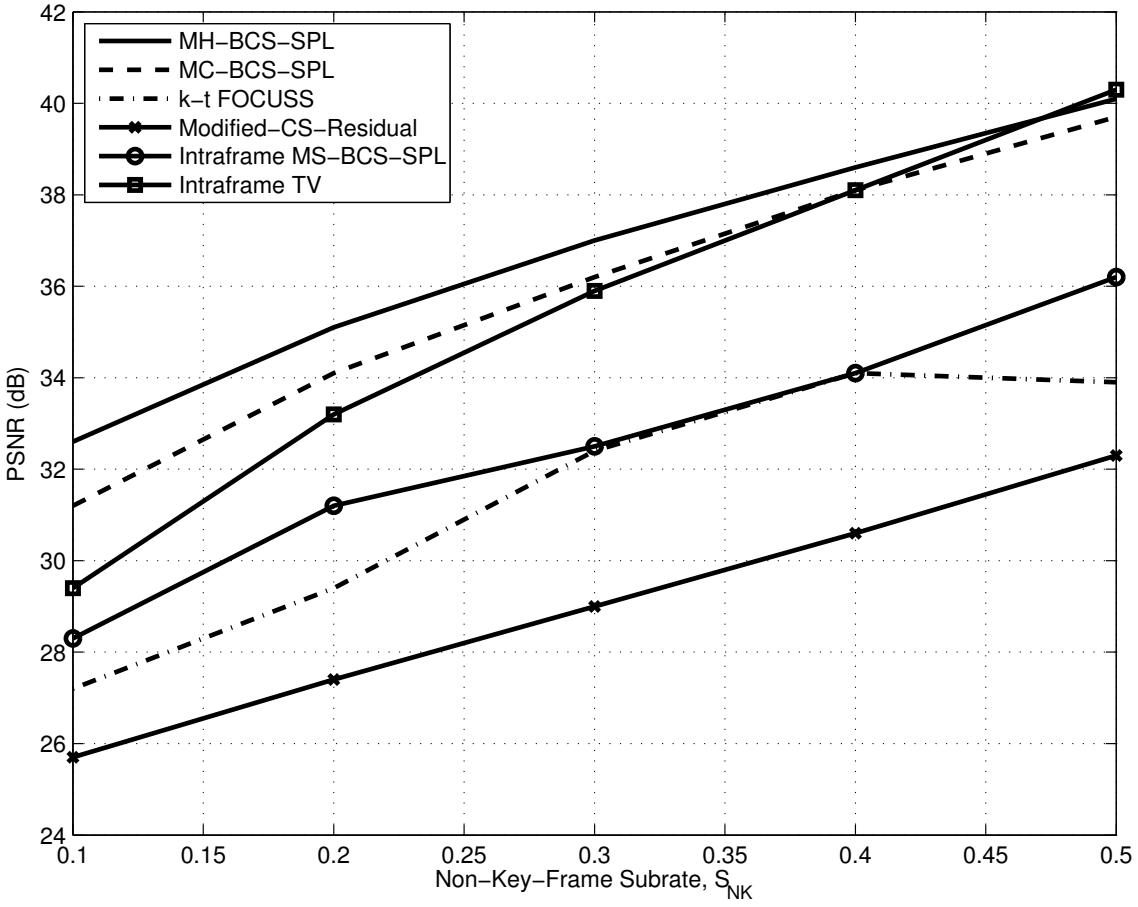


Figure 4.8

Performance of various CS reconstruction algorithms on the 88-frame *Foreman* sequence for $S_{\mathrm{K}} = 0.7$.

PSNR is averaged over all frames of the sequence.

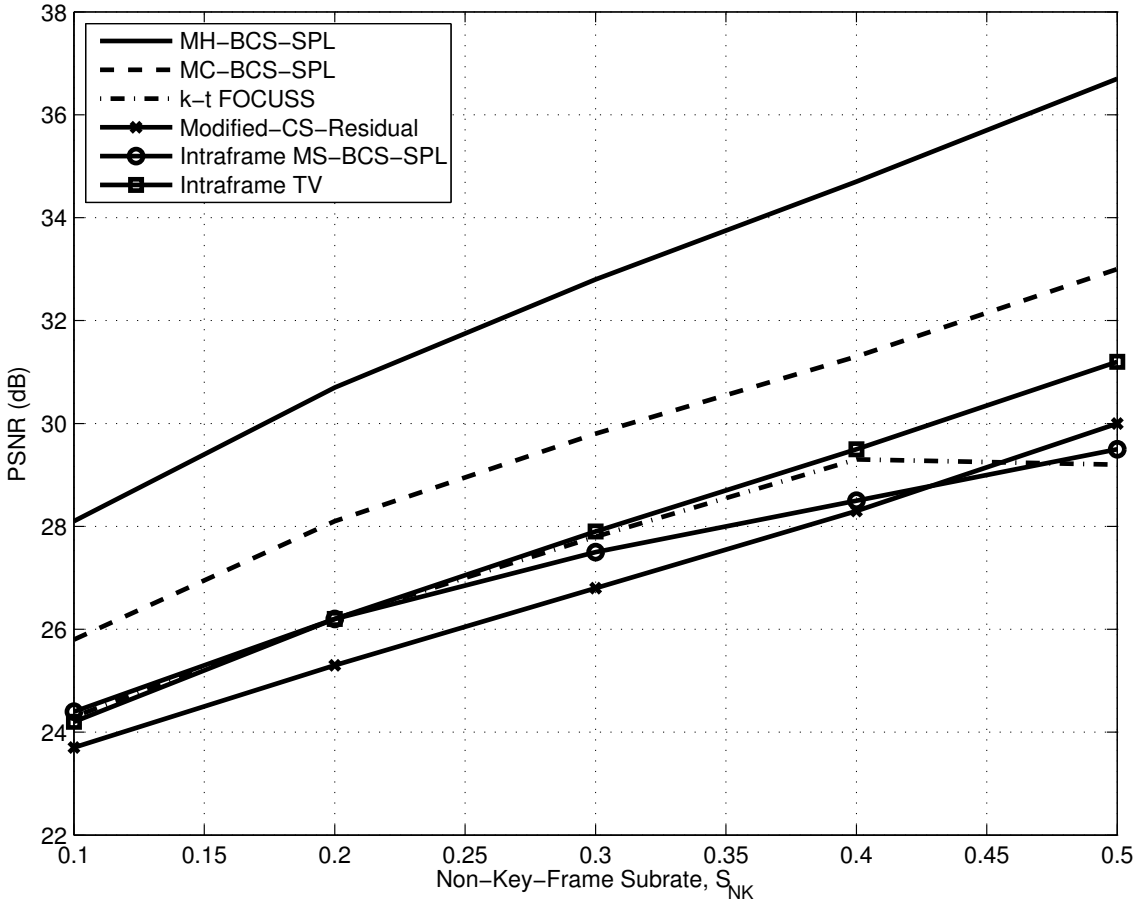


Figure 4.9

Performance of various CS reconstruction algorithms on the 88-frame *Coastguard* sequence for $S_{\mathrm{K}} = 0.7$.

PSNR is averaged over all frames of the sequence.

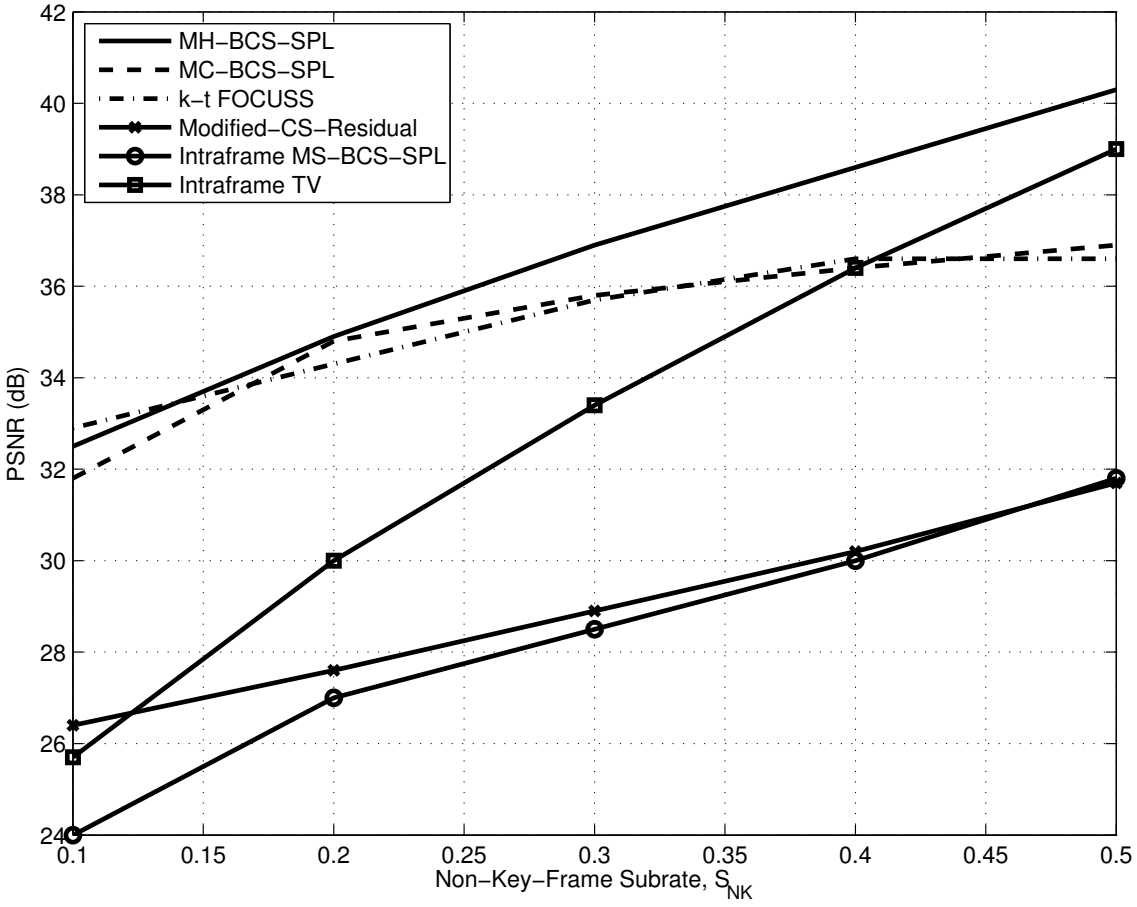


Figure 4.10

Performance of various CS reconstruction algorithms on the 88-frame *Hall Monitor* sequence for $S_{\mathrm{K}} = 0.7$.

PSNR is averaged over all frames of the sequence.

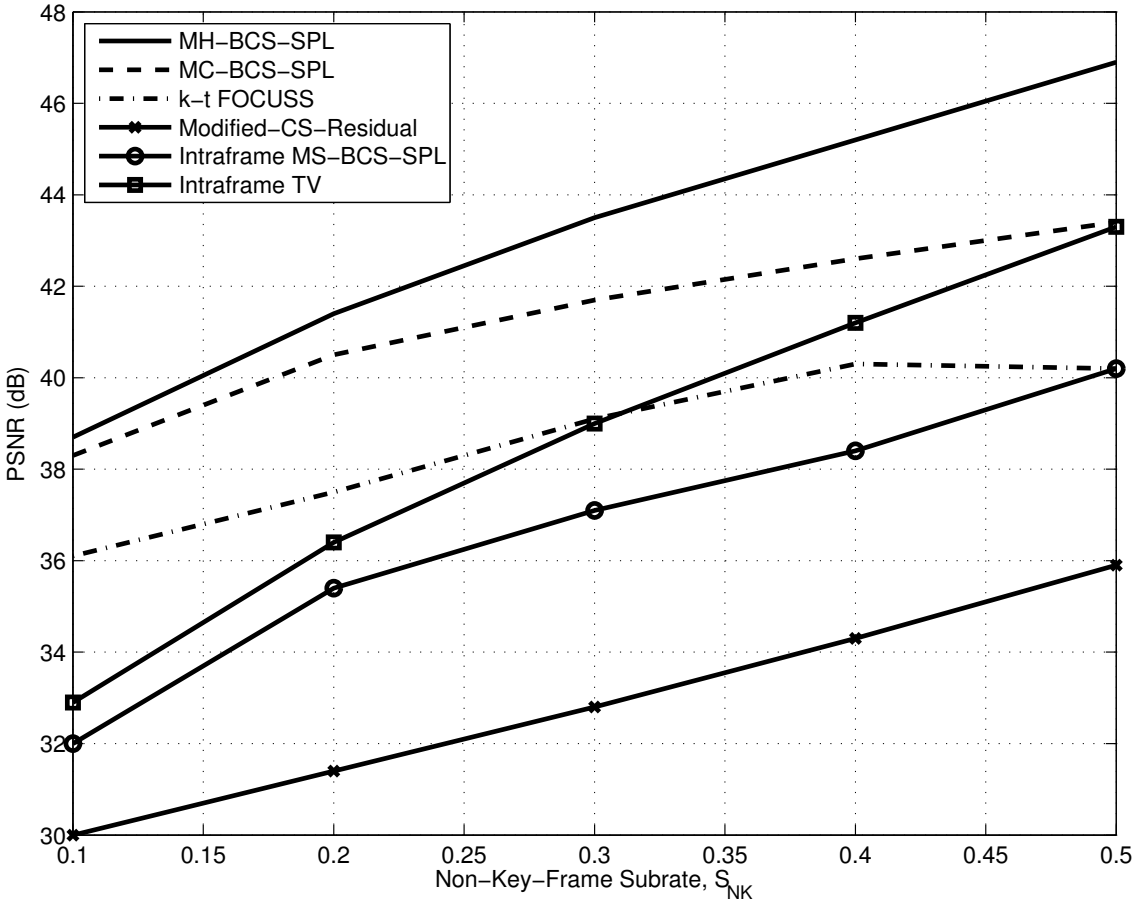


Figure 4.11

Performance of various CS reconstruction algorithms on the 88-frame *Mother and Daughter* sequence for $S_{\mathrm{K}} = 0.7$.

PSNR is averaged over all frames of the sequence.

(a) MH-BCS-SPL (PSNR = 37.0 dB)

(b) MC-BCS-SPL (PSNR = 36.7 dB)

(c) k-t FOCUSS (PSNR = 32.7 dB)

(d) Modified-CS-Residual (PSNR = 29.6 dB)

(e) Intraframe MS-BCS-SPL (PSNR = 33.1 dB)

(f) Intraframe TV (PSNR = 36.7 dB)

Figure 4.12

Reconstructions of frame 4 of the *Foreman* sequence for $S_{\mathrm{K}} = 0.7$ and $S_{\mathrm{NK}} = 0.3$.

CHAPTER 5

NEAREST REGULARIZED SUBSPACE FOR CLASSIFICATION

In the previous chapter, we saw how the use of distance-weighted Tikhonov regularization can benefit video-frame recovery by approximating blocks in each frame via a linear combination of blocks drawn from a key frame. In this setting, Tikhonov regularization was used in conjunction with residual reconstruction to accomplish the recovery of a signal from a set of random projections.

The distance-weighted Tikhonov regularization we demonstrated is not limited to CS recovery alone, however. In this chapter, we present how this form of regularization can also be used for supervised classification tasks. Additionally, present a series of experiments designed to evaluate the accuracy of the proposed classifier in correctly identifying the material corresponding to unknown spectral reflectances. We conduct these tests on ground-truthed hyperspectral-image (HSI) data captured from airborne hyperspectral sensors. We have submitted a journal article [66], and its subsequent reivision, on this topic.

5.1 Classification of Hyperspectral Imagery

Over the last decade, HSI obtained by remote-sensing systems has been investigated at length [62]. HSI provides high-resolution spectral information over a wide range of the electromagnetic spectrum with hundreds of observed spectral bands. Numerous supervised

classification techniques for hyperspectral data have been developed (e.g., [3, 70, 109, 101]) for a variety of application areas, including agricultural monitoring, environment-pollution monitoring, and urban-growth analysis, among others.

The k-nearest-neighbor (k-NN) classifier (e.g., [92, 69]), one of the simplest and oldest classification methods, has been used widely for HSI classification. This non-parametric classifier usually employs a Euclidean distance metric between the training and testing samples, assigning class labels according to the most frequently occurring class of the k nearest training samples. However, the high-dimensional nature of HSI data creates complications for k-NN classification in terms of both computational complexity and classification accuracy. Many dimensionality-reducing techniques have been proposed to combat this so-called "curse of dimensionality," such as the popular linear discriminant analysis (LDA) [38] and its variants (e.g., [64, 86]). Typically, parametric classification is employed after dimensionality reduction, for example the maximum likelihood estimation (MLE) [29] of posterior probabilities. The support vector machine (SVM) [2] is a state-of-the-art classifier which has also been shown to work well for hyperspectral classification tasks. An SVM seeks to separate classes by learning an optimal decision hyperplane which best separates the training samples in a kernel-induced high-dimensional feature space. Variations of the SVM (e.g., [65, 70]) have been proposed to further improve classification performance.

Recently, Wright *et al*. [119] introduced sparse-representation classification (SRC) for face recognition. Later, Chen *et al*. [24] applied the sparse-representation method for HSI classification. In essence, the SRC represents a testing sample by a sparse linear combi-

nation of training samples calculated via ℓ_1 minimization. A similar approach was taken by Zhang *et al*. [122] who proposed collaborative-representation classification (CRC) for face recognition. However, contrary to the ℓ_1-based sparsity-inducing regularization of SRC, CRC uses an ℓ_2-regularized minimization, providing competitive face-recognition accuracy but at significantly lower computational complexity.

In this work, we couple nearest-subspace classification with the distance-weighted Tikhonov regularization from [104, 42]. In resulting system, which can be considered to be a nearest-regularized-subspace (NRS) classifier, an approximation for each testing sample is created via linear combination of all available training samples within each class. In this manner, an approximation of each test sample is generated from training samples of each class independently. The class label is then derived according to the class of the most accurate representation. In a general sense, this NRS classification is similar to both SRC and CRC in that testing samples are approximated via linear combinations of training samples; however, NRS differs in that, not only does it use a non-collaborative approach to the approximation, but it also employs non-uniform regularization.

We also introduce, as a further extension of the proposed NRS paradigm, a discrim–ination-enhancing distance measure [117] designed to improve classification accuracy. Furthermore, a competitive strategy is presented for automatically obtaining optimal performance for the proposed system, thus avoiding involved parameter tuning via cross-validation. Classification results are presented for several HSI datasets to demonstrate the superior classification accuracy of the proposed approach when compared to traditional classification techniques. Ultimately, our work is composed of three main contri-

butions: (1) the NRS classification system based on a distance-weighted Tikhonov regularization (an ℓ_2-regularized term) calculating a representation for each testing sample; (2) a discrimination-enhancing distance measure which improves the Tikhonov biasing term; and (3) a competitive strategy that eliminates the need for involved parameter tuning.

5.2 Related Classification Techniques

5.2.1 Nearest-Neighbor Classification

The nearest-neighbor (NN) algorithm (e.g., [92, 69]) is perhaps the simplest supervised method to predict a testing-sample label. The NN classifier attempts to find the training sample nearest to the testing sample according to a given distance measure, assigning the former's category to the latter. Consider a dataset with training samples $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}$ in $\mathbb{R}^d$ (d-dimensional feature space) and class labels $\omega_i \in \{1, 2, \ldots, C\}$, where C is the number of classes, and n is the total number of training samples. Let n_l be the number of available training samples for the l^{th} class, $\sum_{l=1}^{C} n_l = n$. Commonly Euclidean distance is used, such that the distance measure between training sample $\mathbf{x}_i$ and given testing sample $\mathbf{y}$ is

$$d(\mathbf{x}_i, \mathbf{y}) = \left\|\mathbf{x}_i - \mathbf{y}\right\|_2^2. \tag{5.1}$$

The k-NN classifier is a straightforward extension of the original NN classifier. Instead of using only one sample closest to testing point $\mathbf{y}$, the k-NN classifier chooses the k nearest samples from training data $\mathbf{X}$. Typically, k is an odd number, and majority voting is employed to decide the final label.

5.2.2 Nearest Subspace Classification

Instead of using the nearest training samples as indicators for the classification of a given test sample, the *Nearest Subspace* (NS) approach operates on the assumption that the nearest test samples from the true class form a subspace which the test sample lies near. More specifically, say that we are given a test sample, $\mathbf{y} \in \mathbb{R}^d$ which is drawn from one of C classes. Then, using each class's training samples, $\mathbf{X}_l \in \mathbb{R}^{d \times n_l}$, $l \in \{1, \ldots, C\}$, C subspaces are formed. The correlation between these subspaces is dependent upon the training data used and the generating model of each class. One approach for generating these class-specific subspaces is to use the Gram-Schmidt orthonormalization procedure, whereby an orthogonal basis, $\mathbf{U}_l$, is formed to project onto $\mathrm{span}\,(\mathbf{X}_l)$. In the class featuring information compression (CLAFIC) classifier [83], each $\mathbf{U}_l$ is formed using the first k eigenvectors of the class covariance matrix, $\mathbf{\Sigma}_l = \mathbf{X}_l^T \mathbf{X}_l$. In other words, $\mathbf{U}_l \in \mathbb{R}^{k \times D}$ consists of the first k principal components (PCs) of $\mathbf{X}_l$. $\mathbf{U}_l$ may even be constructed as a random selection of k training samples from class l.

If $\mathbf{U}_l$ is orthogonal, we can use it to construct a $\mathbb{R}^d \mapsto \mathbb{R}^d$ linear mapping operator for class l,

$$\mathbf{P}_l = \mathbf{U}_l {\mathbf{U}_l}^T. \tag{5.2}$$

However, if $\mathbf{U}_l$ is non-orthogonal, we still may form $\mathbf{P}_l$ via the Moore-Penrose pseudo-inverse,

$$\mathbf{P}_l = \mathbf{U}_l \mathbf{U}_l^\dagger = \mathbf{U}_l \left({\mathbf{U}_l}^T \mathbf{U}_l \right)^{-1} {\mathbf{U}_l}^T. \tag{5.3}$$

Using $\mathbf{P}_l$, we find the representation of $\mathbf{y}$ mapped onto $\mathcal{S}_l$, a k-dimensional subspace of $\mathbb{R}^d$,

$$\widehat{\mathbf{y}}_l = \mathbf{P}_l \mathbf{y}. \tag{5.4}$$

In the k-NN classifier, classification is determined according to proximity. For the NS classifier, class assignment is determined by how well $\mathcal{S}_l$ is aligned with $\mathbf{y}$ according to the length of $\widehat{\mathbf{y}}_l$,

$$g_{NS}(\mathbf{y}) = \arg \max_{l=1,\ldots,C} \quad ||\widehat{\mathbf{y}}_l||^2. \tag{5.5}$$

If the basis vectors of $\mathbf{P}_l$ are not of unit length, the classification may be modified to

$$g_{NS}(\mathbf{y}) = \arg \max_{l=1,\ldots,C} \quad \frac{||\widehat{\mathbf{y}}_l||^2}{||\mathbf{y}||^2}. \tag{5.6}$$

Refer to Fig. 5.1 for a demonstration of the NS classifier discriminating between two classes for a two dimension setting.

Laarksonen proposed a different approach to the NS classifier with the *Local Subspace Classifier* (LSC) [61]. The LSC, instead of finding the class which produces the most aligned subspace, sought to find the class which could produce a linear manifold from training points local to $\mathbf{y}$ to best represent $\mathbf{y}$. This was done by selecting the $k+1$ nearest samples from each classes and choosing one additional offset point, μ. This offset was then used to recenter the data such that μ was located at the origin, $\mathbf{X}_{l,\mu} = [\mathbf{x}_{l,1} - \mu, \mathbf{x}_{l,2} - \mu, \ldots, \mathbf{x}_{l,k} - \mu]$. The projection matrix, $\mathbf{P}_{l,\mu}$ was then constructed in an identical fashion to the traditional NS. Finally, the mapping was calculated as

$$\widehat{\mathbf{y}}_{l,\mu} = \mathbf{P}_{l,\mu}(\mathbf{y} - \mu) + \mu. \tag{5.7}$$

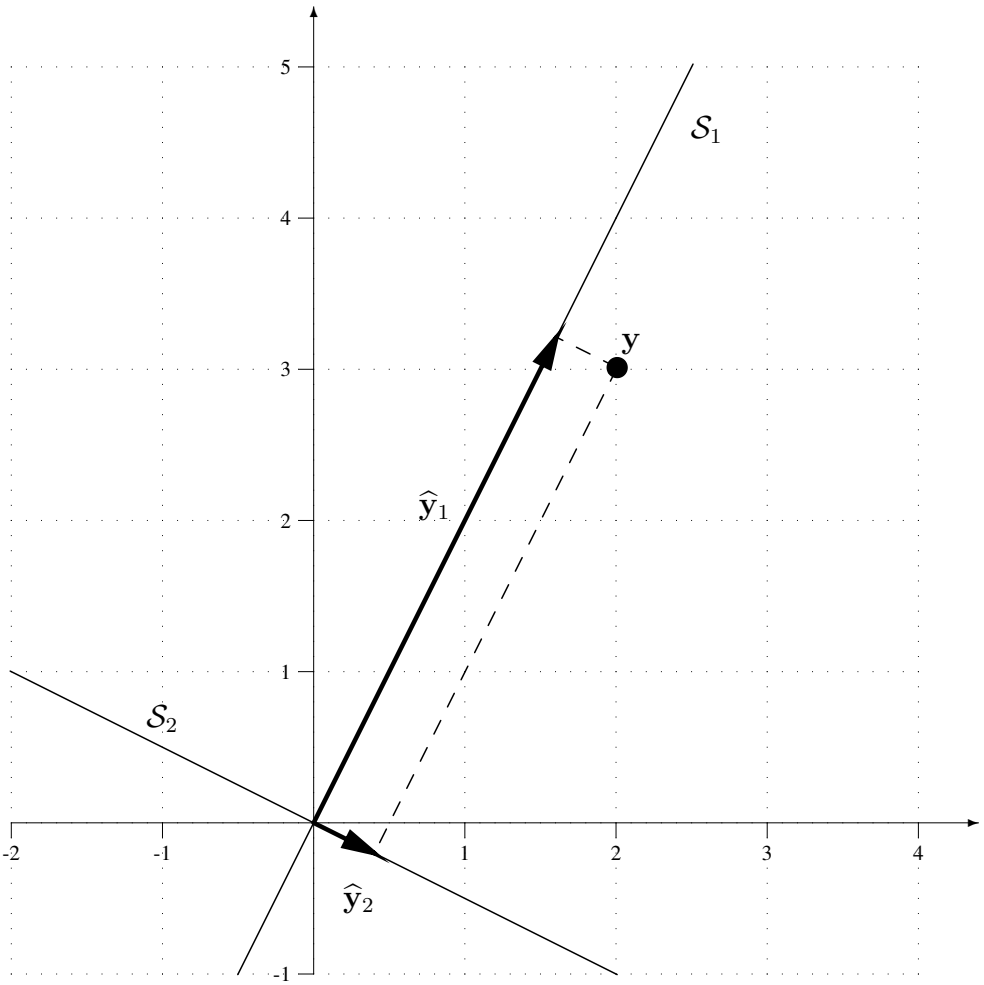


Figure 5.1

Two-dimensional representation of the operation of the NS classifier. Since $||\widehat{\mathbf{y}}_1|| > ||\widehat{\mathbf{y}}_2||$, g_{NS} assigns $\mathbf{y}$ to Class 1 which produced $\mathcal{S}_1$.

The classification was then decided by the class whose generated manifold $\mathcal{S}_{l,\mu}$ provided the best representation of $\mathbf{y}$ by having the smallest residual, which can be calculated as

$$\widetilde{\mathbf{y}}_l = \mathbf{y} - \widehat{\mathbf{y}}_{l,\mu} = (\mathbf{I} - \mathbf{P}_{l,\mu})(\mathbf{y} - \mu). \tag{5.8}$$

The final classification is then

$$g_{LSC}(\mathbf{y}) = \arg \min_{l=1,\ldots,C} \quad ||\widetilde{\mathbf{y}}_l||^2. \tag{5.9}$$

In Laarksonen's work, it was shown that the LSC works very well for the classification of data which can be well characterized by linear manifolds, such as hand-written digit analysis [61]. However, it is not directly apparent that this should be true for the HSI data we wish to classify. For this reason, we mainly focus on the NS approach to classification. We also choose this paradigm because it also closely aligns with other recently proposed classification techniques which we will discuss next.

5.2.3 ℓ_1- and ℓ_2-Regularized Collaborative Representation for Classification

Classification based on sparse representation has been recently studied for both for face recognition [119], and HSI analysis [24]. The SRC approach offers classification which is robust to noise and model errors; for more discussion of the geometrical and graphical interpretations of SRC, we refer the reader to [119].

In essence, an SRC method classifies a testing sample $\mathbf{y}$ according to the class which produces the most accurate sparse representation of $\mathbf{y}$, i.e., the class which produces the most parsimonious description using the training data as the "dictionary" for forming the

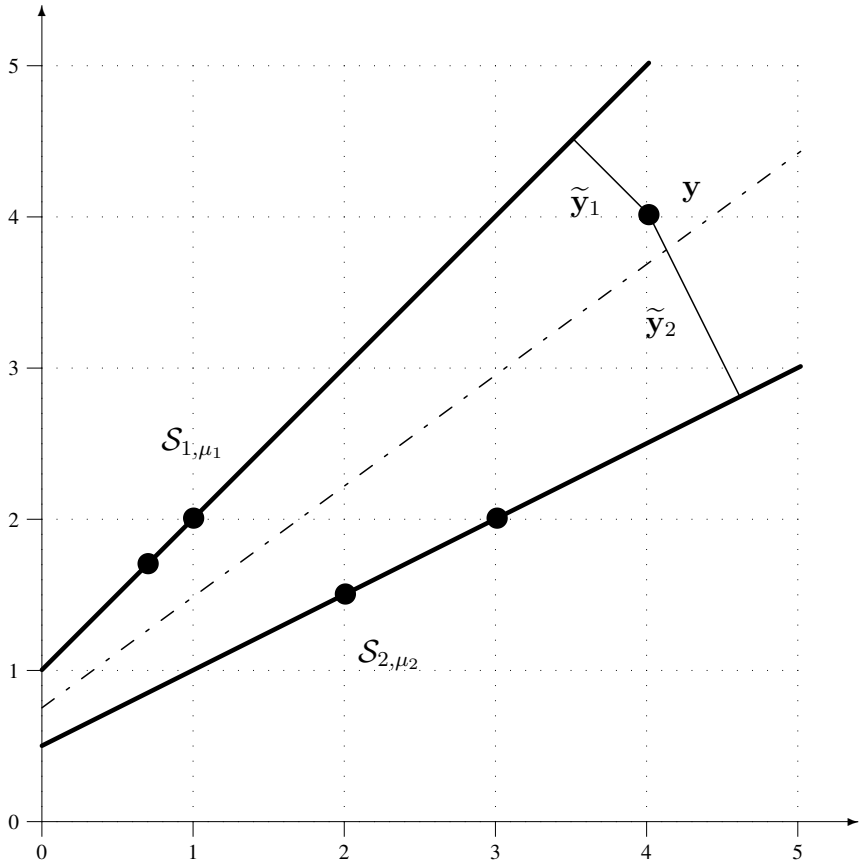


Figure 5.2

Two-dimensional representation of the operation of the LSC classifier.

Here, one-dimensional linear manifolds are defined by two points from each class. The decision boundary caused by g_{LSC} is shown as the dash-dotted line. In this case, since the residual to Class 1, $\widetilde{\mathbf{y}}_1$, is shortest, $\mathbf{y}$ is assigned to Class 1.

representation. First, an approximation of $\mathbf{y}$ is calculated via a sparse linear combination of all available training samples. That is, for training samples arranged column-wise in $\mathbf{X}$,

$$\widetilde{\mathbf{y}} = \mathbf{X}\boldsymbol{\alpha}, \tag{5.10}$$

where $\mathbf{X}$ is of dimensionality $d \times n$, and $\boldsymbol{\alpha}$ is a $n \times 1$ vector of sparse coefficients. Basis pursuit denoising (BPDN) [22] offers one approach for calculating $\boldsymbol{\alpha}$ by solving the ℓ_1-regularized minimization,

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\theta}} \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right\|_2^2 + \lambda\left\|\boldsymbol{\theta}\right\|_1, \tag{5.11}$$

where the regularization parameter, $\lambda > 0$, balances the influence of the residual and sparsity terms. We mention the BPDN formulation in particular here because of its confluence with several regularization techniques we present later. However, other formulations may be equivalently substituted, such as the least absolute shrinkage and selection operator (LASSO) [102] or basis pursuit (BP) [22]. In any event, after $\boldsymbol{\alpha}$ is calculated, a representation for each class, $\widetilde{\mathbf{y}}_l$, is created through a process we term *post-partitioning*.

The post-partitioning approach separates $\mathbf{X}$ into l different sub-dictionaries, $\mathbf{X}_l = \{\mathbf{x}_i \mid \forall i \; s.t. \; \omega_i = l\}$; additionally, the coefficient vector $\boldsymbol{\alpha}$ is also "partitioned" similarly into $\boldsymbol{\alpha}_l = \{\alpha_i \mid \forall i \; s.t. \; \omega_i = l\}$. After this partitioning, class-specific representations, $\widetilde{\mathbf{y}}_l$, are calculated as

$$\widetilde{\mathbf{y}}_l = \mathbf{X}_l\boldsymbol{\alpha}_l. \tag{5.12}$$

We note that this use all the training data concurrently, as in post-partitioning, stands in contrast to the traditional approach used in NS classifiers [61, 67] which use what we call *pre-partitioning*. In such pre-partitioning, the training data is first partitioned into $\mathbf{X}_l$, and

these partitions are instead used to calculate each $\widetilde{\mathbf{y}}_l$ independently, via, e.g., BPDN applied independently for each partition.

In SRC, after calculating each $\widetilde{\mathbf{y}}_l$ via (5.12), the class label of $\mathbf{y}$ is then determined according to the class which minimizes the residual. That is,

$$\text{class}(\mathbf{y}) = \arg \min_{l=1,\ldots,C} (r_l), \tag{5.13}$$

where $r_l = \left\|\widetilde{\mathbf{y}}_l - \mathbf{y}\right\|_2^2$ is the residual between the approximation and corresponding testing sample. A detailed description of the SRC algorithm is given as Fig. 5.3.

input : Training data $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$, class labels ω_i, testing sample $\mathbf{y} \in \mathbb{R}^d$, λ
Calculate $\boldsymbol{\alpha}$ via ℓ_1-minimization of (5.11);
for $l \in \{1, 2, \ldots, C\}$ **do**
    Partition $\mathbf{X}_l$, $\boldsymbol{\alpha}_l$;
    Calculate $\widetilde{\mathbf{y}}_l = \mathbf{X}_l \boldsymbol{\alpha}_l$;
end
Decide $\text{class}(\mathbf{y})$ via (5.13);
output: $\text{class}(\mathbf{y})$

Figure 5.3

The SRC Algorithm

In [119, 24], it was posited that the sparse representation alone led to the observed improvements in classification accuracy. However, both [89] and [122] raise concerns over the SRC framework. In [89], it was shown via analysis of singular values that face datasets are, generally, not a suitable fit for SRC. To show that a sparse approach is unwarranted for face recognition, a QR decomposition was used to calculate each $\widetilde{\mathbf{y}}_l$ instead of sparse

approximation; the resulting performance this technique was competitive with that of the SRC.

Additionally, in [122], it was suggested that the improvement in classification accuracy was not due to sparsity, but rather due to the "collaborative" nature of the approximation. Specifically, it was argued that using the entire training dataset to form approximations via post-partitioning rather than using pre-partitioning as in NS allows for acceptable classification accuracy when signal dimensionality is high or when the number of available training samples are few. To support this argument, [122] proposed the CRC approach which swapped the ℓ_1 penalty of SRC for an ℓ_2 penalty in the style of Tikhonov regularization [103]; i.e.,

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\theta}} \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\right\|_2^2 + \lambda \left\|\boldsymbol{\theta}\right\|_2^2. \tag{5.14}$$

Rather than enforcing a strong assumption about the nature of the dataset's geometry, the ℓ_2 regularization (or *shrinkage*) term instead serves only to overcome the potential for ill-conditioning and ill-posedness in the inverse problem.

One particular advantage of CRC is that (5.14) may be solved with a simple and closed form,

$$\widetilde{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda^2\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{P}_{CRC}\mathbf{y}, \tag{5.15}$$

where the $\mathbf{I}$ is an identity matrix of appropriate size. After calculating $\widetilde{\mathbf{y}}$, the post-partitioning and classification is carried out in a manner identical to the SRC via (5.12)–(5.13). It is noted in [122] that $\mathbf{P}_{CRC}$ is dependent upon only the available training data. Thus, the projector $\mathbf{P}_{CRC}$ may be precomputed to reduce classification time for large volume tasks.

CRC was shown to provide face-recognition accuracy comparable to SRC with much lower computational cost. A detailed description of the CRC is given as Fig. 5.4.

input : Training data $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}$, class labels ω_i, testing sample $\mathbf{y} \in \mathbb{R}^d$, λ
Calculate $\boldsymbol{\alpha}$ via (5.14);
for $l \in \{1, 2, \ldots, C\}$ **do**
  Partition $\mathbf{X}_l$, $\boldsymbol{\alpha}_l$;
  Calculate $\widetilde{\mathbf{y}}_l = \mathbf{X}_l \boldsymbol{\alpha}_l$;
end
Decide $\mathrm{class}(\mathbf{y})$ via (5.13);
output: $\mathrm{class}(\mathbf{y})$

Figure 5.4

The SRC Algorithm

The common element between these works and the sparse approaches of [119, 24] is the assumption of a collaborative, post-partitioning framework for calculating class representations, $\widetilde{\mathbf{y}}_l$. However, this general approach is only loosely justified in previous literature with few significant details given for the departure from the NS approach of pre-partitioning.

We investigate the effects of pre- and post-partitioning empirically for hyperspectral data in Fig. 5.5 using the Indian Pines dataset with 1496 training samples (see Sec. 5.4.1 for a detailed description of this dataset). The classification accuracy is calculated over a range of possible values for the free regularization parameter, λ. We denote the pre-partitioning technique here as CRC-Pre. The only difference between CRC-Pre and the post-partitioning-based CRC is that each $\widetilde{\mathbf{y}}_l$ is calculated in the former using only the training samples from class l, $\mathbf{X}_l$. Even though HSI data resides in the context proposed for

collaborative techniques—namely high-dimensionality data with few training samples—Fig. 5.5 shows collaborative post-partitioning may actually do more harm than good. From these results, it is evident that advances in face recognition using collaborative approximations cannot be applied wholesale to HSI classification. We argue that a different approach is required.

5.3 Nearest Regularized Subspace Classifier

5.3.1 The Basic NRS Algorithm

In this section, we propose the NRS classifier which couples pre-partitioning as in NS with non-uniform Tikhonov regularization for the classification of hyperspectral data when few training samples are available. Like CRC, NRS makes use of Tikhonov regularization [103] to generate each $\widetilde{\mathbf{y}}_l$. However, instead of using uniform regularization as CRC does, we adopt a technique proposed in [104, 42] which is described previously in Chapter 4, therein termed multihypothesis (MH) prediction, which biases atoms of $\mathbf{X}_l$ according to their Euclidean distance from $\mathbf{y}$. In [104, 42], the MH prediction method was used to recover video macroblocks from a small set of random linear measurements taken on the encoder side when a set of high-quality keyframe macroblocks was available on the decoder side via a linear combination of these keyframe macroblocks. The non-uniform nature of the regularization was used to penalize potentially inaccurate macroblocks from being assigned large contributions in the final recovery.

Likewise, in supervised classification, we are given a set of training, or hypothesis, data from which we desire to create approximations via linear combination. Namely, we

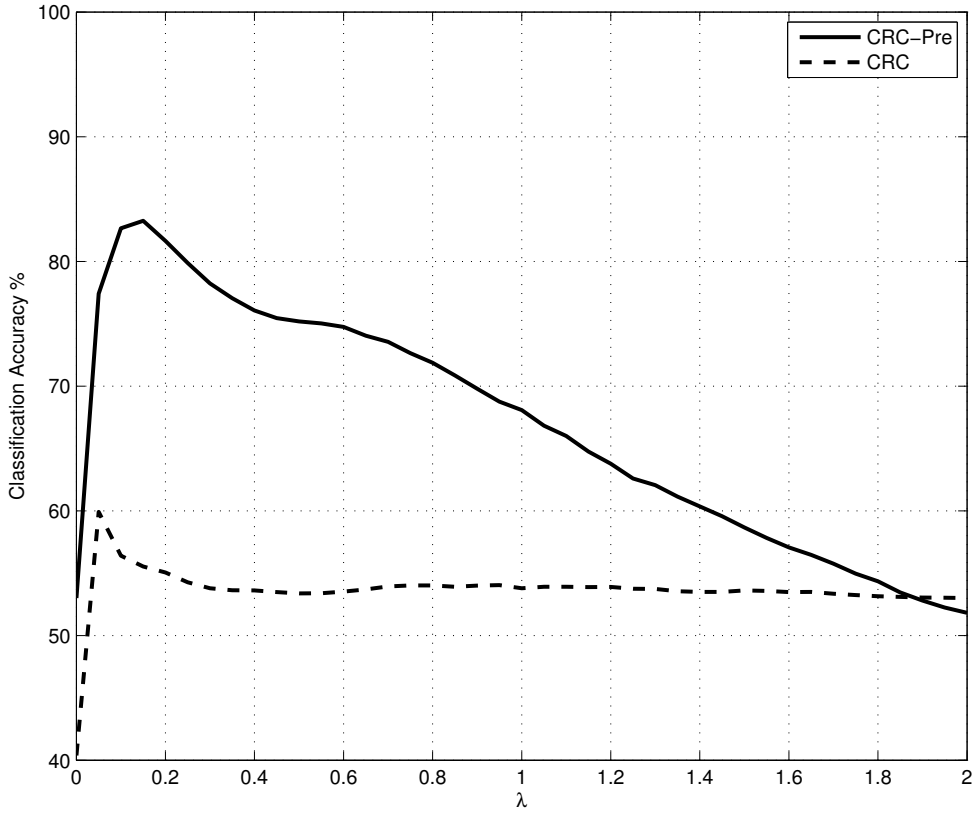


Figure 5.5

Classification accuracy of pre- and post-partitioning (CRC-Pre and CRC, respectively) for the Indian Pines HSI dataset over a range of values for the regularization parameter λ.

seek an approximation of $\mathbf{y}$ for each class, $\widetilde{\mathbf{y}}_l$, calculated only from the training samples particular to class l, $\mathbf{X}_l$. We calculate the per-class coefficients, $\boldsymbol{\alpha}_l$, according to

$$\boldsymbol{\alpha}_l = \arg\min_{\theta} \left\|\mathbf{y} - \mathbf{X}_l\theta\right\|_2^2 + \lambda\left\|\boldsymbol{\Gamma}_{l,y}\theta\right\|_2^2, \tag{5.16}$$

where $\boldsymbol{\Gamma}_{l,y}$ is a biasing Tikhonov matrix specific to each class l and test sample $\mathbf{y}$, and λ is a global regularization parameter which balances the minimization between the residual and regularization terms. Specifically, we use a diagonal $\boldsymbol{\Gamma}_l$ in the form of

$$\boldsymbol{\Gamma}_{l,y} = \begin{bmatrix} \left\|\mathbf{y} - \mathbf{x}_{l,1}\right\|_2 & & 0 \\ & \ddots & \\ 0 & & \left\|\mathbf{y} - \mathbf{x}_{l,n_l}\right\|_2 \end{bmatrix}, \tag{5.17}$$

where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_l}$ are the columns of $\mathbf{X}_l$ for the l^{th} class. According to the minimization defined in (5.16) and the structure of $\boldsymbol{\Gamma}_{l,y}$ given in (5.17), hypotheses which are the most dissimilar to $\mathbf{y}$, in terms of Euclidean distance, should be given much less contribution towards the linear combination than those which are most similar. Using this distance-weighting measure for $\boldsymbol{\Gamma}_{l,y}$ enforces a structural meaning to calculated weights without making as stringent of an assumption as true sparsity. Each testing sample $\widetilde{\mathbf{y}}_l$ can then be calculated in closed form,

$$\widetilde{\mathbf{y}}_l = \mathbf{X}_l(\mathbf{X}_l^T\mathbf{X}_l + \lambda^2\boldsymbol{\Gamma}_{l,y}^T\boldsymbol{\Gamma}_{l,y})^{-1}\mathbf{X}_l^T\mathbf{y}. \tag{5.18}$$

After calculating $\widetilde{\mathbf{y}}_l$ for each class, the class assignment for $\mathbf{y}$ is calculated according to (5.13).

The effect of the ℓ_2-regularization term based on $\boldsymbol{\Gamma}_{l,y}$ is two-fold. First, if the training samples are sufficiently similar in each class, or if a large set of training samples is used

input : Training data $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}$, class labels ω_i, testing sample $\mathbf{y} \in \mathbb{R}^d$, λ
Partition $\mathbf{X}_l$;
for $l \in \{1, 2, \ldots, C\}$ **do**
    Calculate $\mathbf{\Gamma}_{l,y}$ via (5.17);
    Calculate $\widetilde{\mathbf{y}}_l$ via (5.18);
end
Decide $\mathrm{class}(\mathbf{y})$ via (5.13);
output: $\mathrm{class}(\mathbf{y})$

Figure 5.6

Proposed NRS Classifier

($n_l \gg d$), the matrix $\mathbf{X}_l^T\mathbf{X}_l$ will either have poor conditioning or be near-singular. The consequence is that the calculation of its inverse will be inaccurate or impossible, creating a lack of backwards stability in the inverse problem, leading to the calculated weights to be of high variance and to convey little to no meaning. Enforcing the regularization term enforces stability on the problem by effectively inflating the singular values of $\mathbf{X}_l$, improving the conditioning of the problem. Second, the form of the biasing matrix $\mathbf{\Gamma}_{l,y}$ used in the regularization term allows for discrimination between classes. Without this term, it is possible, in certain conditions, for each $\mathbf{X}_l$ to approximate $\mathbf{y}$ with arbitrary accuracy, thus removing any discriminative power from r_l. This situation can be effected by setting $\lambda = 0$, causing (5.16) to become a least-squares (LSQ) problem. As illustrated in Fig. 5.5, a near-zero regularization term destroys the accuracy of the classifier.

Figs. 5.7 and 5.8 show the decision boundaries produced for two synthetic two-dimen–sional datasets using both the proposed NRS as well as SVM classifier using a radial-basis kernel. In both cases, the datasets are not linearly separable and require complex bound-

aries for accurate classification. In Fig. 5.7, both the SVM and NRS classifiers produce a flexible boundary which accurately cuts between the two classes; however, the SVM boundary appears to be a more general fit, with the NRS boundary being much more data dependent. On the other hand, in Fig. 5.8, we see two overlapping classes with shared means. Here, the NRS boundary performs better by cutting much closer to the mean, reducing incorrect classification for samples generated from Class 1 near the mean.

There are several differences between the proposed method and the previously discussed k-NN, SRC, and CRC techniques. Firstly, the NRS classifier, unlike the k-NN classifier, does not limit its classification to the correspondence between testing samples and the provided training data alone. Instead, by forming an approximation from each class, the NRS technique compares the testing sample with what can be considered an imaginary training sample which could have conceivably been drawn from the same process that produced the class training data provided. Secondly, the NRS classifier does not rely on time-consuming iterative sparse-recovery algorithms, as is the case with the SRC and other such sparse techniques for classification. While the recent investigations of sparse regularizations have been of wide interest in signal processing in general, in this area at least, they do not seem to provide significant performance gains to outweigh their computationally expensive implementations. Lastly, while both the NRS and CRC techniques employ Tikhonov regularization to calculate class approximations, the NRS cleaves to the traditional approach of pre-partitioning and calculating class approximations independently. Besides the non-collaborative nature of the NRS, we also propose non-uniform regularization through use of $\mathbf{\Gamma}_{l,y}$. We note that, due to this alteration, a single projec-

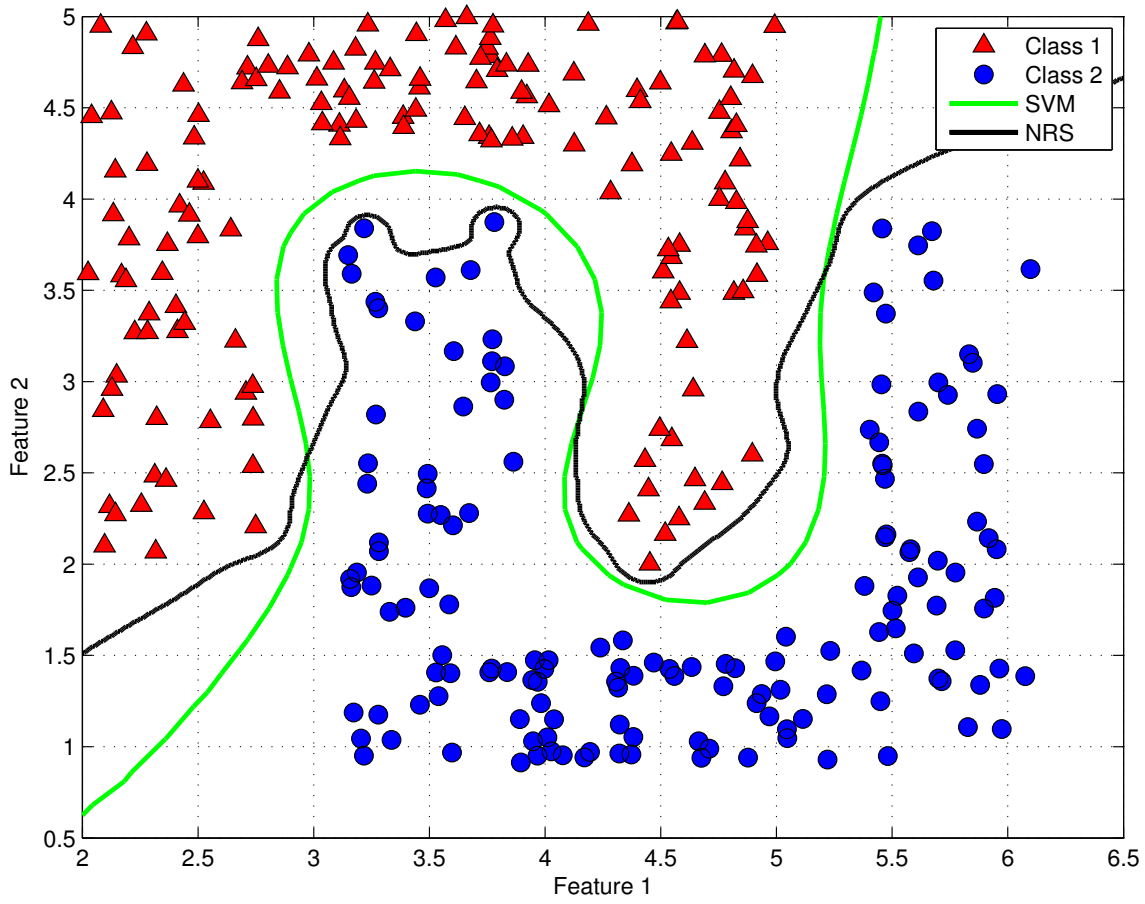


Figure 5.7

Decision boundaries determined for a two-class synthetic dataset.

NRS boundary calculated for $\lambda = 1$. The decision boundary for the SVM classifier using the radial-basis kernel is also shown.

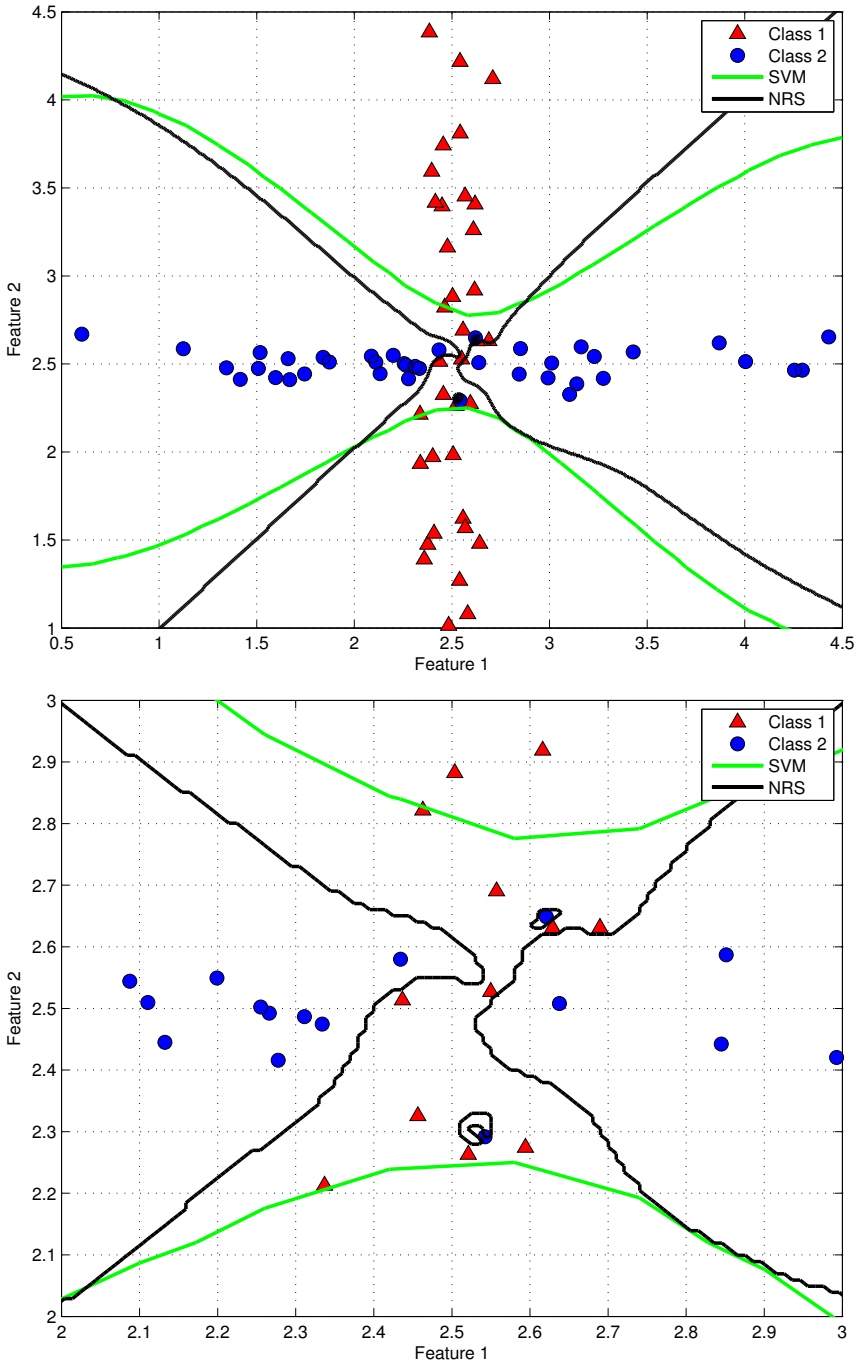


Figure 5.8

Decision boundaries determined for an intersectiong two-class synthetic dataset.

Top: Decision boundaries determined using the NRS classifier for $\lambda = 1$ and the SVM classifier using the radial-basis kernel for two synthetic normally distributed intersecting classes with common mean. Bottom: Closer inspection of the class intersection.

tion operator may not be computed for batch classification tasks when employing the NRS classifier since $\mathbf{\Gamma}_{l,y}$ is specific to each training sample.

When constructing the biasing matrix $\mathbf{\Gamma}_{l,y}$ as in (5.17), we see that only the Euclidean distance between training and test samples is considered. In Sec. 5.4, it is demonstrated that this approach to biasing provides gains in classification accuracy for HSI datasets; however, it is well known that using Euclidean distances for very high-dimensional data can be an exercise in futility for certain data distributions. In the next section, we propose a method to alter the construction of $\mathbf{\Gamma}_{l,y}$ by using a generalized distance measure chosen to maximize class discrimination.

5.3.2 Dynamic Regularization for Classification

From the previous section, we see that the proposed NRS classifier does not estimate or explicitly account for class probability distributions—instead it measures only the ability of each class to approximate a given target sample given a regularization parameter, λ. This regularization parameter is a significant factor in our proposed system, and, in fact, for all regularization-based techniques which make use of weighted-sum penalty functions. From Fig. 5.5, we can see that the setting of this parameter can also greatly affect classification accuracy. Both the SRC and CRC approaches offer little information on how this parameter should be set [24, 122] other than to suggest that cross-validation (CV) approaches could be used—splitting the training set into two parts and testing for a value which maximizes classification accuracy. However, the CV approach might not give an accurate estimation

of the optimal λ when very few training samples are available, or might even be infeasible for extremely small training sets.

We propose to eliminate the need for CV estimations of λ by constructing a classifier which does not require fine tuning of many side variables (for which classifiers such as the SVM are notorious) at the cost of somewhat increased computation. We do this by making the observation that, in the case of classification, we are actually unconcerned with the accuracy of the approximations $\widetilde{\mathbf{y}}_l$; rather, we want just that their proximities to $\mathbf{y}$ are such that they allow us to discriminate the class of $\mathbf{y}$ accurately.

In order to observe the behavior of the NRS classifier with respect to λ, a two-feature synthetic testing environment is considered in Figs. 5.9 and 5.10. For this dataset, all samples exist in only two dimensions, making the visualization of the classifier behavior easier. Three classes of synthetic data randomly drawn from Gaussian distributions are created with a single test sample drawn from one of these three classes. By treating each approximation as a function of λ for a fixed training set and test sample, $\widetilde{\mathbf{y}}_l(\lambda)$, and by varying λ over a range of values (in this case 10^4 to 10^{-10}), a set of approximations over the domain of λ tested, which we term a solution path, is generated for each class.

Looking at the approximation accuracy of the solution paths in Fig. 5.9, an interesting phenomenon becomes apparent. For large values of λ, the regularization term $\left\|\mathbf{\Gamma}_{l,y}\mathbf{w}_l\right\|_2^2$ becomes the dominant term in the cost function of (5.16), and the representations approach the zero vector to minimize this biased norm. However, for small λ, the representations approach to the test sample, $\mathbf{y}$. Between these two modes, an inflection point occurs wherein the solution path rapidly changes direction. For classes whose members best represent $\mathbf{y}$,

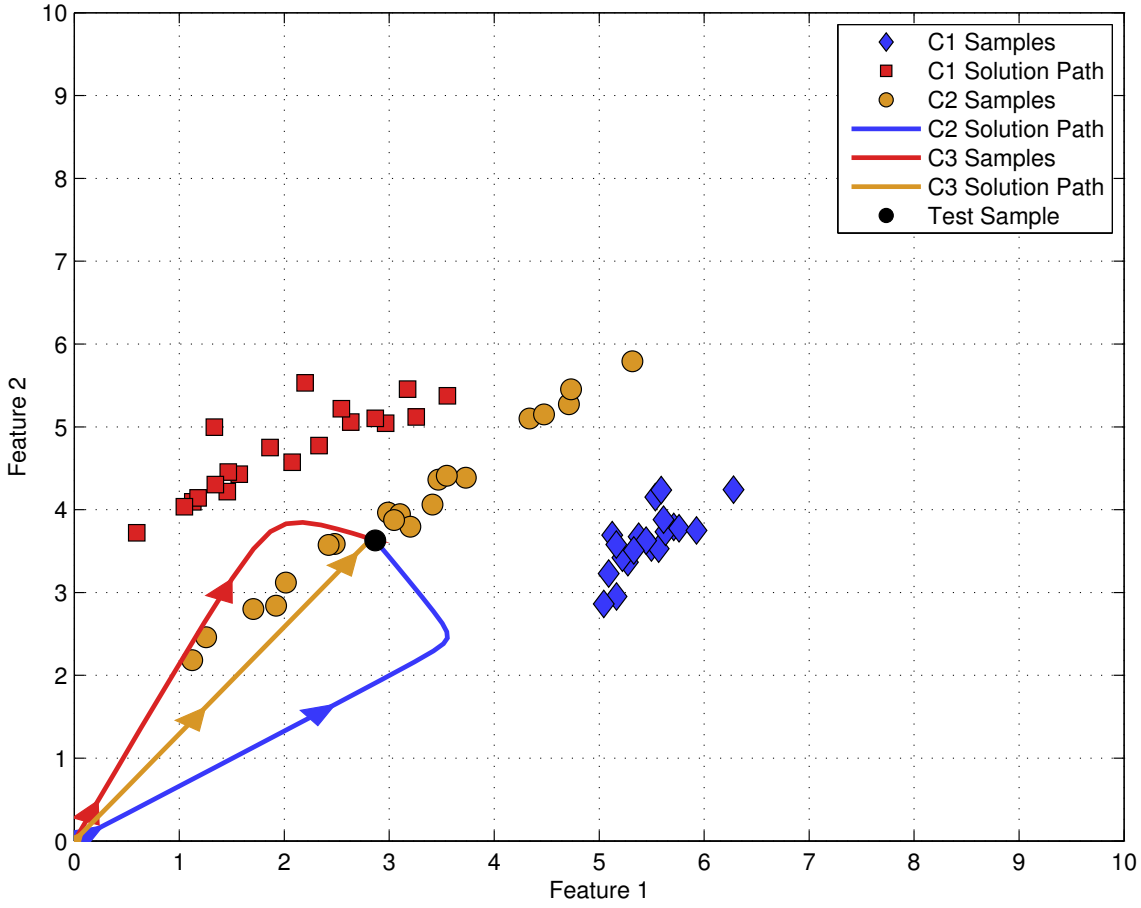


Figure 5.9

Solution paths of the NRS classifier for a synthetic three-class problem in two dimensions for a test sample drawn from class C3.

The 20 training samples per class and the solution paths for each class as λ decreases from 10^4 to 10^{-5}.

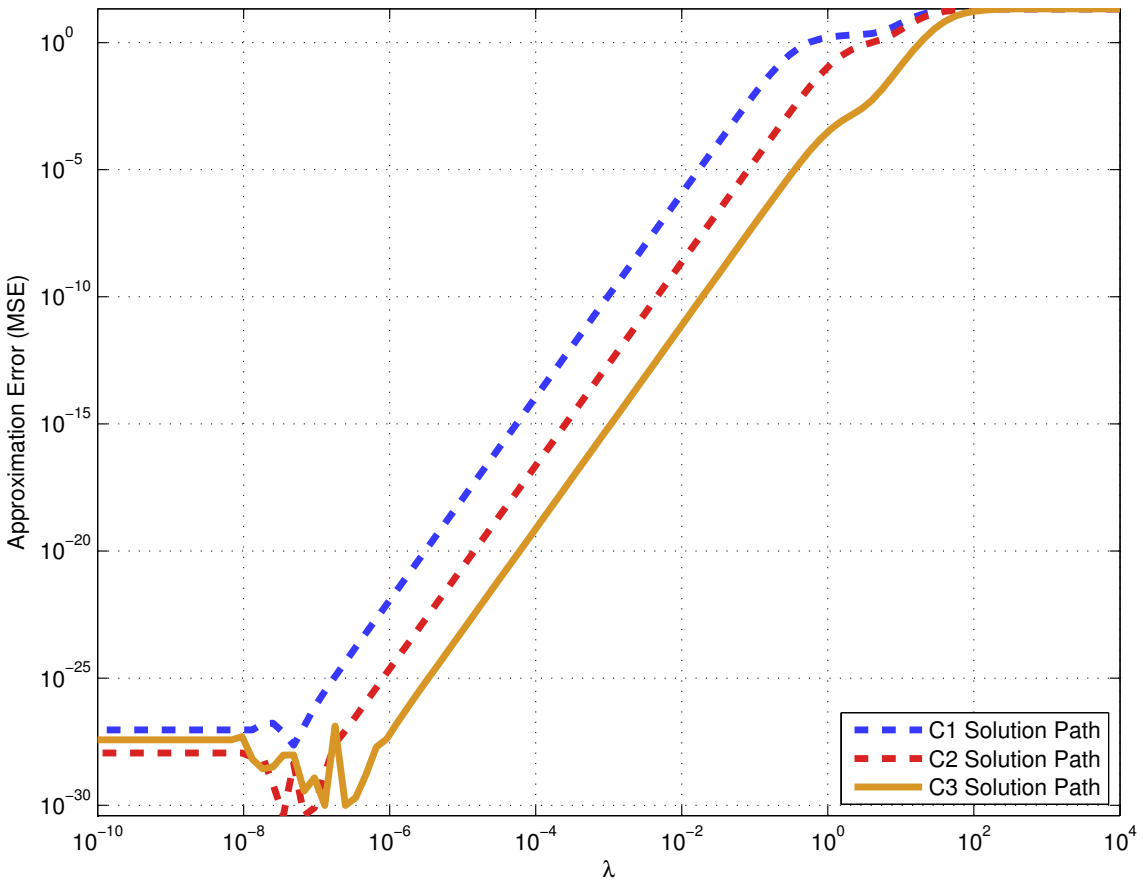


Figure 5.10

Approximation MSE of the NRS classifier for a synthetic three-class problem in two dimensions for a test sample drawn from class C3.

Per-class NRS classifier approximation accuracy. Approximations generated by the true class (C3) are more accurate for $\lambda > 10^{-6}$.

this saddle point is much less pronounced. For classes whose members are most dissimilar, the inflection point is very pronounced, as the "initial" trajectories of these classes are oriented away from $\mathbf{y}$. However, the solution path created by the correct class tends to approach $\mathbf{y}$ much more rapidly, i.e. the approximations for third class, $\widetilde{\mathbf{y}}_3$, are much more accurate for larger values of λ than the approximations generated by the other classes. The rapidity of convergence can be seen in Fig. 5.10.

We propose to use this feature to eliminate the need for setting a fixed value of λ prior to classification. We do this by setting a threshold, ϵ, on the approximation accuracy, $\frac{1}{d}||\widetilde{\mathbf{y}}_l(\lambda) - \mathbf{y}||_2^2$, and determining the classification based upon the first class to pass this threshold as λ is stepped from large to small values, causing the proposed method to resemble a "race" between the classes. From Fig. 5.10, we can see that ϵ is a more robust parameter, as any choice within the range of $[10^{-25}, 10^0]$ would leave the classification unchanged. This is in contrast to the parameter λ, for which, in different test environments, small variations can cause large differences in classification performance. Also, the addition of noise to the dataset can cause the optimal choice for λ to shift away from *a priori* expected values. Instead of indirectly accounting for noise by adjusting λ, an approximation of the noise energy can be used to set ϵ directly. Additionally, if only a small number of training samples are available to drive the classification, the effectiveness of using CV approaches to estimate an optimal fixed setting for λ can be greatly diminished. Also, it is reasonable to assume that not every test sample requires the same value of λ to ensure correct classification. The proposed method accounts for the individuality of each test sample by sidestepping the need for a fixed λ at all, testing each sample's classification across a

range of λ. Together, these features make dynamic regularization more robust than using a fixed λ and ensure stable classifier performance for the practitioner.

5.3.3 Enhancing Discrimination Power

One popular method of enhancing discrimination for hyperspectral classification is through LDA [38]. LDA projects from its natural, perhaps high-dimensional, space into a lower-dimensional subspace via a transform procedure aimed at maximizing between-class scatter while minimizing within-class scatter. Recently, an extension of LDA, locality Fisher's discriminant analysis (LFDA) [97], was proposed. LFDA combines the separability enhancing power of LDA with locality-preserving projections (LPP) [52] to form a transformation, $\mathbf{L}$, which can handle multimodal non-Gaussian class distributions while preserving the local structure of the class distributions in the projected subspace.

In LFDA, we define the *affinity* between $\mathbf{x}_i$ and $\mathbf{x}_j$ as $A_{i,j} = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\gamma_i\gamma_j\right)$, where $\gamma_i = \|\mathbf{x}_i - \mathbf{x}_i^{(k_{\text{nn}})}\|$ denotes the local scaling of data samples in the neighborhood of $\mathbf{x}_i$, and $\mathbf{x}_i^{(k_{\text{nn}})}$ is the k_{nn}-nearest neighbor of $\mathbf{x}_i$. The resulting $\mathbf{A}$ is a symmetric matrix of size $n \times n$, which measures the distance among data samples. In fact, the *local* between-class $\mathbf{S}^{(\text{lb})}$ and within-class $\mathbf{S}^{(\text{lw})}$ scatter matrices of LFDA are the traditional LDA scatter matrices $\mathbf{S}^{(\text{b})}$ and $\mathbf{S}^{(\text{w})}$ scaled appropriately via the affinity matrix $\mathbf{A}$ (see [65]). This weight assignment provides an important benefit to the traditional LDA formulation—if a class-conditional probability distribution function is multi-modal, different modes will contribute to the scatter independently, thereby resulting in a more accurate representation of multi-modal data. This important neighborhood-preserving property ensures that local

neighborhood relationships in the original space are retained in the projected subspace. The LFDA obtains good between-class separation while preserving the within-class local structure simultaneously. The modified Fisher's ratio in LFDA employs these local scatter matrices to estimate the dimensionality-reduction projection as the solution, $\mathbf{L}$, to generalized eigenvalue problem, $\mathbf{S}^{(\text{lb})}\mathbf{L} = \mathbf{\Lambda}\mathbf{S}^{(\text{lw})}\mathbf{L}$. The reader is referred to [65, 97] for more details on LFDA.

In this work, we define a generalized distance measure by comparing the distances between points within the projection space of $\mathbf{L}$, namely,

$$\begin{aligned} D_{LFDA}(\mathbf{x}, \mathbf{y}) &= ||\mathbf{L}\mathbf{x} - \mathbf{L}\mathbf{y}||_2, \\ &= \sqrt{(\mathbf{L}\mathbf{x} - \mathbf{L}\mathbf{y})^T (\mathbf{L}\mathbf{x} - \mathbf{L}\mathbf{y})}, \\ &= \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{M} (\mathbf{x} - \mathbf{y})}, \end{aligned} \tag{5.19}$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors of $d \times 1$, $\mathbf{L}$ is the transformation matrix with size of $d' \times d$ (d' is the reduced dimensionality), $\mathbf{M} = \mathbf{L}^T\mathbf{L}$ is a symmetric positive matrix, and $D_{LFDA}(\mathbf{x}, \mathbf{y})$ is a single scalar. Using (5.19), we modify the construction of the biasing Tikhonov matrix of (5.17) to become

$$\mathbf{\Gamma}_{l,y} = \begin{bmatrix} D_{LFDA}(\mathbf{y}, \mathbf{x}_{l,1}) & & 0 \\ & \ddots & \\ 0 & & D_{LFDA}(\mathbf{y}, \mathbf{x}_{l,n_l}) \end{bmatrix}. \tag{5.20}$$

We refer to the classifier using this construction of $\mathbf{\Gamma}_{l,y}$ as NRS-LFDA. By comparing distance relationships within the LFDA-projected space, we gain two distinct advantages when biasing our Tikhonov regularization of (5.16). First, by reducing the dimensionality

of the space in which distances are calculated, distances become more meaningful to the classification task, rather than having all distances being large. Second, the space is chosen in such a manner that inter-class separability is increased, further penalizing classes whose memberships lie mostly distant from the target point. Additionally, the LPP of LFDA means that samples which are truly neighbors of $\mathbf{y}$ are also seen as neighbors within the projected space. Without such locality preservation, calculating distances within a lower-dimensional space (such as that produced by LDA) might not give any information on within-class distance relationships with $\mathbf{y}$ and might offer little benefit in terms of classification accuracy. In the next section, we present results which demonstrate that the NRS-LFDA technique presented here does indeed improve classification accuracy as compared to the original NRS which uses Euclidean distances in the original space.

5.4 Experiments

5.4.1 Experimental Hyperspectral Data

In this section, we demonstrate the effectiveness of the both proposed NRS and NRS-LFDA classifiers on HSI datasets. The first HSI dataset in our tests was acquired using NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor and was collected over northwest Indiana's Indian Pines test site in June 1992[1]. The image represents a vegetation-classification scenario with 145×145 pixels and 220 spectral bands in the 0.4- to 2.45-μm region of the visible and infrared spectrum with a spatial resolution of 20 m. The two main crops, soybean and corn, shown in the HSI are in their early growth stage. The notation *no till*, *min till*, and *clean till* indicate the amount of previous crop residue

[1] `ftp://ftp.ecn.purdue.edu/biehl/MultiSpec`

remaining. There are 16 different land-cover classes in original ground truth; however, we conduct our experiments with eight classes, allowing for more training samples from a statistical viewpoint [73]. Approximately 8600 labeled pixels are employed to train and validate the efficacy of the proposed classification methods. This data is partitioned into approximately 1496 training pixels and 7102 testing pixels.

The other two HSI datasets used in this work were collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor [43]. The image, covering the city of Pavia, Italy, was collected under the HySens project managed by DLR (the German Aerospace Agency). The images have 115 spectral bands with a spectral coverage from 0.43- to 0.86-μm, and a spatial resolution of 1.3 m. Two scenes are used in our experiment. The first one of these is the university area which has 103 spectral bands with a spatial coverage of 610×340 pixels. The second one is the Pavia city center which has 102 spectral bands with 1096×715 pixels formed by combining two separate images representing different areas of the Pavia city. The numbers of training and testing samples used for the University of Pavia data set are 1476 and 7380, respectively. The numbers of training and testing samples used for the Pavia Centre data set are 1477 and 8862, respectively.

5.4.2 Experiments

We compare our proposed methods with k-NN, SRC[2], CRC-Pre, SVM, and the recently proposed LFDA-SVM [65] classifiers. For the k-NN classifier, we find that $k = 3$ usually provides better classification performance compared to other values (such as 1, 5,

[2] ℓ_1-minimization is implemented by `l1_ls.m` from `http://www.stanford.edu/~boyd/software.html`

7, etc.). For SRC, we chose the parameter $\lambda = 0.01$ in our experiments. For CRC-Pre, the optimal parameter λ is 0.2 for the Indian Pines dataset, 0.25 for the University of Pavia dataset, and 0.6 for the Pavia Centre dataset. The optimal parameters for SVM and LFDA-SVM can be found in [65]. For NRS-LFDA, the dimensionality of LFDA is around 10 for experimental datasets, and we found it is not sensitive to sample size. Additionally, for both NRS and NRS-LFDA, a threshold of $\epsilon = 10^{-3}$ was used. In practical situations, the number of available training samples is often insufficient for each class. We illustrate the sensitivity of each classifier to the number of available training samples by testing over different percentages of the dataset used for training while retaining the prior probability of each class. To avoid any bias, we randomly choose a subset of training samples for each sample-size value and repeat the experiment 10 times, reporting the average classification accuracy.

It is obvious from Fig. 5.11 that the proposed methods—the NRS and NRS-LFDA classifiers—outperform other approaches, especially under the small training-size classification scenario. The k-NN classifier has the worst classification accuracy, while SVM does not perform as well as either CRC-Pre or SRC do for the cases of small-training-samples-size. It is worthwhile mentioning that NRS-LFDA classifier has on average 3% better accuracy than the NRS classifier and even greater improvements in accuracy over the other tested classifiers, which verifies that the discriminant enhancing LFDA distance metric works well for hyperspectral data. Figs. 5.12–5.13 show the overall accuracy as a function of number of training samples for the University of Pavia and Pavia Centre datasets, respectively. For these two Pavia datasets, SRC and CRC-Pre have unfavorable classifica-

tion accuracies, even lower than k-NN. The proposed NRS-LFDA and NRS classifiers still provide the best classification accuracy of the tested classifiers for these datasets.

Fig. 5.14 provides a visual inspection of the classification maps generated using the whole HSI scene for the Indian Pines dataset (145×145, including unlabeled pixels). To facilitate easy comparison between classification methods, only areas for which we have ground truth are shown in these maps. In Fig. 5.14, our proposed techniques show the best spatial homogeneity of the tested approaches. This homogeneity is most pronounced within the "Soybean-min till" and "Soybean-clean till" areas.

Finally, we compare the computational complexity of the classification methods. All the experiments are carried out using MATLAB on a 3.2-GHz machine with 5.8 GB of RAM. As an example, the execution times (in seconds) to train and validate with the Indian Pines dataset is shown in Table 5.1. We find that the NRS classifier generally runs around 15 times slower than CRC-Pre, but around 10 times faster than SRC. Notice that both CRC-Pre and SRC require either prior information on the optimal parameter λ, or for a CV approach to be used to estimate this parameter. However, the NRS and NRS-LFDA classifiers do not require such fine tuning. If we were to provide the optimal λ for them, the execution time decreases accordingly (NRS: 135 s, NRS-LFDA: 346 s).

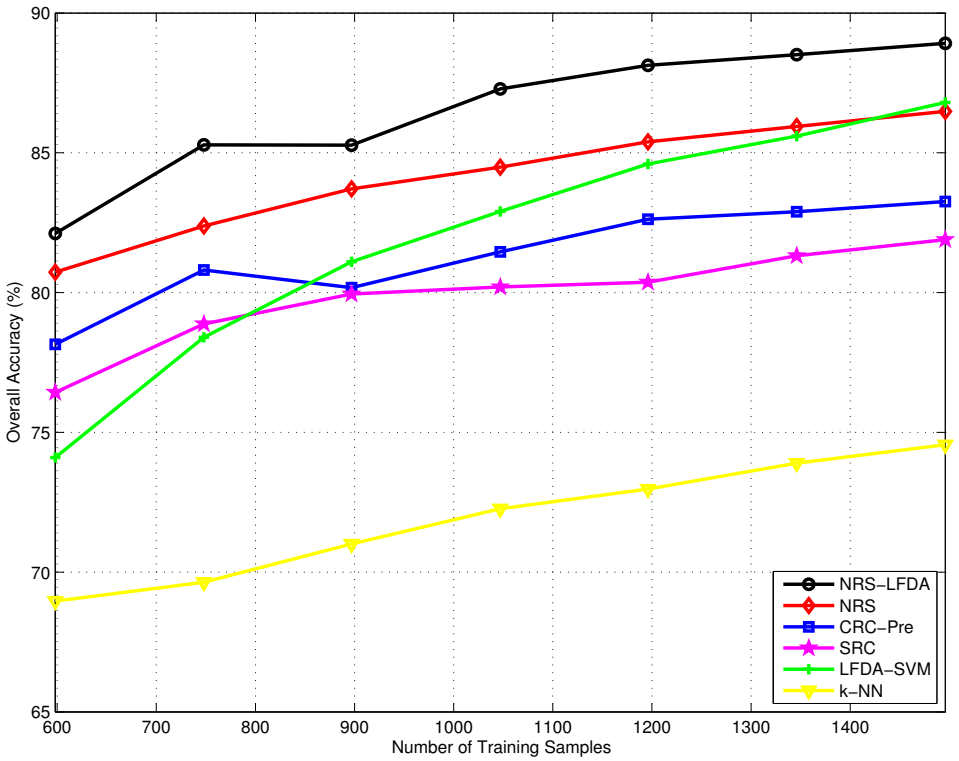


Figure 5.11

Classification accuracy versus the number of training samples for the Indian Pines dataset.

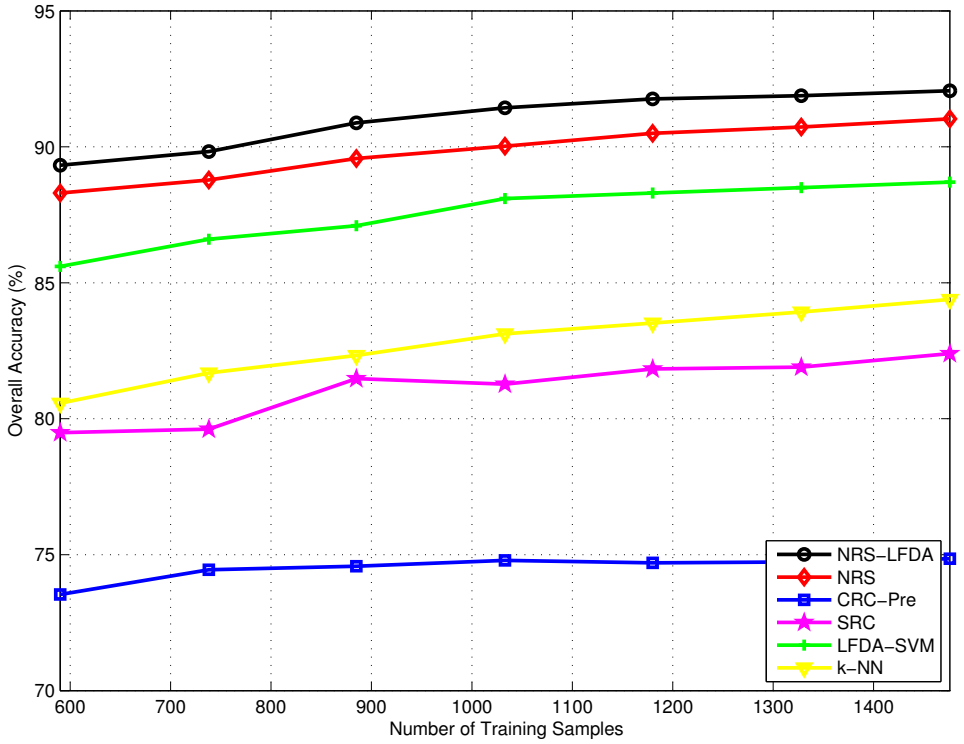


Figure 5.12

Classification accuracy versus the number of training samples for the University of Pavia dataset.

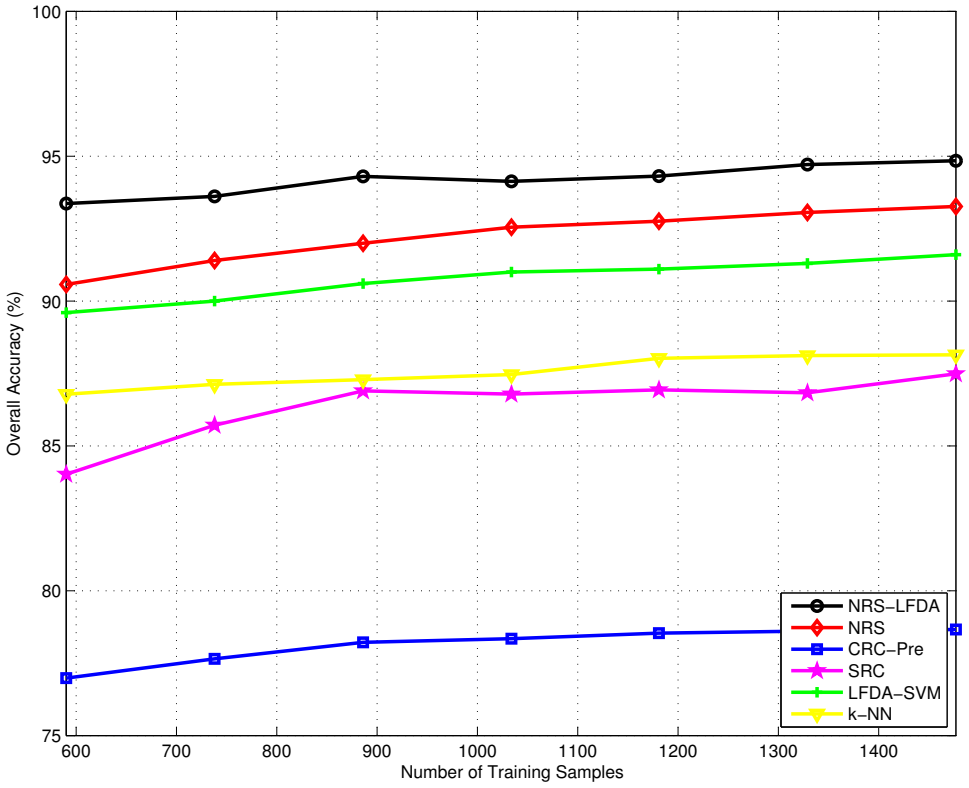
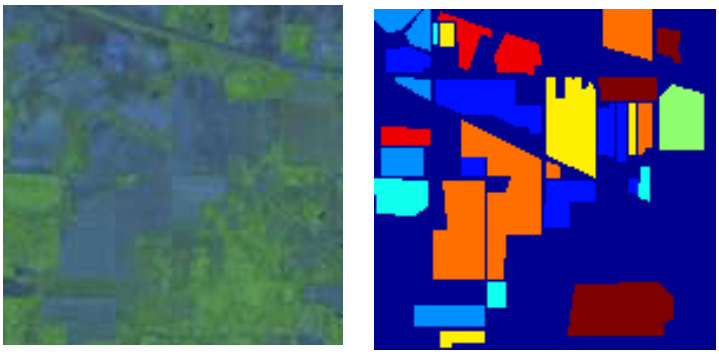
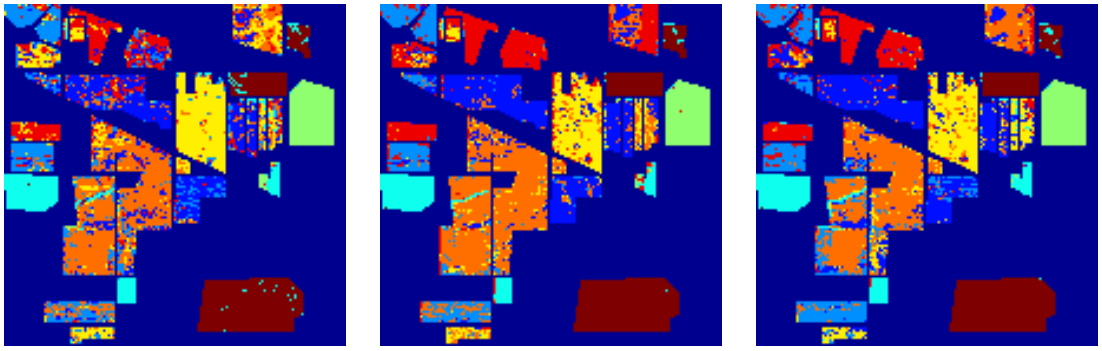


Figure 5.13

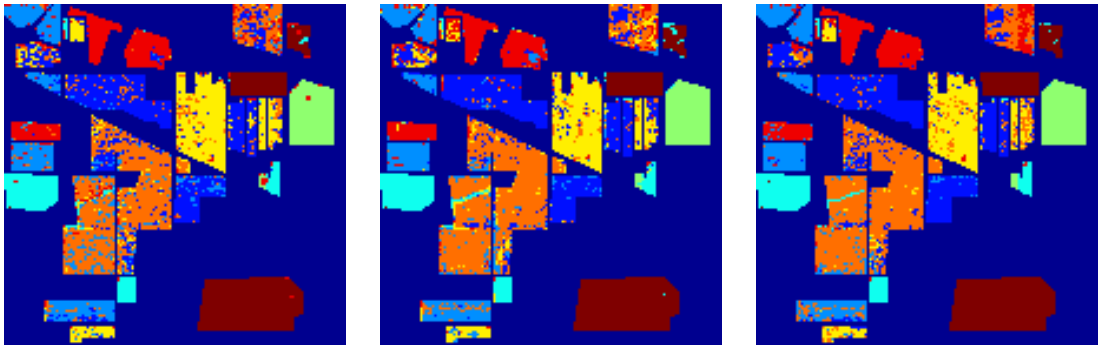
Classification accuracy versus the number of training samples for the Pavia Centre dataset.



(a) False Color Image (b) Ground Truth



(c) k-NN (d) CRC-Pre (e) SRC



(c) LFDA-SVM (d) NRS (e) NRS-LFDA

Corn-no till	Corn-min till	Grass/Pasture	Hay-windowed
Soybean-no till	Soybean-min till	Soybean-clean till	Woods

Figure 5.14

Thematic maps resulting from classification using 748 training samples for the *Indian Pines* HSI dataset.

Table 5.1

Execution time (in seconds) to train and validate with the Indian Pines dataset using 748 samples for training and the whole scene for testing.

Algorithm	*Time (s)*
k-NN	24
CRC-Pre	132
NRS	2210
SVM	5364
LFDA-SVM	5367
NRS-LFDA	9633
SRC	23245

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

In this dissertation, we investigated how best to approach video in terms of CS acquisition and recovery. We researched different methods of recovering video signals from CS acquired measurements, such as frame by frame, volumetric, and residual recovery. From these experiments, we found residual recovery to provide superior distortion performance as compared to the two other approaches. In order to employ residual recovery in practical way, a method of creating accurate predictions from only measurement data was needed. We sought to do this without changing any of the CS imaging hardware suggested in Sec. 2.3.1.

We showed that using a block-match method within the projected, or measurement, domain served as the best method for creating frame predictions. We extended this method from a SH case to a MH one. We further investigated MH prediction by researching different methods of finding hypothesis weighting vectors using both ℓ_1 and Tikhonov regularization. We found that our proposed method of Tikhonov regularization created predictions which significantly enhanced the performance of the residual recovery in relation to the ℓ_1-regularized MH method, SH prediction, and frame-by-frame recovery.

However, in this work, we did not investigate tuning the Tikhonov regularization for finding the MH weights. This regularization could be optimized on a block-by-block basis based upon statistics such as the ratio of maximum and minimum singular values of the hypothesis matrix at each block, H_b. Also, the function used to generate the values on the diagonal of Γ could be investigated further. We used a Euclidean distance, but there is no reason that a different metric might not perform better. We have continued to pursue this topic, but at the time of this writing, this work is not yet complete.

We specifically investigate blocks from temporally neighboring frames serving as hypotheses for a current frame as a form of inter-prediction, but there is also no reason that some form of intra-prediction could not also work well, as spatially neighboring blocks within a frame are also likely to have correlated content. The H.264 video-coding standard allows for intra-predicted blocks and there could be some benefit to investigating their use in the case of CS recovery. This topic was covered in one of our works [21] in the context of still image recovery, but similar intra-prediction strategies and sub-block matching techniques could be employed in video recovery, as well.

Additionally, the use of signal predictions created from side information or from measurements of other highly correlated signals does not need to be limited to only the recovery of video signals. This general concept as applied to a number of different application areas such as bioinformatics or distributed sensor networks could be explored. We have shown in our a number of our works [105, 106, 107] that CS of multiview images is another application to which our techniques are well suited. Since multiview imaging is a specific case of plenoptic, or light-field, imaging for coarse angular resolution, our techniques might

also be adapted for use in the recovery of entire sets of dynamic light-field data acquired using the CS framework.

Concurrent to the writing of this document, a recent work on CS video acquisition and recovery was proposed, termed CS multi-scale video (CS-MUVI) [93]. The CS-MUVI framework eliminates the need for frame-based video acquisition, instead opting for a model of sampling in which each measurement is treated as sampling of dynamic scene content at a different point in time. In this manner, video may be recovered at variable frame rates determined at recovery time, rather than a frame rate fixed by the encoding device. Further work may be done to determine the effectiveness of block-based sampling schemes within this environment. Additionally, the method of [93] might be extended according to the bootstrap methods of [106, 107].

For the NRS classifier, while we demonstrate classification performance competitive with state-of-the-art HSI classification techniques, there are still a number of extensions which could be made to the work. Firstly, the classifier can be extended to compensate for corrupted, mislabeled, or otherwise inaccurate training samples. One approach to this problem might be to use redundant DWT (RDWT) coefficients of different scales as independent classifications and then use these scale classifications to derive a collaborative final classification as is done in [87]. Additionally, alternate methods of determining the parameters of $\mathbf{\Gamma}_{l,y}$ can be investigated to enhance classification accuracy by enhancing the discriminatory power of the NRS classifier.

The NRS classifier also need not be applied only to the task of HSI classification. It might be further adapted to handle other high-dimensionality supervised classification

tasks when very few samples are available for training. Face recognition is one such task which the NRS might be well suited to. The NRS technique might also be advantageous for unsupervised clustering and semi-supervised classification tasks.

The topics covered in this dissertation show many open doors in emerging areas of high-dimensional signal processing. CS as a field, though having matured somewhat, still leaves many questions unanswered in terms of its ultimate applicability to real-life sensing systems. However, CS has already been shown to be immediately applicable in areas such as medical imaging, significantly decreasing MRI and CT scan times. CS has also been shown to be useful in wideband spectrum sensing [75, 74] and even astronomical imaging [9], as evidenced by the CS experiments conducted by the *Herschel* space telescope. In terms of CS imaging, some companies, such as InView in Austin, TX, are actively investigating the SPC framework for hyperspectral [91] and shortwave infrared (SWIR) imaging. These developments point towards the continuing growth of CS as a dynamic and fruitful area of research in signal processing.

REFERENCES

[1] D. Achlioptas, "Database-Friendly Random Projections: Johnson-Lindenstrauss With Binary Coins," *Journal of Computer and System Science*, vol. 66, no. 4, June 2003, pp. 671–687.

[2] R. Archibald and G. Fann, "Feature Selection and Classification of Hyperspectral Images With Support Vector Machines," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, October 2007, pp. 674–677.

[3] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of Hyperspectral Images with Regularized Linear Discriminant Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, March 2009, pp. 862–873.

[4] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-Based Compressive Sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, April 2010, pp. 1982–2001.

[5] R. G. Baraniuk, V. Cevher, and M. B. Wakin, "Low-Dimensional Models for Dimensionality Reduction and Signal Recovery: A Geometric Perspective," *Proceedings of the IEEE*, vol. 98, no. 6, June 2010, pp. 959–971.

[6] R. G. Baraniuk, K. F. Kelly, S. Krishna, and R. F. Bridge, "Compressive sensing architecture advances infrared camera design," *Laser Focus World*, vol. 47, no. 6, June 2011.

[7] 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, December 2007, pp. 2992–3004.

[8] T. Blumensath and M. E. Davies, "Iterative Thresholding for Sparse Approximations," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, December 2008, pp. 629–654.

[9] J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed Sensing in Astronomy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, Oct. 2008, pp. 718 –726.

[10] E. Candès and J. Romberg, "Signal Recovery from Random Projections," *Computational Imaging III*, San Jose, CA, March 2005, Proc. SPIE 5674, pp. 76–86.

[11] E. Candès, 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, February 2006, pp. 489–509.

[12] E. Candès, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, August 2006, pp. 1207–1223.

[13] E. Candès and T. Tao, "Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, December 2006, pp. 5406–5425.

[14] E. J. Candès, "Compressive Sampling," *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, August 2006, vol. 3, pp. 1433–1452.

[15] E. J. Candès, "The Restricted Isometry Property and Its Implications for Compressed Sensing," *Comptes Rendus de l'Académie des Sciences*, Paris, France, 2008, pp. 589–592.

[16] E. J. Candès and T. Tao, "Decoding by Linear Programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, December 2005, pp. 4203–4215.

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

[18] A. Chambolle and P.-L. Lions, "Image Recovery via Total Variation Minimization and Related Problems," *Numerische Mathematik*, vol. 76, no. 2, April 1997, pp. 168–188.

[19] T. F. Chan, S. Esedoglu, F. Park, and A. Yip, *Recent Developments in Total Variation Image Reconstruction*, Tech. Rep. CAM Report 05-01, Department of Mathematics, UCLA, 2004.

[20] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A Single-Pixel Terahertz Imaging System Based on Compressive Sensing," *Applied Physics Letters*, vol. 93, no. 12, September 2008, p. 121105.

[21] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-Sensing Recovery of Images and Video Using Multihypothesis Predictions," *Proceedings of the* 45^{th} *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2011, pp. 1193–1198.

[22] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, August 1998, pp. 33–61.

[23] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43, no. 1, March 2001, pp. 129–159.

[24] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, October 2011, pp. 3973–3985.

[25] L. Şendur and I. W. Selesnick, "Bivariate Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, November 2002, pp. 2744–2756.

[26] S. Dasgupta and A. Gupta, "An Elementary Proof of a Theorem of Johnson and Lindenstrauss," *Random Structures and Algorithms*, vol. 22, no. 1, January 2003, pp. 60–65.

[27] I. Daubechies, M. Defrise, and C. De Mol, "An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, November 2004, pp. 1413–1457.

[28] I. Daubechies, R. DeVore, M. Fournasier, and C. S. Güntürk, "Iteratively Reweighted Least Squares Minimization for Sparse Recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, January 2010, pp. 1–38.

[29] S. Di Zenzo, R. Bernstein, S. D. Degloria, and H. C. Kolsky, "Gaussian Maximum Likelihood and Contextual Classification Algorithms for Multicrop Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 25, no. 6, November 1987, pp. 805–814.

[30] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed Compressed Video Sensing," *Proceedings of the International Conference on Image Processing*, Cairo, Egypt, November 2009, pp. 1393–1396.

[31] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity Adaptive Matching Pursuit Algorithm for Practical Compressed Sensing," *Proceedings of the* 42^{th} *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, October 2008, pp. 581–587.

[32] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and Efficient Compressive Sensing Using Structurally Random Matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, January 2012, pp. 139–154.

[33] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, *Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit*, Tech. Rep., Stanford University, 2006.

[34] M. F. Duarte and R. G. Baraniuk, "Kronecker Compressive Sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 2, February 2012, pp. 494–504.

[35] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-Pixel Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, March 2008, pp. 83–91.

[36] M. F. Duarte, C. Hegde, V. Cevher, and R. G. Baraniuk, "Recovery of Compressible Signals in Unions of Subspaces," *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, March 2009, pp. 175–180.

[37] M. F. Duarte, M. Wakin, and R. Baraniuk, "Wavelet-Domain Compressive Signal Reconstruction Using a Hidden Markov Tree Model," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, April 2008, pp. 5137–5140.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd edition, John Wiley & Sons, Inc., New York, 2001.

[39] B. Efron, T. Hastie, and I. Johnstone, "Least Angle Regression," *Annals of Statistics*, vol. 32, no. 2, June 2004, pp. 407–499.

[40] 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 on Selected Areas in Communications*, vol. 1, no. 4, December 2007, pp. 586–597.

[41] J. E. Fowler, S. Mun, and E. W. Tramel, "Multiscale Block Compressed Sensing with Smoothed Projected Landweber Reconstruction," *Proceedings of the European Signal Processing Conference*, Barcelona, Spain, August 2011, pp. 564–568.

[42] J. E. Fowler, S. Mun, and E. W. Tramel, "Block-Based Compressed Sensing of Images and Video," *Foundations and Trends in Signal Processing*, vol. 4, no. 4, March 2012, pp. 297–416.

[43] P. Gamba, "A Collection of Data for Urban Area Characterization," *Proceedings of the International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, September 2004, vol. 1, pp. 69–72.

[44] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed Sensing in Dynamic MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 2, February 2008, pp. 365–373.

[45] L. Gan, "Block Compressed Sensing of Natural Images," *Proceedings of the International Conference on Digital Signal Processing*, Cardiff, UK, July 2007, pp. 403–406.

[46] L. Gan, T. T. Do, and T. D. Tran, "Fast Compressive Imaging Using Scrambled Block Hadamard Ensemble," *Proceedings of the European Signal Processing Conference*, Lausanne, Switzerland, August 2008.

[47] B. Girod, "Motion-Compensating Prediction with Fractional-Pel Accuracy," *IEEE Transactions on Communications*, vol. 41, no. 4, April 1993, pp. 604–612.

[48] D. Goldfarb and W. Yin, "Second-Order Cone Programming Methods for Total Variation-Based Image Restoration," *SIAM Journal on Scientific Computing*, vol. 27, no. 2, 2005, pp. 622–645.

[49] J. Haupt and R. Nowak, "Signal Reconstruction from Noisy Random Projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, September 2006, pp. 4036–4048.

[50] L. He and L. Carin, "Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, September 2009, pp. 3488–3497.

[51] L. He, H. Chen, and L. Carin, "Tree-Structured Compressive Sensing With Variational Bayesian Analysis," *IEEE Signal Processing Letters*, vol. 17, no. 3, March 2010, pp. 233–236.

[52] X. He and P. Niyogi, "Locality Preserving Projections," *Advances in Neural Information Processing System*, S. Thrun, L. Saul, and B. Schölkopf, eds., MIT Press, Cambridge, MA, 2004.

[53] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, vol. 12, no. 1, February 1970, pp. 55–67.

[54] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, June 2008, pp. 2346–2356.

[55] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz Mappings into a Hilbert Space," *Contemporary Mathematics*, vol. 26, 1984, pp. 189–206.

[56] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t FOCUSS: A General Compressed Sensing Framework for High Resolution Dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, January 2009, pp. 103–116.

[57] H. Jung and J. C. Ye, "Motion Estimated and Compensated Compressed Sensing Dynamic Magnetic Resonance Imaging: What We Can Learn from Video Compression Techniques," *Imaging Systems and Technology*, vol. 20, no. 2, June 2010, pp. 81–98.

[58] L.-W. Kang and C.-S. Lu, "Distributed Compressive Video Sensing," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, April 2009, pp. 1169–1172.

[59] Y. Kim, M. S. Nadar, and A. Bilgin, "Compressed Sensing Using a Gaussian Scale Mixtures Model in Wavelet Domain," *Proceedings of the International Conference on Image Processing*, Hong Kong, September 2010, pp. 3365–3368.

[60] N. G. Kingsbury, "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals," *Journal of Applied Computational Harmonic Analysis*, vol. 10, May 2001, pp. 234–253.

[61] J. Laaksonen, *Subspace Classifiers in Recognition of Handwritten Digits*, doctoral dissertation, Helsinki University of Technology, May 1997.

[62] D. A. Landgrebe, "Hyperspectral Image Data Analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, January 2002, pp. 17–28.

[63] C. Li, *An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing*, master's thesis, Rice University, September 2009.

[64] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-Preserving Discriminant Analysis in Kernel-Induced Feature Spaces for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, September 2011, pp. 894–898.

[65] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-Preserving Dimensionality Reduction and Classification for Hyperspectral Image Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, April 2012, pp. 1185–1198.

[66] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest Regularized Subspace for Hyperspectral Classification," *IEEE Transactions on Geoscience and Remote Sensing*, Submitted April 2012. Revised August 2012.

[67] Y. Liu, S. S. Ge, C. Li, and Z. You, "$\boldsymbol{k}$-NS: A Classifier by the Distance to the Nearest Subspace," *IEEE Transactions on Neural Networks*, vol. 22, no. 8, August 2011, pp. 1256–1268.

[68] W. Lu and N. Vaswani, "Modified Compressive Sensing for Real-Time Dynamic MR Imaging," *Proceedings of the International Conference on Image Processing*, Cairo, Egypt, November 2009, pp. 3045–3048.

[69] L. Ma, M. M. Crawford, and Jinwen, "Local Manifold Learning-Based k-Nearest-Neighbor for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, November 2010, pp. 4099–4109.

[70] F. A. Mainji and Y. Zhang, "Robust Hyperspectral Classification Using Relevance Vector Machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, June 2011, pp. 2100–2112.

[71] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy Continuation for Sparse Signal Representation," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, March 2005, vol. 5, pp. 733–736.

[72] R. Marcia and R. Willett, "Compressive Coded Aperture Video Reconstruction," *Proceedings of the European Signal Processing Conference*, Lausanne, Switzerland, August 2008.

[73] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images with Support Vector Machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, August 2004, pp. 1778–1790.

[74] M. Mishali and Y. C. Eldar, "Wideband Spectrum Sensing at Sub-Nyquist Rates," *IEEE Signal Processing Magazine*, vol. 28, no. 4, July 2011, pp. 102–135.

[75] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to Digital at Sub-Nyquist Rates," *IET Circuits, Devices and Systems*, vol. 5, no. 1, January 2011, pp. 8–20.

[76] S. Mun and J. E. Fowler, "Block Compressed Sensing of Images Using Directional Transforms," *Proceedings of the International Conference on Image Processing*, Cairo, Egypt, November 2009, pp. 3021–3024.

[77] S. Mun and J. E. Fowler, "Block Compressed Sensing of Images Using Directional Transforms," *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, eds., Snowbird, UT, March 2010, p. 547.

[78] S. Mun and J. E. Fowler, "Residual Reconstruction for Block-Based Compressed Sensing of Video," *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, eds., Snowbird, UT, March 2011, pp. 183–192.

[79] D. Nedell and R. Vershynin, "Signal Recovery from Inaccurate and Incomplete Measurements via Regularized Orthogonal Matching Pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, April 2010, pp. 310–316.

[80] D. Needell and J. A. Tropp, "CoSAMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, May 2009, pp. 301–321.

[81] G. N. Nielson, R. H. Olsson, P. R. Resnick, and O. B. Spahn, "High-Speed MEMS Micromirror Switching," *Proceedings of the Conference on Lasers and Electro-Optics*, Baltimore, MD, May 2007, pp. 1–2.

[82] S. Nogaki and M. Ohta, "An Overlapped Block Motion Compensation for High Quality Motion Picture Coding," *Proceedings of the IEEE International Symposium on Circuits and Systems*, San Diego, CA, May 1992, vol. 1, pp. 184–187.

[83] E. Oja, *Subspace Methods of Pattern Recognition*, Research Studies Press, Letchworth, England, 1983.

[84] M. T. Orchard and G. J. Sullivan, "Overlapped Block Motion Compensation: An Estimation-Theoretic Approach," *IEEE Transactions on Image Processing*, vol. 3, no. 5, September 1994, pp. 693–699.

[85] J. Y. Park and M. B. Wakin, "A Multiscale Framework for Compressive Sensing of Video," *Proceedings of the Picture Coding Symposium*, Chicago, IL, May 2009.

[86] S. Prasad and L. M. Bruce, "Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, May 2008, pp. 1448–1456.

[87] S. Prasad, W. Li, J. E. Fowler, and L. M. Bruce, "Information Fusion in the Redundant-Wavelet-Tranform Domain for Noise-Robust Hyperspectral Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 9, September 2012, pp. 3474–3486.

[88] C. Qiu, W. Lu, and N. Vaswani, "Real-Time Dynamic MR Image Reconstruction Using Kalman Filtered Compressed Sensing," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, April 2009, pp. 393–396.

[89] R. Rigamonti, M. A. Brown, and V. Lepetit, "Are Sparse Representations Really Relevant for Image Classification?," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 2011, pp. 1545–1552.

[90] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D*, vol. 60, no. 1-4, November 1992, pp. 259–268.

[91] T. A. Russell, L. McMackin, B. Bridge, and R. Baraniuk, "Compressive Hyperspectral Sensor for LWIR Gas Detection," *Compressive Sensing, Proc. SPIE 8365*, F. Ahmad, ed., Baltimore, MD, June 2012, p. 83650C.

[92] L. Samaniego, A. Bárdossy, and K. Schulz, "Supervised Classification of Recovery Sensed Image Using a Modified k-NN Technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, July 2008, pp. 2112–2125.

[93] A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk, "CS-MUVI: Video Compressive Sensing for Spatial-Multiplexing Cameras," *Proceedings of the IEEE International Conference on Computational Photography*, Seattle, WA, April 2012, pp. 1–10.

[94] A. C. Sankaranarayanan, P. K. Turaga, R. G. Baraniuk, and R. Chellappa, "Compressive Acquisition of Dynamic Scenes," *Proceedings of the European Conference on Computer Vision, Part I*, Crete, Greece, September 2010, pp. 129–142.

[95] A. Secker and D. Taubman, "Lifting-based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression," *IEEE Transactions on Image Processing*, vol. 12, no. 12, December 2003, pp. 1530–1542.

[96] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, December 1993, pp. 3445–3462.

[97] M. Sugiyama, "Dimensionaltity Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis," *Journal of Machine Learning Research*, vol. 8, no. 5, May 2007, pp. 1027–1061.

[98] G. J. Sullivan, "Multi-hypothesis Motion Compensation for Low Bit-rate Video Coding," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, MN, April 1993, vol. 5, pp. 437–440.

[99] R. Sundaresan, Y. Kim, M. S. Nadar, and A. Bilgin, "Motion-Compensated Compressed Sensing for Dynamic Imaging," *Applications of Digital Image Processing XXXIII*, A. G. Tescher, ed., San Diego, August 2010, Proc. SPIE 7798, p. 77980A.

[100] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A New Compressive Imaging Camera Architecture Using Optical-Domain Compression," *Computational Imaging IV*, C. A. Bouman, E. L. Miller, and I. Pollak, eds., San Jose, CA, January 2006, Proc. SPIE 6065, p. 606509.

[101] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, August 2009, pp. 2973–2987.

[102] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, 1996, pp. 267–288.

[103] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington, D.C., 1977.

[104] E. W. Tramel and J. E. Fowler, "Video Compressed Sensing with Multihypothesis," *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, eds., Snowbird, UT, March 2011, pp. 193–202.

[105] M. Trocan, T. Maugey, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Compressed Sensing of Multiview Images Using Disparity Compensation," *Proceedings of the International Conference on Image Processing*, Hong Kong, September 2010, pp. 3345–3348.

[106] M. Trocan, T. Maugey, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Multistage Compressed-Sensing Reconstruction of Multiview Images," *Proceedings of the IEEE Workshop on Multimedia Signal Processing*, Saint-Malo, France, October 2010, pp. 111–115.

[107] M. Trocan, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Compressed-Sensing Recovery of Multiview Image and Video Sequences using Disparity Compensation," *Multimedia Tools and Applications*, 2012, Submitted March 2012, under revision.

[108] J. Tropp and A. Gilbert, "Signal Recovery from Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, December 2007, pp. 4655–4666.

[109] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning Relevant Image Features With Multiple-Kernel Classifications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, October 2010, pp. 3780–3791.

[110] N. Vaswani, "Kalman Filtered Compressed Sensing," *Proceedings of the International Conference on Image Processing*, San Diego, CA, October 2008, pp. 893–896.

[111] N. Vaswani, "LS-CS-Residual (LS-CS): Compressive Sensing on Least Squares Residual," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, August 2010, pp. 4108–4120.

[112] N. Vaswani and W. Lu, "Modified-CS: Modifiying Compressive Sensing for Problems with Partially Known Support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, September 2010, pp. 4595–4607.

[113] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An Architecture for Compressive Imaging," *Proceedings of the International Conference on Image Processing*, Atlanta, GA, October 2006, pp. 1273–1276.

[114] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressive Imaging for Video Representation and Coding," *Proceedings of the Picture Coding Symposium*, Beijing, China, April 2006.

[115] M. B. Wakin, J. Y. Park, H. L. Yap, and C. J. Rozell, "Concentration of Measure for Block Diagonal Measurement Matrices," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Dallas, TX, March 2010, pp. 3614–3617.

[116] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A New Alternating Minimization Algorithm for Total Variation Image Reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, 2008, pp. 248–272.

[117] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Journal of Machine Learning Research*, vol. 10, February 2009, pp. 207–244.

[118] T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, February 1999, pp. 70–84.

[119] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, February 2009, pp. 210–227.

[120] X. Wu, X. Zhang, and J. Wang, "Model-Guided Adaptive Recovery of Compressive Sensing," *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, eds., Snowbird, UT, March 2009, pp. 123–132.

[121] W. Yin, S. Osher, J. Darbon, and D. Goldfarb, "Bregman Iterative Algorithms for Compressed Sensing and Related Problems," *SIAM Journal on Imaging Scienes*, vol. 1, no. 1, 2008, pp. 143–168.

[122] L. Zhang, M. Yang, and X. Feng, "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?," *Proceedings of the International Conference on Computer Vision*, Barcelona, Spain, November 2011, pp. 471–478.