Recovery of Compressive Sensed Images With Piecewise Autoregressive Modeling

RECOVERY OF COMPRESSIVE SENSED IMAGES WITH PIECEWISE AUTOREGRESSIVE MODELING

BY

REZA POURNAGHI, B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2009)	McMaster University
(Electrical & Computer Engineering)	Hamilton, Ontario, Canada

TITLE:	Recovery of Compressive Sensed Images With Piecewise Autoregressive Modeling
AUTHOR:	Reza Pournaghi B.Sc., (Information Technology Engineering) Sharif University of Technology, Tehran, Iran
SUPERVISOR:	Dr. Xiaolin Wu
NUMBER OF PAGES:	xi, 75

To my family, Farideh, Parviz and Mona who always offered me unconditional love and support.

To my best friend Azin without whose caring support it would not have been possible.

Abstract

This thesis concerns with recovery of compressive sampled images. Since many natural signals such as images are non-stationary, the sparse space varies in time/spatial domain. Therefore, compressive sensing (CS) recovery should be carried on locally adaptive, signal-dependent spaces to answer the fact that the CS measurements are not dependant to the signal structures. Existing CS reconstruction algorithms use fixed basis such as wavelets and DCT for the signals. To address this problem, we proposed new technique for model guided adaptive recovery of compressive sensing. The proposed algorithms are based on two dimensional piecewise autoregressive model and can adaptively recover compressive sampled images. In addition, proposed algorithms offer a powerful mechanism to characterize structured sparsity of natural images. This mechanism greatly restricts the CS solution space. Simulation results show the preeminent effect of our algorithms in the recovery of wide range of natural images. In average our best algorithm improves the reconstruction quality of existing CS methods by $2dB$.

Acknowledgements

I would like to express my great gratitude to my supervisor, Dr. Xiaolin Wu for his support and encouragement. This thesis would not have been possible without his guidance, encouragement and patience. I would also like to thank the examiner committee members, Dr. Dumitrescu and Dr. Zhang for their help and time that they put on reviewing this thesis.

My appreciation goes to my colleagues Navid Shahdi, Amin Behnad, Navid Samavati, AliReza Shoa, Mohammadreza Dadkhah, Sahar Alipour, Mahdy Nabaee, Xiangjun, Ying, Mingkai and all my friends who supported me and made my time a lot more fun. Also, I would like to acknowledge Cheryl, Helen, Cosmin and Terry for their friendly assistance and expert technical support.

Notation and abbreviations

CS	Compressive Sensing
TV	Total Variation
NEAR	Nonlinear Estimation for Adaptive Restoration
PAR	Piecewise Autoregressive
AR	Autoregressive
ARMA	Autoregressive Moving Average
DMD	Digital Micro-mirror Device
RNG	Random Number Generator
CCD	Charge-Coupled Device
CMOS	Complementary MetalOxideSemiconductor
WSN	Wireless sensor networks
FFT	Fast Fourier Transform
RMF	Random Markov Field
ADC	Analog to Digital Converter
MRI	magnetic resonance imaging
RIP	Restricted Isometry Property
OMP	Orthogonal Matching Pursuit
HHS	Heavy Hitters on Steroids

IID	Independent and Identically Distributed
PDF	Probability Distribution Function
STLS	Structured Total Least Squares
MARX	Model-based Adaptive Recovery of Compressive Sensing
SOCP	Second Order Cone Problem
EMARX	Extended Model-based Adaptive Recovery of Compressive Sensing
DFT	Discrete Fourier Transform
PSNR	Peak Signal-to-Noise Ratio

Contents

List of Figures

Chapter 1

Introduction and Problem Statement

In our modern information technology era digital images and videos have become ubiquitous in all walks of the life. Image data acquisition, storage, and communication are indispensable for all applications of digital images. In the current state of the art, digital image acquisition devices (e.g., digital cameras, camcorders, scanners) typically consist of two cascade modules:

1. a dense sensor array that takes a large number of samples (pixels) of the light field.
2. a compressor that reduces the large amount of raw data created by the sensor array.

According to Shannon's sampling theorem, the reconstruction of the original continuous image signal requires the sampling frequency to be at least twice as high as the highest frequency component of the image. In many high-end applications, such as medical imaging, remote sensing, space, reconnaissance, digital cinema, etc., users need to image very fine details and hence demand imaging devices of very high resolution. But the cost and complexity of the camera system increase drastically in spatial resolution. Furthermore, the amount of energy required to acquire and compress images also increases in spatial resolution. In some applications scenarios, such as on board of a satellite in outer space, the camera system is limited in computing power and energy supply. Under such resource constraints one naturally wonders if a mathematical approach exists that allows high-quality recovery of images with a far fewer number of samples.

To answer this question, we first notice a well-known fact in the field of image/video compression. A typical image can be compressed to 10% of the size of the original raw format of two-dimensional sample array without inducing perceptible loss of quality. This indicates a high amount of data redundancy in the dense sample array. Given such high degree of redundancy, why should we adopt the approach of oversampling followed by massive dumping, i.e., compression, as in current practice? This approach makes the imaging devices complex, expensive and with high energy consumption. Recently, the wisdom of above outlined conventional signal acquisition systems is questioned by researchers in a new field of study in signal processing, called compressive sensing (CS). The research on compressive sensing was pioneered by Candes and Donoho [1, 2]. It addresses the inefficiency of the conventional sample-then-process technique by introducing a new data acquisition method that captures and compresses data simultaneously. The CS theory claims that under certain conditions, it can reconstruct a signal with high probability from a small number of random measurements. The conditions on which CS relies are signal sparsity and incoherentness of the random measurements to the original signal.

One can represent any signal $\mathbf{x} \in \mathbb{R}^N$ in terms of a basis $\{\psi_i\}_{i=1}^N$ where each ψ_i is a vector of length N. Defining basis matrix $\mathbf{\Psi} = [\psi_1|\psi_2|...|\psi_N]$ with vectors ψ_i as columns of $\mathbf{\Psi}$, the signal $\mathbf{x}$ can be represented in this basis as

$$\mathbf{x} = \mathbf{\Psi}\mathbf{s}$$

where $\mathbf{s}$ is the coefficient vector of length N.

A signal $\mathbf{x} = \{x_i\}_{i=1}^N$ of length N is said to be sparse in basis $\mathbf{\Psi} = \{\psi_i\}_{i=1}^N$, if the transform coefficients $\langle \mathbf{x}, \psi_i \rangle, 1 \leq i \leq N$, are mostly zero. Incoherence means that sampled sparse signal should have spread out representation in the domain in which it is acquired. Putting differently, if $\mathbf{\Phi}$ is defined as the random measurement matrix used to acquire the signal, then the rows $\{\phi_j\}$ of $\mathbf{\Phi}$ should have an extremely dense representation in $\mathbf{\Psi}$. CS can exactly recover a K-sparse signal of length N, which is a signal with exactly K nonzero coefficients in $\mathbf{\Psi}$, from $M = O(K \log{(N/K)})$ measurements with high probability.

CS recovers signal $\mathbf{x}$ of length N from M measurements, $\mathbf{y}$, by solving the following constrained optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{\Psi}^T\mathbf{x}\|_{l_1} \quad \text{such that} \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{1.1}$$

where $\mathbf{\Phi}$ is the random measurement matrix of size $M \times N$ and $\mathbf{\Psi}$ is the basis in which signal $\mathbf{x}$ has sparse representation. One of the most challenging research topic in compressive sensing is the design of recovery algorithm. Different CS recovery algorithms were recently proposed: gradient projection sparse reconstruction [3], matching pursuit [4], and iterative thresholding [5].

The outstanding property of CS is that it can compactly encode a signal $\mathbf{x}$ independent of the structure of the signal, i.e., CS can use the same random measurement matrix $\mathbf{\Phi}$ on all signals which might have different characteristics. On the other hand, CS recovery process should be optimized for a specific type of signals to recover the signal $\mathbf{x}$ with higher quality. Indeed, the tricky part of recovery step of CS is the selection of the space $\mathbf{\Psi}$ in which a particular signal $\mathbf{x}$ has sparse representation. Thus, CS transfers the task of signal-dependent code optimization from the encoder to the decoder. For optimal CS recovery, finding the space $\mathbf{\Psi}$ in which signal $\mathbf{x}$ has sparse representation is as challenging as finding an adaptive transform to completely decorrelate $\mathbf{x}$. The disappointing performance of CS recovery methods asserts the major challenges CS faces, despite the fact that CS introduces the brilliant idea of changing the dominant practice of "oversampling followed by massive dumping" in image acquisition and compression.

Nevertheless, by shifting the burden of code optimization to the decoder the CS-based data acquisition devices can be greatly simplified. The encoder simply makes a small number of random projections of the signal, quantizes and transmits the projection values. This asymmetric design is highly desired when the data acquisition devices must be simple and operate on limited power budget.

Wireless sensor networks (WSN) is one of the applications that can greatly benefit from CS. WSN is a collection of low-cost, low-power sensor nodes that communicate in short distance and collaborate together to reach the objective of a WSN application. Some applications of WSN are environmental monitoring, biomedical research, human imaging and tracking and military applications [6]. In cases where the number of samples are large, compression must be done prior to transmission. In these applications there are limitations on computational power of sampling devices or communication channel bandwidth. Therefore reducing the complexity and power consumption of the senors is desirable. Since sampling process of a compressive sensing device is simple and the collected measurements are already compressed, the devices built based on compressive sensing need less computational power and have cheaper embedded hardware. Another advantage of CS based sensors is that the encoder can be designed in such a way that data stream is robust against packet loss when the communication channel is noisy. The reason is that signal information is evenly distributed amongst the measurements and if some measurements are dropped during communication, it is still possible to recover the signal using received measurements[7].

CS based simple data acquisition devices that require fewer number of samplers are also highly desired when very expensive sampling sensors are used in the capturing device (e.g., infrared imaging) or when high-density sampling can harm the object being captured (e.g., medical imaging). Medical imaging devices cannot have as many sensors as needed to capture samples from whole patient's body either because of costly sensors or harmful high-density sampling. Therefore, the patient is required to be moved through the device during data capturing process which could add artifacts to the captured image due to patient movements. Since CS based devices require fewer number of sensors, it is possible to build devices that can take images from the whole body at once.

To see how CS can reduce the number of sensors in the capturing device a practical example of a single-pixel compressive digital camera is presented. The CS based camera is introduced by Baraniuk in [8]. Figure 1.1 shows the model of this CS based camera which requires only one photodiode instead of millions of sensors used in a conventional digital camera. Instead of having N photodiode sensors, this camera uses a digital micro-mirror device (DMD) which consists of an array of N tiny mirrors. The light-field is reflected off the DMD and is then collected by a second lens and focused onto the single photodiode. The direction of each mirror is randomly set to or away from the photodiode by using a random number generator (RNG). The orientation of the mirrors creates a measurement vector that is used to calculate one measurement of the desired image. The process of setting the orientation of the mirrors and calculating the measurement of the image is repeated M times to obtain all desired measurement values. These measurements will later be used at the decoder to regenerate the image by solving an optimization problem. Figure 1.1(b) is the image taken by conventional digital camera and Figure 1.1(c),(d) are the 64×64 images taken by the single-pixel compressive camera using 800 and 1600 measurements respectively.

Despite the enthusiastic idea of CS which challenges the wisdom of conventional data acquisition systems, the performance of existing CS-based compression methods is not satisfactory. The poor performance of current CS recovery techniques compared to conventional coding techniques is caused by oversimplified assumption in the problem formulation (1.1) for CS recovery. A natural signal $\mathbf{x}$ is typically non-stationary, and there exists no space $\mathbf{\Psi}$ in which all parts of $\mathbf{x}$ have sparse representation. The problem is particularly intense for images. For a non-stationary two dimensional $m \times n$ image signal $x(i,j) \in \mathbb{N}^{m \times n}$, in two different areas $\mathcal{A}_k$ and $\mathcal{A}_l$ of the spatial domain, sub-images $x_k(i,j)$ and $x_l(i,j)$ can have very different waveforms

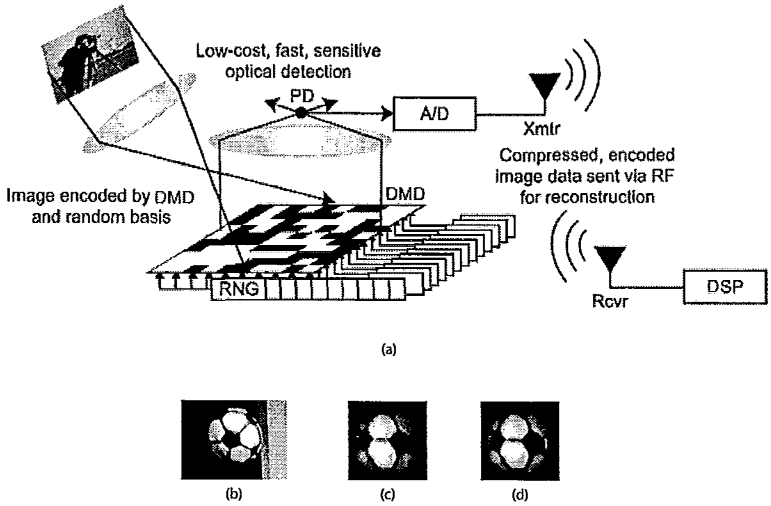

Figure 1.1: (a) The model of single-pixel compressive camera. (b) Original 64×64 image taken by conventional digital camera. (c) 64x64 image taken by single-pixel camera with 800 measurements. (d) 64x64 image taken by single-pixel camera with 1600 measurements. The image in (b) is not meant to be aligned with images in (c) and (d).

(e.g., smooth shade vs. strong edge), and hence they are sparse in different spaces Ψ_k and Ψ_l. Thus, performing CS recovery in a fixed space Ψ, such as that of DCT, a wavelet, or total variation, can and do fail in parts of the image. To address this problem we use locally adaptive strategy to recover CS-acquired images.

In this thesis we propose a new framework of model-based adaptive recovery of compressive Sensing (MARX) to solve the problem current CS recovery formulation (1.1) faces. The feature of MARX which makes it distinguished from other CS recovery methods is a locally adaptive sparse signal representation based on a piecewise autoregressive (PAR) model. The PAR model is defined by

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \epsilon. \tag{1.2}$$

where $\mathbf{x}$ is the vector representation of an image $x(i,j) \in \mathbb{N}^{m \times n}$ by stacking all $N = m \times n$ pixels of it, and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a real-valued square matrix with all elements on the main diagonal being zero. The term $\epsilon \in \mathbb{R}^N$ is a random vector that is the excitation of the 2D autoregressive process. $\mathbf{a}_k$ is the k^{th} row vector of $\mathbf{A}$ where $a_{kk} \equiv 0$, $1 \leq k \leq N$. Assuming an image to be a random Markov field (RMF) of a modest order, the number of nonzero elements of each row vector $\mathbf{a}_k$ is very small, i.e., it is sparse. The nonzero elements of $\mathbf{a}_k$ comprise the 2D support of the regression relation $x_k = \mathbf{a}_k\mathbf{x} + \epsilon$ for pixel x_k. The image waveform at the pixel location k gives the spatial configuration and the order of the regression support for x_k [9].

Assuming that a natural image is a non-stationary RMF, MARX allows the PAR model parameters $\mathbf{a}_k$ to vary in k. The PAR model offers a sparse and yet adaptive representation of image signal $\mathbf{x}$. Thus, the following problem of l_1 minimization can be used for the CS recovery of $\mathbf{x}$:

$$\min_{\mathbf{A},\mathbf{x}} \sum_{k=1}^{N} \|\mathbf{a}_k\|_{l_1} \text{ subject to } \mathbf{y} = \mathbf{\Phi}\mathbf{A}\mathbf{x} \tag{1.3}$$

Note that $\mathbf{A}$ in (1.3) is variable where as $\mathbf{\Psi}$ in (1.1) is predetermined. The proposed MARX sparsity mechanism can fit image local structures (e.g., edges, textures, smooth shades, etc.) much better than wavelet, curvelet, DCT or whatever predetermined basis of $\mathbf{\Psi}$ [9].

Clearly, the proposed MARX objective function is computationally more complex than the current CS problem formulation. MARX jointly estimates the pixel values of the image and PAR model parameters in contrast to current CS problem formulation which estimates $\mathbf{x}$ in a fixed space $\mathbf{\Psi}$. The added search space of $\mathbf{A}$ makes the inverse problem of CS recovery severely under-determined. In our research we developed algorithm techniques to overcome this difficulty, making the MARX solution feasible and robust. To achieve maximum bonds on the solution space for (1.3), we used structured sparsities due to self similarities of natural images. The resulting technique makes the MARX process computationally tractable and greatly improves the performance of existing CS recovery algorithms. Indeed, experimental results show the superior recovery quality of MARX over other CS methods.

The results of proposed MARX algorithm have superior quality specifically along the edges where there is high frequency in the local window. However, since MARX algorithm is not robust to noise, it generates worm-like artifacts in areas where no dominant edge is present (e.g., smooth or noisy areas). In this thesis we also proposed hybrid TV-MARX algorithm that addresses the problems of MARX algorithm. This algorithm dynamically switches between TV and PAR depending on the characteristics of the local context. The pixels in the image are put into two categories: 1) $\mathbf{x}_{PAR}$ which is the set of pixels on dominant edges and 2) $\mathbf{x}_{TV}$ which is the set of pixels in smooth or noisy neighborhood. For the pixels in $\mathbf{x}_{PAR}$, PAR model is applied during recovery process and for the ones in $\mathbf{x}_{TV}$ TV is used.

To further improve the quality of recovered image, weighted PAR models are used during restoration process. The weights are calculated such that they favor the PAR models that are in the dominant edge direction. Experimental results showed that in average the quality of proposed hybrid TV-MARX recovery algorithm is $0.6dB$ higher than the results of MARX algorithm in peak signal-to-noise ratio (PSNR).

1.1 Literature Review

Signal processing is an area of applied mathematics that deals with operations on discrete or continues signals. Data compression, data transmission, denoising, prediction, filtering, smoothing and deblurring are examples of such operations [10]. Signals that these operations are applied on include sound, image and sensor data. Signal processing is used in a wide range of applications such as medicine, communications, entertainment and military systems. As communications systems become wireless, mobile and multi-functional, the importance of sophisticated signal processing grows. Rapid evolution of digital computers and development of important theoretical algorithms such as fast fourier transform (FFT) attracted more attentions to the field of digital signal processing from 1960s. The fundamental part of digital signal processing is signal sampling during which a continuous-time signal is converted into a discrete-time signal through sampling [11].

The sampling process is based on sampling theorem proposed by Shannon in 1949. He stated the theorem as:

If a function $f(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced $1/2W$ seconds apart.

"cps" is cycle per second which Shannon used instead of Hertz. Shannon named the upper band on the signal frequency Nyquist frequency due to his important contributions to the communication theory. As Shannon admitted himself, it was not the work he did himself rather it is a fact which is common knowledge in communication art [12]. Nonetheless Shannon should get a full credit for formalizing it for the first time. Similar theorem also appeared in mathematical literature due to the works done by Whittaker in [13]. Also in Russian literature, Kotelnikov in [14] introduced this theorem to communication theory [15].

This sampling theorem, regardless of who proposed it first, imposes a lower bound on the sampling rate. This requirement should be satisfied during the sampling process to avoid aliasing effect which corrupts signal characteristics. In certain cases it might be impossible or difficult to build sampling devices that meet Nyquist rate requirement due to hardware limitation, i.e. it is practically impossible to design an analog to digital converters (ADC) that samples at Nyquist rate using state-of-the-art technologies or it is simply too expensive to design such devices. Therefore, many sampling schemes with different properties have been investigated in digital signal processing field to address this issue. Most of these researches are based on nun-uniform sampling strategies [16, 17, 18, 19, 20, 21, 22].

Bilinskis in [21], states that for some signals, it is possible to acquire samples below required Nyquist rate without suffering from aliasing effect by using non-uniform sampling. In an example in his book Bilinskis showed that a signal with highest frequency equal to 1.185 GHz can be fully reconstructed by using non-uniform sampling with sampling at a mean rate of 80 MS/s (megasamples per second). Note that uniform sampling rate for this signal should be at least 2.370 GHz. He then says that once aliasing is somehow eliminated the rate of sampling doesn't depend on the highest frequency component of the signal and non-uniform sampling is one way of removing aliasing effect. The problem of high sampling rate for uniform sampling is that the excessive sampling values don't add information and are only used to remove overlapping signal spectral components.

Compressive sensing (CS) also known as compressed sensing and compressive sampling is a data acquisition technique developed in recent years that acquires data with sampling rate below the Nyquist rate. The idea was first used in earth sciences during 1970s where seismologists used it to construct images of reflective layers in the earth from data that did not seem to satisfy the Nyquist rate. The idea of CS as is known today is originated from the works done by Emmanuel J. Candes, a mathematician at

Caltech, in 2004. While working on a problem in magnetic resonance imaging (MRI), Candes discovered that a test image can be exactly reconstructed even though the available data is less than what is required by Nyquist criterion. To make sure this result is not accidental, he collaborated with Justin Romberg and published much of the underlying theory in [23, 24, 2, 25]. Candes also started to work with Terence Tao of UCLA and published a series of papers [26, 27, 28, 29, 30, 31] to set forth the basic principles of compressive sensing.

The results from Candes works attracted wide attention even before Candes, Romberg and Tao published their papers. David Donoho, Candes' Ph.D. advisor at Stanford University, made outstanding contributions to the theory and applications of CS [1, 32, 33, 34, 35, 36, 37]. Richard Baraniuk at Rice University leads a large and active group in CS research area and published wide range of papers on this topic [38, 39, 40, 41, 42]. Now there is a worldwide community that conducts workshops and conferences to contribute to this rapidly expanding research area [43].

Unfortunately, given the same compression ratio, the image quality of CS recovery algorithm is not as good as the result of state-of-the-art compressing techniques such as JPEG-2000. The quality of reconstruction depends on signal compressibility, incoherence of measurement matrix with sparsity basis and chosen reconstruction algorithm [39]. One of the most challenging research topic in compressive sensing is the design of recovery algorithm. One category of recovery algorithms builds approximate signal in each step by making locally optimal choices. Orthogonal matching pursuit (OMP) [4], stage-wise OMP [34] and regularized OMP [44, 45] are some examples of this type of algorithms. Convex relaxation algorithms such as interior-point method

[28, 46], projected gradient methods [3] and iterative thresholding [5] are other techniques that solve a convex program with known minimizer. Fourier sampling [47, 48], chaining pursuit [49] and heavy hitters on steroids (HHS) pursuit [50] are a third type of algorithms that support rapid reconstruction via group testing [51].

The focus of all these algorithms described above is on improving image quality during signal recovery step of CS. Another way of improving the quality of reconstructed image is to use a recovery algorithm on the CS acquired images. In other words, we use one of the algorithms mentioned above to reconstruct the image from its measurements and further improve the quality of the image using a recovery algorithm. In this thesis we proposed two recovery algorithms to improve the quality of the CS reconstruction. Our algorithms are based on non-linear estimation for adaptive restoration (NEAR) algorithm proposed in [52].

1.2 Contributions

In this thesis, we proposed model-based adaptive recovery (MARX) for CS. This algorithm is a modification of NEAR that uses measurement matrix as the degrading function. It uses axial and diagonal piecewise autoregressive (PAR) models for each pixel to recover the CS acquired image using the measurement matrix as the matrix representation of a degrading function.

The second algorithm is hybrid TV-MARX CS recovery algorithm. The core of this technique is very similar to the MARX algorithm. Some improvements are made in hybrid TV-MARX to address the problems MARX has.

Due to the sensitivity of MARX algorithm to the noise, this algorithm generates worm-like artifacts in smooth and noisy local context because of data over-fitting.

Hybrid TV-MARX algorithm solves this problem by mixing TV and PAR during recovery process and dynamically switches between TV and PAR based on the presence of edge(s) near the current pixel.

Hybrid TV-MARX further improves MARX algorithm by applying weights on PAR models to favor the ones that are in the dominant edge direction. The direction of dominant edge is determined using an edge detector algorithm. The weights are chosen using Gabor function in four direction ($0°$, $45°$, $90°$, $135°$).

The comparison study verifies that MARX algorithm improves the quality of CS reconstruction by $1.7dB$ in average. Hybrid TV-MARX algorithm achieves superior results compared to MARX algorithm and removes most of the artifacts created by this algorithm.

1.3 Organization

This thesis is organized as follows. Chapter 2 presents mathematical background of compressive sensing and NEAR restoration techniques. Details of MARX and hybrid TV-MARX algorithms are proposed in Chapter 3. Chapter 4 presents the experimental results to compare the proposed algorithms. Finally, conclusion and future work remarks are presented in Chapter 5.

Chapter 2

Mathematical Background

2.1 Introduction

In this chapter the mathematical background for our proposed algorithms is presented. The first section is a mathematical overview of compressive sensing theory. The necessary conditions for a measurement matrix to recover signal $\mathbf{x}$ of length N from $M << N$ measurements are discussed. Different types of measurement matrices used in CS applications are reviewed. The description of the CS reconstruction algorithm is also presented. The second section starts with a brief introduction of the image restoration problem and talks about mathematical approaches to solve it. Then it proceeds to outline the NEAR restoration technique on which the proposed MARX algorithm is based.

2.2 Compressive Sampling

Conventional data acquisition systems are based on Shannon-Nyquist sampling theory. This theorem imposes lower bound on the number of samples that should be taken from a signal in order to fully recover it. According to Shannon-Nyquist sampling theory, for a bandlimited signal the sampling rate should be at least twice as high as the highest frequency component of the signal. For signals such as images and videos the number of samples is still so high that working directly with sampled data is inefficient. Thus, the large body of samples have to be compressed prior to transmission or storage.

For an image $x(i,j) \in \mathbb{N}^{m\times n}$, lets represent it by a vector $\mathbf{x}$ whose elements are stacked $N = m \times n$ pixels. Suppose $\mathbf{\Psi}$ is an $N \times N$ orthonormal basis matrix. $\mathbf{x}$ can be defined in Ψ as

$$\mathbf{x} = \mathbf{\Psi}\mathbf{s}$$

where $\mathbf{s}$ is the coefficients vector of length N which represents signal $\mathbf{x}$ in domain $\mathbf{\Psi}$.

A signal $\mathbf{x}$ is K-sparse in basis $\mathbf{\Psi}$ if all but K number of coefficients are zero in $\mathbf{\Psi}$. The compression can be done by transforming the signal into $\mathbf{\Psi}$ basis and sending just those K nonzero coefficients and their location instead of sending all N samples. In real world it is almost impossible to find natural signals that are strictly K-sparse. However, most signals can be approximated well by a sparse signal by truncating small coefficients of the signal to zero. This approximation will not degrade the reconstruction quality and is used in practice, which is in fact the foundation of transform coding [53].

As an example let examine image compression in wavelet domain. As illustrated in

Figure 2.1 (b), the transformed image generates lots of very small coefficients. During compression all coefficients less than specified threshold are set to zero. Incrementing the threshold increases the compression ratio but reduces the quality of compressed image. Because of high rate of redundancy in natural images wavelet domain can be used to generate high quality images with very low bit-rate (Figure 2.1(c, d)).

The steps of a typical digital acquisition system are:

1. Acquire N samples from continuous signal to generate discrete signal $\mathbf{x}$.

2. Transform signal $\mathbf{x}$ into basis $\boldsymbol{\Psi}$ where the representation of the signal is nearly sparse.

3. Keep K largest coefficients and their locations and discard all other $(N - K)$ small coefficients.

4. Encode these K values and their locations.

Baraniuk in [8] called this procedure the sample-then-compress framework. This framework requires to capture all N samples, which is very large, even though the number of coefficients being kept is small. Also to calculate K large coefficients, all other ones should be calculated even tough they will be discarded later. Finally, the location of K large coefficients should be encoded along with the values of these coefficients. This adds an overhead to the system.

Compressive sampling addresses these inefficiencies by introducing a new data acquisition process that combines sampling and compressing in one step. The basic idea behind this theory is that one can reconstruct a signal from fewer number of measurements than what is required by Shannon-Nyquist theorem if

1. the signal is sparse in some basis $\boldsymbol{\Psi}$

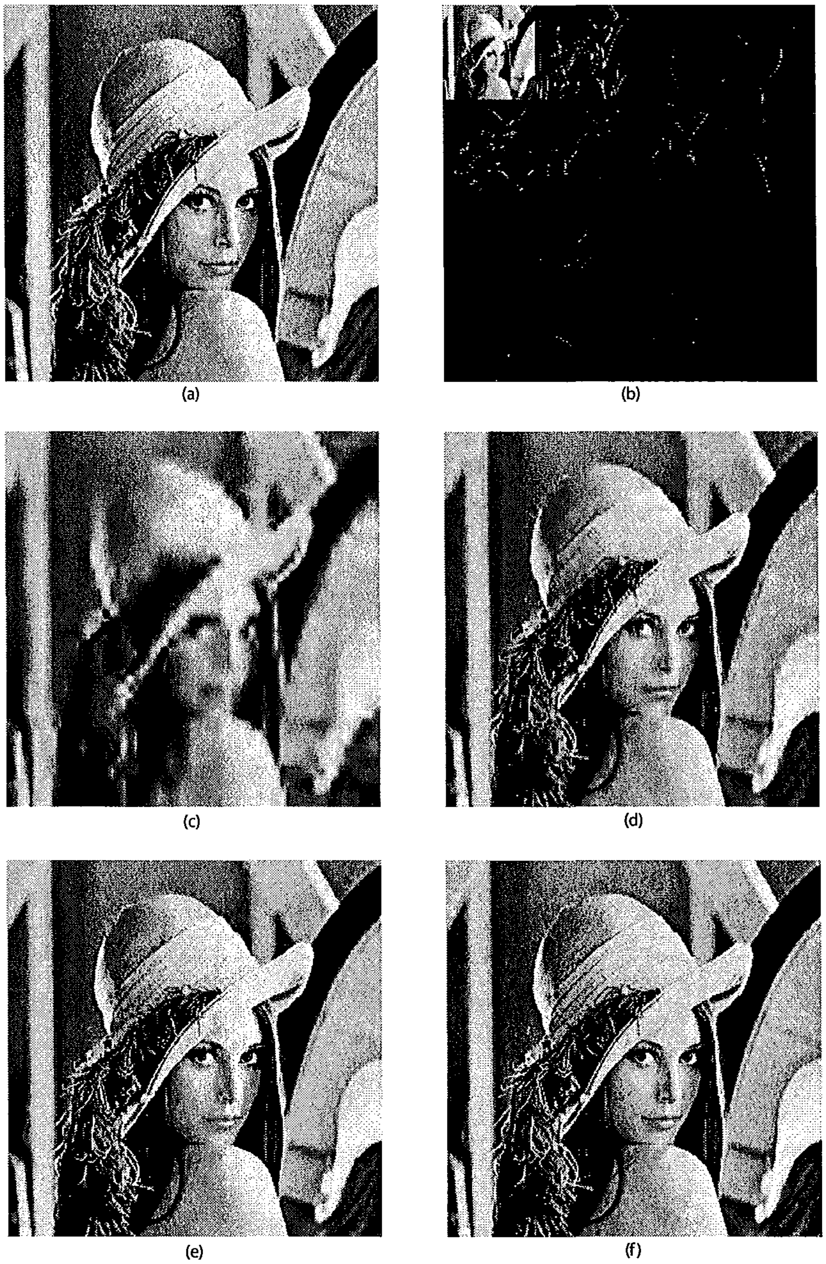

Figure 2.1: Comparison of lossy JPEG2000 encoder with different bit-rates. (a) Original image. (b) Wavelet coefficients for two level of transforms. (c) Compressed image with 0.08 BPP bit-rate (Compressing ratio = 1:100). (d) Compressed image with 0.4 BPP bit-rate (Compressing ratio = 1:20). (e) Compressed image with 0.8 BPP bit-rate (Compressing ratio = 1:10). (f) Compressed image with 4 BPP bit-rate (Compressing ratio = 1:2). Images are taken from [54]

2. the measurement matrix $\mathbf{\Phi}$ is stable so that it can capture almost all the important information in any K-sparse signal without damaging it by the dimensionality reduction from $\mathbf{x} \in R^N$ to $\mathbf{y} \in R^M$.

For now we focus on signals that are K-sparse, i.e., all other $(N-K)$ coefficients of the signal are exactly zero. Later we will discuss how CS theory can be applied on nearly K-sparse signals. A signal is said to be nearly K-sparse in basis $\mathbf{\Psi}$ if it has K large coefficients in $\mathbf{\Psi}$ and all other $(N-K)$ coefficients are small values but not necessarily zero.

Suppose $\mathbf{x}_K$ is the sampled signal of length N which is K-sparse in basis $\mathbf{\Psi}$ and $\mathbf{\Phi}$ is an $M \times N$ measurement matrix. If we define $\mathbf{y}$ as measurement vector, the calculation can be written in matrix form as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}_K$$

using sparse representation of signal $\mathbf{x}_K$ in $\mathbf{\Psi}$ we can write above equation as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s}_K = \mathbf{\Theta}\mathbf{s}_K \tag{2.1}$$

where $\mathbf{s}_K$ is the K-sparse representation of signal $\mathbf{x}_K$ and $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$ is an $M \times N$ matrix. The goal of CS is to use measurement vector $\mathbf{y}$ and matrix $\mathbf{\Theta}$ to regenerate coefficients vector $\mathbf{s}_K$. Once $\mathbf{s}_K$ is known the original signal $\mathbf{x}_K$ can be recovered as

$$\mathbf{x}_K = \mathbf{\Psi}\mathbf{s}_K$$

The following questions should be answered in order to define a practical CS based data acquisition system:

1. how to design a measurement matrix $\mathbf{\Phi}$ such that the ill-posed inverse problem in (2.1) has stable solution

2. what is the reconstruction algorithm that can recover signal $\mathbf{x}_K$ from $M \approx K << N$ measurements

The first question is concerned with the design of a stable measurement matrix $\mathbf{\Phi}$ that can capture almost all the information about signal $\mathbf{x}_K$. The measurement matrix $\mathbf{\Phi}$ should allow the reconstruction of signal $\mathbf{x}_K$ of length N from $M < N$ measurements. The problem of finding N unknowns from M measurements is ill-posed. The sparsity of signal $\mathbf{x}_K$ can be used to reduce the number of unknowns from N to $K \leq M$. If the places of K nonzero coefficients are known, a necessary and sufficient condition for the ill-posed inverse problem in (2.1) to have stable solution is that for any vector $\mathbf{v}_K$ sharing the same K nonzero elements as $\mathbf{s}_K$ the following inequality holds for some $\epsilon > 0$ [8]

$$1 - \delta \leq \frac{\|\mathbf{\Theta}\mathbf{v}_K\|_2}{\|\mathbf{v}_K\|_2} \leq 1 + \delta \tag{2.2}$$

However, the assumption of knowing the positions of nonzero elements of $\mathbf{s}_K$ is not desired since it adds overhead to the encoding process and wastes bandwidth. Candes, Romberg and Tao in [28] proposed restricted isometry property (RIP) to define a sufficient condition for stable solution for $\mathbf{s}_K$ when the position of nonzero elements are not known. RIP states that a sufficient condition to have a stable solution for a K-sparse signal is that $\mathbf{\Theta}$ satisfies (2.2) for any $3K$-sparse vector $\mathbf{v}_{3K}$. If matrix $\mathbf{\Theta}$ satisfies the (2.2) for K-sparse signal, we say that $\mathbf{\Theta}$ obeys RIP of order K.

To construct a measurement matrix $\mathbf{\Phi}$ which obeys RIP of order K, one needs to verify (2.2) for each of the $\binom{N}{3K}$ possible combinations of 3K nonzero elements of vector $\mathbf{v}_{3K}$. Unfortunately, for a large N this approach is not practical. Also, checking matrix $\boldsymbol{\Theta}$ to have RIP, makes encoder dependent to the type of signal. This is because the encoder should know the basis $\boldsymbol{\Psi}$ in which the signal has sparse representation in order to generate matrix $\boldsymbol{\Theta}$.

In [55] Candes stated that if $\boldsymbol{\Phi}$ is a Gaussian measurement matrix and

$$M \geq CK \log \frac{N}{K} \tag{2.3}$$

for some constant C, then the measurement matrix $\boldsymbol{\Phi}$ has RIP with probability $1 - O(e^{-\lambda^N})$ for some $\lambda > 0$. Therefore, if the measurement matrix is a Gaussian projection and if the number of measurements M satisfies the inequality in (2.3), then one can recover any K-sparse signal $\mathbf{x}_K$ with high probability. The measurement matrix $\boldsymbol{\Phi}$ is Gaussian matrix if the elements of it are independent and identically distributed (IID) random variables from a Gaussian probability distribution function (PDF) with zero mean and variance $1/M$, By using the Gaussian measurement matrix, signal $\mathbf{x}_K$ can be exactly recovered with probability $1 - O(e^{-\lambda^N})$ for some $\lambda > 0$.

Another random measurement matrix that has RIP is Fourier measurement matrix. Fourier measurement matrix is a partial Fourier matrix obtained by selecting M rows uniformly at random. The columns of the M selected rows should be normalized so that they have unit norm. In case of Fourier measurement Candes stated that if

$$M \geq CK \log^4 N \tag{2.4}$$

then Fourier measurement matrix has RIP with 'overwhelming probability'. Specifically the 'overwhelming probability' here means that given constant $C = 22(\sigma + 1)$, the probability of exactly recovering $\mathbf{x}_K$ exceeds $1 - O(N^{-\sigma})$. Therefore, using a Fourier measurement matrix with enough number of measurements M that satisfies the inequality in (2.4), then one can recover any K-sparse signal $\mathbf{x}_K$ with high probability.

The above discussions address how to choose a measurement matrix that is incoherent to the representation basis. The measurement matrix can be a random matrix. If the number of measurements, M, satisfies the inequality in (2.4), the representation matrix $\mathbf{\Psi}$ is incoherent to the sensing matrix $\mathbf{\Phi}$ with high probability. Therefore we can expect to recover the signal with high probability.

Once measurements are made by using some random measurement matrix such as Gaussian measurement matrix or Fourier measurement matrix, then the problem becomes how to recover full-length signal from these measurements. This requires to solve an under-determined system of equations with some constraints. The reconstruction algorithm needs measurement vector $\mathbf{y}$, random measurement matrix $\mathbf{\Phi}$ and basis $\mathbf{\Psi}$ as input. Notice that the whole random measurement matrix does not have to be sent to the reconstruction algorithm. Instead, a pseudo random measurement matrix with a seed can be used. The seed is used during reconstruction to regenerate matrix $\mathbf{\Phi}$. Since $M < N$ the equation $\mathbf{\Theta s} = \mathbf{y}$ has infinite number of solutions in $H = \mathcal{N}(\mathbf{\Theta}) + \mathbf{s}_K$ space, where $\mathcal{N}(\mathbf{\Theta})$ is the null space of $\mathbf{\Theta}$. The reconstruction algorithm needs to search for the sparse coefficient vector of signal $\mathbf{x}_K$ in space H.

The classical way of solving an under-determined system of equations uses l_2 minimization to find a solution with minimum energy. The l_2 minimization has a closed-form solution $\hat{\mathbf{s}} = \Theta^T(\Theta\Theta^T)^{-1}\mathbf{y}$. However for CS reconstruction, the l_2 minimization solution ia not sparse, not satisfying the K-sparsity property of the signal.

Since we know that the the signal is sparse in space Ψ, the following l_0 minimization is the optimal criterion for CS reconstruction

$$\hat{\mathbf{s}} = argmin\|\mathbf{s}\|_0 \qquad \text{such that} \qquad \Theta\mathbf{s} = \mathbf{y} \tag{2.5}$$

The l_0 minimization selects a solution from solution space H which has the minimum number of nonzero elements. Unfortunately, the problem (2.5) is known to be NP-complete [56]. An alternative algorithm is needed to make the problem tractable.

Surprisingly, the following optimization problem based on l_1 minimization can exactly recover K-sparse signal with high probability [28].

$$\hat{\mathbf{s}} = argmin\|\mathbf{s}\|_1 \qquad \text{such that} \qquad \Theta\mathbf{s} = \mathbf{y} \tag{2.6}$$

The difference between (2.6) and (2.5) is that the former uses the sum of magnitudes instead of size of support [57]. This optimization problem can be recast to one of linear programming which can be solved in $O(N^3)$ [55, 1].

In summary, the CS-based data acquisition and communication consist of as follows components:

1. Design a random sampling device that directly generates measurement vector $\mathbf{y}$ (A single-pixel camera is an example of such device that was described in introduction chapter), bearing in mind to take at least the minimum number of measurements required.

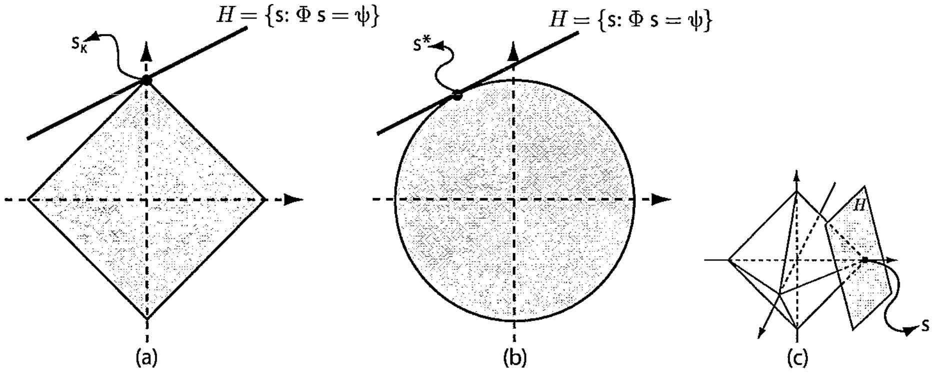

Figure 2.2: Geometry of l_1 minimization. (a) l_1 ball finds desired sparse vector **s** from solution space H in two dimensional space. (b) l_2 ball finds a solution with many nonzero elements instead of the sparse vector in two dimensional space. (c) l_1 minimization in three dimensional space. The graphs are taken from [53]

2. Send the measurement vector $\mathbf{y}$ along with the random seed used to generate random measurement matrix to the decoder.

3. Decoder uses basis $\mathbf{\Psi}$ that represents signal in sparse form to generate matrix $\mathbf{\Theta} = \mathbf{\Phi\Psi}$. Then the decoder solves convex optimization problem in (2.6) to exactly recover signal $\mathbf{x}_K$ with 'overwhelming probability'.

One of the interesting characteristics of CS is that during signal capturing the encoder does not need to know what basis is used to recover signal. Selection of basis $\mathbf{\Psi}$ can be done at decoder and even different basis can be used on different parts of the image to increase the performance of the reconstruction algorithm.

A geometrical intuition on why l_2 minimization fails to find the sparse solution in contrast to l_1 minimization can be had as follows. Precise mathematical proofs can be found in [29, 58, 59]. Figure 2.2(a) shows an l_1 minimization process in two dimensions. The gray square is the l_1 ball in $\mathbb{R}^2$. The gray region contains all $\mathbf{s} \in \mathbb{R}^2$ such that $|\mathbf{s}(1) + \mathbf{s}(2)| \leq r$ where r is the radius of the l_1 ball. The line marked with H is the solution space and point $\mathbf{s}_K$ is the desired sparse solution. One can imagine

l_1 minimization process as blowing l_1 ball by starting from small radius and gradually expanding it until it hits line H for the first time. The first intersection of l_1 ball and line H is the solution to the l_1 minimization problem. As illustrated in the Figure 2.2(a), l_1 minimization finds the desired sparse vector.

Figure 2.2(b) shows an l_2 minimization process in two dimensions. The gray circle is the l_2 ball with radius r and the gray region contains all $\mathbf{s} \in \mathbb{R}^2$ such that $\mathbf{s}(1)^2 + \mathbf{s}(2)^2 \leq r^2$. The line marked with H is the solution space. The process of l_2 minimization can be imagined to be expanding a circle by gradually increasing the radius. The first intersection between the l_2 ball and line H is the solution to the l_2 minimization problem. But this solution is not necessarily sparse.

So far we considered signal $\mathbf{x}_K$ to be K-sparse but in real world applications signals are nearly sparse. Also in any real applications measured data is corrupted by some amount of noise (such as quantization noise and sensor noise). To be practical, CS needs to deal with the cases where signals are approximately sparse and noise is presented during the capturing process. The question, however, is weather it is possible to accurately reconstruct nearly sparse signals with the presence of noise from highly under-sampled measurements?

We first examine the accuracy of CS for nearly sparse signals without the presence of noise. Then, noise element is added to the system and the error bound for CS reconstruction is determined based on the theorems provided in [29]. Suppose $\mathbf{\Phi}$ is a measurement matrix that obeys the RIP in (2.2). For K-sparse signals $\mathbf{x}_K^1$ and $\mathbf{x}_K^2$ with sparse representations $\mathbf{s}_K^1$ and $\mathbf{s}_K^2$ in $\mathbf{\Psi}$, the following inequality holds for all K-sparse signals when δ_{2K} is sufficiently less than one (which is the case when stable measurement matrices such as Gaussian or Fourier measurement matrices are used)

[60]

$$1-\delta_{2K} \leq \frac{||\Theta\mathbf{s}_K^1 - \Theta\mathbf{s}_K^2||_2}{||\mathbf{s}_K^1 - \mathbf{s}_K^2||_2} \leq 1+\delta_{2K} \tag{2.7}$$

where $\Theta = \Phi\Psi$. Suppose $\mathbf{x}$ is a nearly K-sparse signal and $\mathbf{s}$ is its representation in Ψ. Define $\mathbf{s}_K$ as the K-sparse approximation of $\mathbf{x}$ in Ψ by truncating all (N-K) small coefficients of $\mathbf{s}$ to zero. Without the presence of noise the measurement vector $\mathbf{y}$ of signal $\mathbf{x}$ is calculated as

$$\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\mathbf{s} = \Theta\mathbf{s}$$

and following optimization problem is solved to find sparse representation of signal $\mathbf{x}$

$$\hat{\mathbf{s}} = argmin||\tilde{\mathbf{s}}||_1 \qquad \text{such that} \qquad \Theta\tilde{\mathbf{s}} = \mathbf{y} \tag{2.8}$$

Theorem:

Assume that $\delta_{2K} < \sqrt{2}-1$ in (2.7). Then the solution $\hat{\mathbf{s}}$ to (2.8) obeys

$$||\hat{\mathbf{s}} - \mathbf{s}||_2 \leq C_0 \frac{||\mathbf{s} - \mathbf{s}_K||_1}{\sqrt{K}}$$

$$\text{and} \tag{2.9}$$

$$||\hat{\mathbf{s}} - \mathbf{s}||_1 \leq C_0.||\mathbf{s} - \mathbf{s}_K||_1$$

for some constant C_0. The above theorem was proposed by Candes, Romber and Tao in [29] to characterize the CS recovery of nearly sparse signals as well as sparse ones. If $\mathbf{x}$ is K-sparse, that is $\mathbf{x} = \mathbf{x}_K$, then the recovery would be exact. If signal $\mathbf{x}$ is not K-sparse, then (2.9) states that the quality of the recovered signal is as good as the one that would be obtained by knowing everything about the $\mathbf{s}$ and selecting its K largest entries [29]. Therefore, a nearly sparse signal can be recovered from its random measurements with high accuracy by solving the l_1 minimization problem.

Suppose that we are given a noisy measurement vector $\mathbf{y}$ of a nearly sparse signal $\mathbf{x}$

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{z} = \mathbf{\Phi \Psi s} + \mathbf{z} = \mathbf{\Theta s} + \mathbf{z},$$

where $\mathbf{z}$ is an unknown noise term added to the system and $\mathbf{s}$ is the nearly sparse representation of $\mathbf{x}$ in $\mathbf{\Psi}$. The l_1 minimization problem in (2.6) with relaxed constraint can be written as

$$\hat{\mathbf{s}} = argmin\|\tilde{\mathbf{s}}\|_1 \qquad \text{such that} \qquad \|\mathbf{\Theta}\tilde{\mathbf{s}} - \mathbf{y}\|_2 \leq \epsilon \tag{2.10}$$

where ϵ is the bound on the noise element. The following theorem in [29] shows that even for noisy measurements of a nearly sparse $\mathbf{x}$, the reconstruction error is bounded.

Theorem:

Assume that $\delta_{2K} < \sqrt{2} - 1$ in (2.7). Then the solution $\hat{\mathbf{s}}$ for (2.10) obeys

$$\|\hat{\mathbf{s}} - \mathbf{s}\|_2 \leq C_0 \frac{\|\mathbf{s} - \mathbf{s}_K\|_1}{\sqrt{K}} + C_1.\epsilon \tag{2.11}$$

for some constants C_0 and C_1. The above theorem declares that the reconstruction error is bounded by two error terms. One term is the error due to the fact that the signal is nearly sparse (other than being fully sparse) and the other is because of the added noise and is proportional to the noise level.

Being able to recover nearly sparse signal in the presence of noise with high probability and acceptable error makes CS a practical and robust sensing mechanism. It can work with all types of signals including nearly sparse ones and can handle noise elegantly [60].

In image processing applications the total variation (TV) method, which is closely related to the l_1 minimization, is widely used for solving inverse problems and it is also used as the criterion for CS reconstruction. TV was originally proposed by Rudin, Osher, and Fatemi in [61] and gained popularity in the literature. For a two dimensional image $x(i,j)$, TV is defined

$$TV(\mathbf{x}) = \sum_{i,j} \sqrt{(x(i+1,j) - x(i,j))^2 + (x(i,j+1) - x(i,j))^2}$$

which is the sum of the gradient magnitude at every pixel

$$TV(\mathbf{x}) = \sum_{i,j} |\nabla x(i,j)|$$

Let $\mathbf{x}$ be the vectorized representation of $x(i,j)$, then the TV-based CS reconstruction can be stated as

$$\hat{\mathbf{x}} = argmin\,(TV(\mathbf{x})) \qquad \text{such that} \qquad \|\Phi\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \tag{2.12}$$

The TV criterion favors signals of bounded variation (e.g., smooth images), without sufficient consideration of signal transients (e.g., image edges). Due to the fact that most natural signals have an exponentially decay power spectrum, TV performs better than the l_1 minimization in most cases. However, the TV performance deteriorates quickly in signal segments of discontinuities and high frequency components.

To overcome this drawback of TV, we introduce a methodology of model-based adaptive recovery of compressive sensing (MARX). The defining feature of MARX, which distinguishes it from other CS recovery techniques, is a locally adaptive sparse signal representation facilitated by a piecewise autoregressive (PAR) model. This algorithm is based on the non-linear estimation for adaptive restoration (NEAR) technique proposed in [52]. In the next section the NEAR technique is briefly reviewed.

2.3 Non-linear Estimation for Adaptive Restoration

Restoration is the process of recovering an image from its degraded version by using prior knowledge of the degrading function [62]. Suppose $x(i,j)$ is the original image and h is the degrading function in spatial domain. The degraded image y is given by

$$y = h * x + \eta \tag{2.13}$$

where η is the additive noise and $*$ indicates convolution. A restoration technique obtains an estimate $\hat{x}$ of the original image x by applying an inverse process to the degraded image y. A model of image degradation and restoration processes is illustrated in Figure 2.3.

For an image $x(i,j)$ of size $m \times n$, lets represent image $x(i,j)$, degraded image $y(i,j)$ and additive noise $\eta(i,j)$ in vectorized forms as $\mathbf{x}$, $\mathbf{y}$ and η respectively. Suppose $\mathbf{H}$ is an $mn \times mn$ matrix which corresponds to degrading function h. The elements of $\mathbf{H}$ are given by the elements of the convolution in (2.13). The matrix form of the

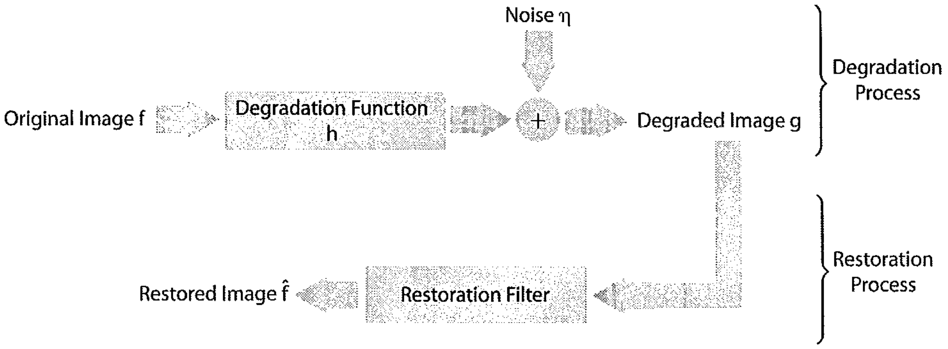

Figure 2.3: A model of image degradation and restoration processes.

degradation process is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \eta$$

The objective of restoration process is to find restored image $\mathbf{z}$ using some pre-knowledge about degrading function and additive noise. The more we know about $\mathbf{H}$ and η the closer $\mathbf{z}$ will be to $\mathbf{x}$ [62]. The simplest approach to find $\mathbf{z}$ is to apply inverse filtering. If $\mathbf{H}$ is known the restored image can be calculated as

$$\mathbf{z} = \mathbf{H}^{-1}\mathbf{y} = \mathbf{H}^{-1}\mathbf{H}\mathbf{x} + \mathbf{H}^{-1}\eta = \mathbf{x} + \mathbf{H}^{-1}\eta \tag{2.14}$$

where $\mathbf{H}^{-1}$ is the inverse of $\mathbf{H}$. Without presence of noise the above equation can exactly restore $\mathbf{x}$ since the last term in (2.14) is zero. In general, however, there is always some additive noise in the system. It means that even if we now the degradation function, the degraded image $\mathbf{y}$ cannot be exactly recovered because of the presence of noise.

Different restoration algorithms are proposed to restore degraded image $\mathbf{y}$ in presence of noise. The general restoration algorithm is constrained least square filtering.

This method requires the knowledge of the mean and the variance of the noise which can be determined from a given degraded image. The sensitivity of $\mathbf{H}$ to noise is the issue that should be considered in this algorithm. One way of reducing the impact of noise over $\mathbf{H}$ is to base the optimality of restoration process on a measure of smoothness. A typical measure of smoothness, which is used in restoration process, is the second order derivative of the image [62]. Two dimensional discrete second order Laplacian operator can be defined as

$$\nabla^2(\mathbf{x}) = \sum_{i,j}[x(i+1,j) + x(i-1,j) + x(i,j+1) + x(i,j-1) - 4x(i,j)]$$

by using above equation as the criterion function the problem of restoration process can be defined as

$$\min([\nabla^2(\mathbf{x})]^2) \qquad \text{such that} \qquad \|\mathbf{Hx} - \mathbf{y}\|_2 \leq \epsilon \tag{2.15}$$

where ϵ is the bound on the noise element. Using criterion functions that model the image as close to the reality as possible can improve the performance of the restoration process. NEAR is a powerful restoration technique proposed in [52]. This algorithm models an image as a piecewise autoregressive (PAR) process. Since most images are far from being stationary, models that assume images as stationary autoregressive processes are ill in practice [63]. However, most images are highly correlated in local windows, i.e., formed by spatially coherent contiguous pixels. This property of natural images makes it possible to model images as PAR processes and to estimate the PAR model parameters using sample statistics in a moving window. For an image $x(i,j)$, the PAR model can be defined as

$$x(i,j) = \sum_{(i,j)\in W} \sum_{u,v} a_{u,v} x(i-u, j-v) + \eta(i,j)$$

where W is the local window, $a_{u,v}$ are PAR model parameters and $\eta(i,j)$ is the measurement noise. In [52] it is stated that if image $x(i,j)$ is a two dimensional autoregressive(AR) process and is degraded by degrading function h, then the degraded image $y(i,j)$ is a two dimensional autoregressive moving average (ARMA) process. AR part of the degraded image has the same parameters as those of the original image, but its MA part is determined by h. The question is how one can estimate the PAR parameters to use during restoration process. Original image cannot be used as we only have the degraded image. Also, since the degraded image has signal dependent MA part, the PAR parameters $a_{u,v}$ cannot be estimated from it either. To resolve this problem NEAR restoration technique jointly estimates the parameters of the PAR model and the pixels of the restored image [52]. The constrained optimization problem for NEAR restoration technique can be defined as

$$\min_{\mathbf{x},a} \sum_{(i,j)\in W} \left(x(i,j) - \sum_{u,v} a_{u,v} x(i-u, j-v) \right)^2$$

$$\text{such that} \tag{2.16}$$

$$\|\mathbf{Hx} - \mathbf{y}\|_2 \leq \epsilon$$

where ϵ is the bound on noise element, $\mathbf{x}$, $\mathbf{y}$ are vectorized representation of $x(i,j)$ and $y(i,j)$ respectively and $\mathbf{H}$ is the convolution matrix which is generated based on degrading function h.

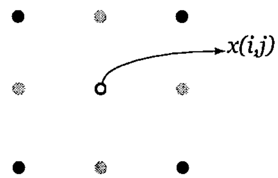

Figure 2.4: Two PAR models of Order 4 proposed in [52]. The pixel at center is image pixel at i_{th} row and j_{th} column. Black pixels are diagonal model and gray pixels are axial model.

If the size of window $|W|$ and the order of AR process t are not chosen with caution, data over-fitting might occur because of fewer number of equations compared to the number of unknowns. To overcome this problem, Wu and Zhang [52], suggested to use multiple PAR models in different direction. They proposed two PAR models of order 4 which are called diagonal model and axial model. Figure 2.4 show spatial configuration of the two PAR models. For the image pixel at (i, j), we define $\mathbf{x}_{\times}(i, j)$ as diagonal model which consists of four 8-connected neighbors of pixel $x(i, j)$

$$\mathbf{x}_{\times}(i, j) = [x(i-1, j-1), x(i-1, j+1), x(i+1, j+1), x(i+1, j-1)]^T$$

$\mathbf{x}_{+}(i, j)$ is the axial model of the image which consists of four 4-connected neighbors of $x(i, j)$

$$\mathbf{x}_{+}(i, j) = [x(i, j-1), x(i-1, j), x(i, j+1), x(i+1, j)]^T$$

If we define $\mathbf{a}_{\times}$ and $\mathbf{a}_{+}$

$$\mathbf{a}_{\times} = \left[a_{\times}^1, a_{\times}^2, a_{\times}^3, a_{\times}^4\right]^T$$

$$\mathbf{a}_{+} = \left[a_{+}^1, a_{+}^2, a_{+}^3, a_{+}^4\right]^T$$

as diagonal and axial model parameters, the modified restoration problem in (2.16) can be written as

$$\min_{\mathbf{x},\mathbf{a}_\times,\mathbf{a}_+} \left\{ \sum_{(i,j)\in W} w_\times . \left(x(i,j) - \mathbf{a}_\times^T \mathbf{x}_\times(i,j)\right)^2 + \sum_{(i,j)\in W} w_+ . \left(x(i,j) - \mathbf{a}_+^T \mathbf{x}_+(i,j)\right)^2 \right\}$$

$$\text{such that} \tag{2.17}$$

$$\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon$$

where $w_\times$ and w_+ are optimal weights of the two models. The above equation can be converted to an unconstrained nonlinear least square problem which can be solved by an iterative algorithm of structured total least squares (STLS) [52].

In this thesis we applied NEAR restoration technique in proposed MARX and hybrid TV-MARX algorithms to recover CS reconstruction results. To reduce the artifacts generated in the recovered image, a stronger criterion function is defined in hybrid TV-MARX that mixes TV and PAR. Details of our algorithms will be discussed in the next chapter.

Chapter 3

Compressive Sensing Recovery via Piecewise Autoregressive Modeling

3.1 Introduction

In this chapter details of the MARX algorithm are presented. The first section describes model-based adaptive recovery of compressive sensing (MARX) algorithm. MARX uses piecewise autoregressive model (PAR) to recover the CS-acquired signals (images specifically in this thesis). It is an iterative process that jointly estimates the PAR model parameters and pixel values of recovered image. The second section discusses about how we improve the MARX algorithm by combining the TV prior with the PAR model. This hybrid TV-MARX CS recovery algorithm dynamically switches between TV and PAR depending on whether there is presence of edge(s) near the current pixel, and as such improves the reconstruction quality over either TV or PAR alone. It also uses weighted PAR parameters based on edge structure of a local window to further improve the results of recovery process. In order to run the hybrid algorithm on large image blocks, we use Fourier measurement matrix instead of Gaussian measurement matrix.

3.2 Model-based Adaptive Recovery of Compressive Sensing

As stated in the introduction the performance of current CS recovery techniques is poor compared with conventional coding techniques. This is because existing CS recovery techniques assume the signal to be a stationary, which is not true for natural images. In contrast, the proposed MARX algorithm is a locally adaptive algorithm that can recover the image in different sparse spaces by altering model parameters. For an image $\mathbf{x}$ the TV optimization for CS reconstruction is

$$\hat{\mathbf{x}} = argmin\,(TV(\mathbf{x})) \qquad \text{such that} \qquad \|\Phi\mathbf{x}-\mathbf{y}\|_2 \leq \epsilon \tag{3.1}$$

where Φ is the measurement matrix and $\mathbf{y}$ is the measurement vector.

The MARX algorithm considers the measurement matrix Φ as the matrix representation of a degrading function h and the output of TV reconstruction $\hat{\mathbf{x}}$ as the degraded image. It solves the following optimization problem to recover the original image from $\hat{\mathbf{x}}$

$$\min_{\mathbf{z},\mathbf{a}_\times,\mathbf{a}_+} \left\{ \sum_{(i,j)\in W} w_\times . \left(z(i,j) - \mathbf{a}_\times^T \mathbf{z}_\times(i,j)\right)^2 + \sum_{(i,j)\in W} w_+ . \left(z(i,j) - \mathbf{a}_+^T \mathbf{z}_+(i,j)\right)^2 \right\}$$

$$\text{such that} \tag{3.2}$$

$$\|\mathbf{\Phi z} - \hat{\mathbf{x}}\|_2 \leq \epsilon$$

where $\mathbf{z}$ is the restored image, $\mathbf{a}_\times$ and $\mathbf{a}_+$ are diagonal and axial PAR model parameters of pixel $z(i,j)$ and $\mathbf{z}_\times(i,j)$ and $\mathbf{z}_+(i,j)$ are four 8-connected and 4-connected neighbors of pixel $z(i,j)$ respectively. $w_\times$ and w_+ are the optimal weights of the two PAR models. The recovery algorithm is an iterative process that solves a nonlinear least square optimization problem to jointly estimate the unknowns $\mathbf{z}$, $\mathbf{a}_\times$ and $\mathbf{a}_+$.

This algorithm recovers CS-acquired image $\hat{\mathbf{x}}$ by solving the optimization problem defined in (3.2). The optimization problem can be rewritten in unconstrained nonlinear least square form as follows

$$\begin{aligned}\min_{\mathbf{z},\mathbf{a}_\times,\mathbf{a}_+} & \sum_{(i,j)\in W} w_\times \left(z(i,j) - \mathbf{a}_\times^T \mathbf{z}_\times(i,j)\right)^2 + \\ & \sum_{(i,j)\in W} w_+ \left(z(i,j) - \mathbf{a}_+^T \mathbf{z}_+(i,j)\right)^2 + \\ & \lambda \|\mathbf{\Phi z} - \hat{\mathbf{x}}\|_2^2\end{aligned}$$

The Lagrangian multiplier λ is adjusted such that the solution $\mathbf{z}$ satisfies $\|\mathbf{\Phi z} - \hat{\mathbf{x}}\|_2 \leq \epsilon$.

We can define matrices $\mathbf{C}_\times$ and $\mathbf{C}_+$ of size $W \times W$ to represent the above summations in the matrix form

$$\min_{\mathbf{z},\mathbf{a}_\times,\mathbf{a}_+} w_\times \|\mathbf{z} - \mathbf{C}_\times \mathbf{z}\|_2^2 + \|\mathbf{z} - \mathbf{C}_+ \mathbf{z}\|_2^2 + \lambda \|\mathbf{\Phi z} - \hat{\mathbf{x}}\|_2^2 \tag{3.3}$$

The elements of matrices $\mathbf{C}_\times$ and $\mathbf{C}_+$ can be determined from the elements of $\mathbf{a}_\times$ and

$\mathbf{a}_+$. Lets define residue vector $\mathbf{r}(\mathbf{z}, \mathbf{a}_\times, \mathbf{a}_+)$ as

$$\mathbf{r}(\mathbf{z}, \mathbf{a}_\times, \mathbf{a}_+) = \begin{bmatrix} \mathbf{r}_\times(\mathbf{z}, \mathbf{a}_\times) \\ \mathbf{r}_+(\mathbf{z}, \mathbf{a}_+) \\ \mathbf{r}_\Phi(\mathbf{z}) \end{bmatrix}$$

where

$$\mathbf{r}_\times(\mathbf{z}, \mathbf{a}_\times) = \sqrt{w_\times}(\mathbf{I} - \mathbf{C}_\times)\mathbf{z}$$

$$\mathbf{r}_+(\mathbf{z}, \mathbf{a}_+) = \sqrt{w_+}(\mathbf{I} - \mathbf{C}_+)\mathbf{z}$$

$$\mathbf{r}_\Phi(\mathbf{z}) = \sqrt{\lambda}(\hat{\mathbf{x}} - \mathbf{C}_\Phi\mathbf{z})$$

with this definition, equation (3.3) can be represented in quadratic from as follows

$$\min_{\mathbf{z}, \mathbf{a}_\times, \mathbf{a}_+} \mathbf{r}(\mathbf{z}, \mathbf{a}_\times, \mathbf{a}_+)^T \mathbf{r}(\mathbf{z}, \mathbf{a}_\times, \mathbf{a}_+) \tag{3.4}$$

The above problem can be solved using structured total least square (STLS) iterative algorithm. If $\Delta\mathbf{z}$, $\Delta\mathbf{a}_\times$ and $\Delta\mathbf{a}_+$ represent small changes in $\mathbf{z}$, $\mathbf{a}_\times$ and $\mathbf{a}_+$ respectively, then the linearized residue vector $\mathbf{r}(\mathbf{z}, \mathbf{a}_\times, \mathbf{a}_+)$ can be written as

$$\mathbf{r}(\mathbf{z}, \mathbf{a}_\times, \mathbf{a}_+) = \begin{bmatrix} \mathbf{r}_\times(\mathbf{z} + \Delta\mathbf{z}, \mathbf{a}_\times + \Delta\mathbf{a}_\times) \\ \mathbf{r}_+(\mathbf{z} + \Delta\mathbf{z}, \mathbf{a}_+) + \Delta\mathbf{a}_+ \\ \mathbf{r}_\Phi(\mathbf{z} + \Delta\mathbf{z}) \end{bmatrix} = \begin{bmatrix} \mathbf{r}_\times(\mathbf{z}, \mathbf{a}_\times) + \frac{\partial \mathbf{r}_\times}{\partial \mathbf{z}}\Delta\mathbf{z} + \frac{\partial \mathbf{r}_\times}{\partial \mathbf{a}_\times}\Delta\mathbf{a}_\times \\ \mathbf{r}_+(\mathbf{z}, \mathbf{a}_+) + \frac{\partial \mathbf{r}_+}{\partial \mathbf{z}}\Delta\mathbf{z} + \frac{\partial \mathbf{r}_+}{\partial \mathbf{a}_+}\Delta\mathbf{a}_+ \\ \mathbf{r}_\Phi(\mathbf{z}) + \frac{\partial \mathbf{r}_\Phi}{\partial \mathbf{z}}\Delta\mathbf{z} \end{bmatrix} \tag{3.5}$$

Therefore, given the current estimates of $\mathbf{z}$, $\mathbf{a}_\times$ and $\mathbf{a}_+$ the problem in (3.4) reduces to

$$\min_{\Delta\mathbf{z},\Delta\mathbf{a}_\times,\Delta\mathbf{a}_+} \left\| \begin{bmatrix} \frac{\partial \mathbf{r}_\times}{\partial \mathbf{z}} & \frac{\partial \mathbf{r}_\times}{\partial \mathbf{a}_\times} & 0 \\ \frac{\partial \mathbf{r}_+}{\partial \mathbf{z}} & 0 & \frac{\partial \mathbf{r}_+}{\partial \mathbf{a}_+} \\ \frac{\partial \mathbf{r}_\Phi}{\partial \mathbf{z}} & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{z} \\ \Delta\mathbf{a}_\times \\ \Delta\mathbf{a}_+ \end{bmatrix} + \begin{bmatrix} -\mathbf{r}_\times(\mathbf{z},\mathbf{a}_\times) \\ -\mathbf{r}_+(\mathbf{z},\mathbf{a}_+) \\ -\mathbf{r}_\Phi(\mathbf{z}) \end{bmatrix} \right\|_2^2 \tag{3.6}$$

For the next iteration estimated $\mathbf{z}$, $\mathbf{a}_\times$ and $\mathbf{a}_+$ are updated using $\Delta\mathbf{z}$, $\Delta\mathbf{a}_\times$ and $\Delta\mathbf{a}_+$ from (3.6).

The remaining question is what are the initial estimates for the image and PAR model parameters. The initial estimates for $\mathbf{a}_\times$ and $\mathbf{a}_+$ can be obtained from the degraded image $\hat{\mathbf{x}}$ by solving the following least square problems

$$\mathbf{a}_\times^0 = argmin \left\{ \sum_{(i,j)\in W} \left(\hat{x}(i,j) - \mathbf{a}_\times^T \mathbf{x}_\times(i,j)\right)^2 \right\}$$

$$\mathbf{a}_+^0 = argmin \left\{ \sum_{(i,j)\in W} \left(\hat{x}(i,j) - \mathbf{a}_+^T \mathbf{x}_+(i,j)\right)^2 \right\}$$

where $\mathbf{x}_\times$ and $\mathbf{x}_+$ are four 8-connected and 4-connected neighbors of $\hat{x}(i,j)$. The initial estimate for the image can be calculated using initialized PAR model parameters and CS reconstruction output $\hat{\mathbf{x}}$ as follows

$$\mathbf{z}^0 = argmin \left\| \begin{bmatrix} 0 \\ 0 \\ \lambda\hat{\mathbf{x}} \end{bmatrix} - \begin{bmatrix} -\mathbf{I} + \mathbf{C}_\times \\ -\mathbf{I} + \mathbf{C}_+ \\ \lambda\mathbf{C}_\Phi \end{bmatrix} \mathbf{z} \right\|_2^2 \tag{3.7}$$

Having the initial values for the image and two PAR model parameters, an iterative restoration process can be executed to recover the image that has been degraded by the measurement matrix. Even though the MARX recovery algorithm outperforms existing CS recovery methods by $0.5-3dB$, it can be further improved if we resolve the issues this algorithm suffers from. In the next section hybrid TV-MARX is proposed that improves the performance of MARX algorithm.

3.3 Hybrid TV-MARX Algorithm

As demonstrated in [9] the MARX algorithm drastically improves the quality of CS reconstruction. However, this algorithm has the following weaknesses

1. MARX algorithm uses two PAR models in diagonal and axial directions separately, and weights the two PAR models equally. Using the same weight for PAR models is not optimal since the structure of a dominant edge in a local context cannot be described by equal contributions of the PAR models. For example, suppose the current pixel in a local context is on a strong edge in the 45° direction. It is not optimal to assign same weight to the 90° PAR model as the model in the 45° direction.

2. The MARX algorithm is not robust against noises. In particular in smooth regions the MARX algorithm tends to over-fit noises and creates work like artifacts.

In Figure 3.1 the output of CS and MARX algorithms are shown. The weaknesses of MARX algorithm can be seen in this figure. Figure 3.1(c) clearly shows the worm-like artifacts created by MARX algorithm.

Proposed hybrid TV-MARX algorithm improves MARX performance by addressing these issues. In hybrid TV-MARX algorithm the first problem of MARX is addressed by giving more weight to the PAR model that is in the strong edge direction.

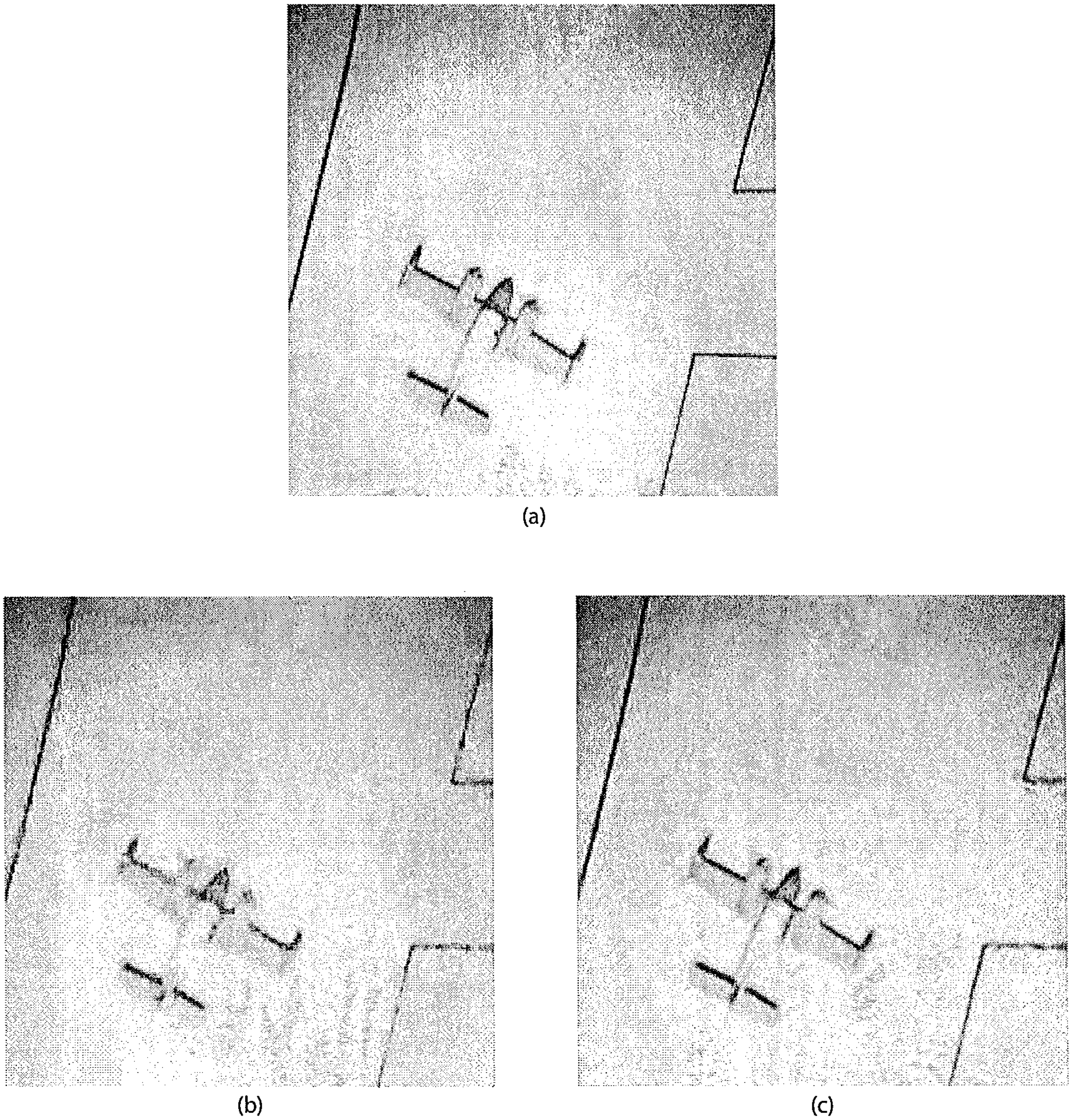

Figure 3.1: MARX drawbacks. (a) is the original 255×255 image. (b) is the output of TV reconstruction algorithm for CS. The block size used for this algorithm is 15×15. (c) is the output of MARX algorithm. The local window size is 5×5.

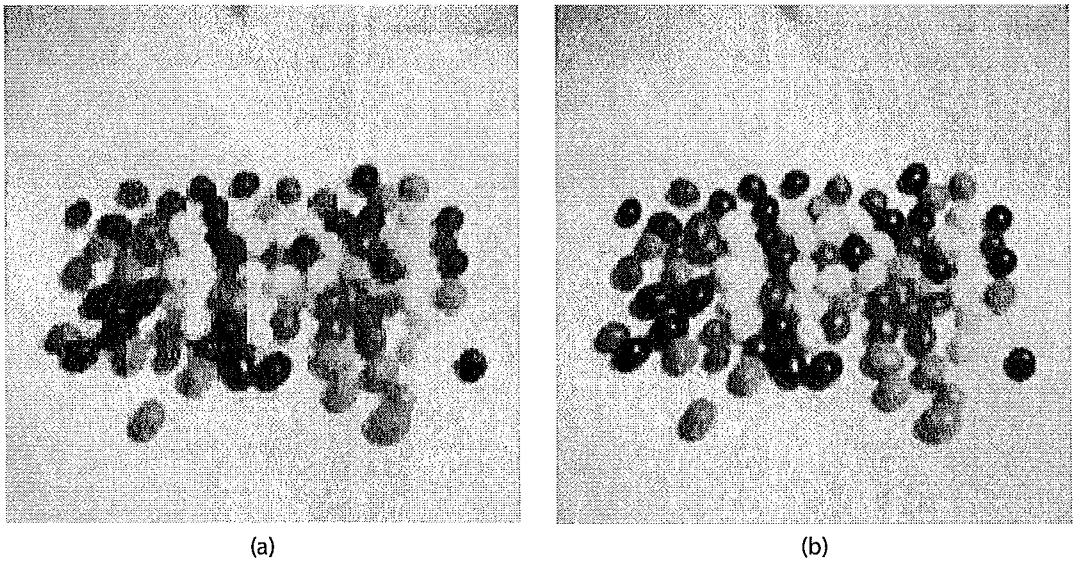

Figure 3.2: Blocking effect in TV reconstruction algorithm for CS. (a) The output of TV reconstruction algorithm for CS with small block size. The size of blocks is 15×15. (b) The output of TV reconstruction algorithm for CS with large blocks of size 128×128.

To do so, an edge detection algorithm is used on the CS acquired image. The edge detection algorithm finds the orientation and the strength of the edges for each pixel. First, the gradients of the pixels in horizontal and vertical directions are calculated by convolving the image with $[-1\ \ 0\ \ 1]$ and $[-1\ \ 0\ \ 1]^T$ masks respectively. Horizontal and vertical gradients of an image at position (i,j) can be calculated as

$$d_x(i,j) = x(i,j+1) - x(i,j-1)$$

$$d_y(i,j) = x(i+1,j) - x(i-1,j)$$

The orientation of the edge $\theta(i,j)$ and its gradient magnitude $r(i,j)$ are computed

$$\theta(i,j) = -\arctan(\frac{d_x(i,j)}{d_y(i,j)}) + \frac{\pi}{2}$$

$$\text{and} \tag{3.8}$$

$$r(i,j) = \sqrt{d_x(i,j)^2 + d_y(i,j)^2}$$

$r(i,j)$ and $\theta(i,j)$ are used along with Gabor functions to calculate weights for each PAR model in different directions. Gabor function is basically a Gaussian function modulated by a complex sinusoid

$$G(x,y,\theta,f) = \exp\left(-\frac{1}{2}(\frac{x'}{\sigma_{x'}})^2 + (\frac{y'}{\sigma_{y'}})^2\right)\cos(2\pi f x')$$

where

$$x' = x\cos(\theta) + y\sin(\theta)$$

$$y' = y\cos(\theta) - x\sin(\theta)$$

Four Gabor functions are in directions (0°, 45°, 90°, 135°) are used to compute the weights. The gradient directions are first rounded to the nearest 45° direction. Rounded gradient directions are then used to select the corresponding Gabor function. For example if the gradient direction of a pixel is 45° then the Gabor function with $\theta = 45°$ is used to compute the weights for the PAR model for that pixel. The least square problems that finds the initial weighted PAR models parameters can be written as

$$\mathbf{a}_\times^0 = argmin\left\{\sum_{(i,j)\in W}\left(g_c^d.\hat{x}(i,j) - \mathbf{a}_\times{}^T\mathbf{G}^d\mathbf{x}_\times(i,j)\right)^2\right\}$$

$$\tag{3.9}$$

$$\mathbf{a}_+^0 = armin\left\{\sum_{(i,j)\in W}\left(g_c^d.\hat{x}(i,j) - \mathbf{a}_+{}^T\mathbf{G}^d\mathbf{x}_+(i,j)\right)^2\right\}$$

where g_c^d is the weight given to the center pixel by the Gabor function and $\mathbf{G}^d$ is a diagonal matrix. The elements of main diagonal of $\mathbf{G}^d$ are the weights given to the neighborhood of the center pixel using the corresponding Gabor function. d in g_c^d and G^d specifies the gradient direction of the center pixel ($d \in \{0°, 45°, 90°, 135°\}$). For example if the gradient direction of a pixel is 45° then $\mathbf{G}^{45}$ matrix is used in (3.9). The elements in the main diagonal of $\mathbf{G}^{45}$ are computed using Gabor function with $\theta = 45°$. Solving updated least square problems for PAR models gives new initial PAR model parameters that are regularized based on the direction of the edge at each pixel.

To address the low performance of MARX algorithm on smooth regions, mixed TV and PAR are used in criterion function of the nonlinear least square problem. Hybrid TV-MARX algorithm dynamically switches between TV and PAR. If the neighborhood of the current pixel is smooth then TV is used but if the current pixel resides on a dominant edge, then PAR is used to recover that pixel value. To distinguish between these neighborhoods, the edge detection algorithm described before is used. The gradient directions are categorized in four groups: $(0°, 45°, 90°, 135°)$. If the gradient magnitude of the current pixel $x(i, j)$ in the local window is above a threshold then that pixel is marked as being on a strong edge. All other pixels that are not marked are either in a smooth or a noisy neighborhood.

During recovery process PAR model is used for the marked pixels. TV is used for all other pixels that has a smooth or noisy neighborhood. If we define $\mathbf{z}_D$ as the set of pixels in image $\hat{\mathbf{x}}$ that are on a strong edge and $\mathbf{z}_O$ as the set of all other pixels then, the constrained optimization problem for hybrid TV-MARX can be written as

$$\min_{\mathbf{z},\mathbf{a}_\times,\mathbf{a}_+} \left\{ \sum_{z(i,j)\in\mathbf{z}_D} \left(\sum_{(i,j)\in W} w_\times . \left(z(i,j) - \mathbf{a}_\times^T \mathbf{z}_\times(i,j)\right)^2 + \right.\right. \tag{3.10}$$

$$\left.\left. \sum_{(i,j)\in W} w_+ . \left(z(i,j) - \mathbf{a}_+^T \mathbf{z}_+(i,j)\right)^2 \right) + \sum_{z(i,j)\in\mathbf{z}_O} TV(z(i,j)) \right\}$$

such that

$$\|\mathbf{\Phi z} - \hat{\mathbf{x}}\|_2 \leq \|\epsilon\|_2$$

Hybrid TV-MARX algorithm addresses MARX issues using the above constrained optimization problem and the least square problems in (3.9).

Theoretically, CS reconstruction and MARX algorithms don't impose any limitation on the size of the image being processed. However, because of the computational constrains, it is impractical to work with the whole image. Therefore, the image has to be split in two smaller blocks and the CS reconstruction and MARX algorithms are executed on each block separately. Using small blocks creates blocking effect artifacts in the results of CS and MARX algorithms. The blocking effect is clearly visible in the result of CS reconstruction algorithm in Figure 3.1 and 3.2. The MARX algorithm can highly reduce the blocking effect in the recovered image (see Figure 3.1 (c)). However, since MARX algorithm is sensitive to the noises, the noise added to the results of CS algorithm due to the blocking effect degrades the performance of MARX algorithm.

The blocking effect is solved by changing Gaussian measurement to Fourier measurement matrix. Using Fourier measurement matrix instead of Gaussian measurement matrix makes it possible to increase the size of the blocks to the image size.

The reason is that one can define Fourier measurement matrix as a function. Since the measurement matrix is defined as a function, we don't need to store the whole matrix in the memory. Whenever we need to multiply the measurement matrix by a vector, the Fourier measurement function is called which returns the result of the multiplication. But for the Gaussian measurement matrix one needs to store the whole matrix in the memory. In such case, when the block sizes is large, the computational complexity of CS reconstruction increases drastically which makes the CS implementation impractical.

To clarify why Gaussian measurement matrix cannot be used on large blocks an example is provided. Suppose we want to use CS on an image of size 512×512. If the measurement ratio (the ratio of the number of measurements to the size of the image) is $r = 20\%$, then the number of measurements is $M = \lceil 0.20 \times 512 \times 512 \rceil = 52429$. To work with the whole image, i.e., using one large block of the image size, the size of the measurement matrix would be $52429 \times (512 \times 512) = 13743947776$. If the elements of the matrix are defined as double which takes 8 bytes, then the amount of memory needed to store measurement matrix is $13743947776 * 8 \approx 100 Gbytes$. It is clear that it is impractical from both physical and computational points of view to run the algorithms that require this amount of memory. Thus, the image has to be split into smaller blocks to make CS algorithm feasible to implement.

To apply Fourier measurement matrix in CS reconstruction step, we used the code written by Justin Romberg that defines the measurement matrix as a function. The new code can deal with large blocks using Fourier measurement matrix instead of Gaussian measurement. By using this code we were able to increase the size of the blocks to reduce the blocking effect in CS reconstruction. This change in CS algorithm removed the artifacts, which were related to the blocking effect, from the results of recovery step. Figure 3.2 shows the differences between the results of CS reconstruction algorithm with different block sizes. The output of CS reconstruction algorithm that uses blocks of size of 15×15 is shown in Figure 3.2(a). The blocks are visibly separated by boundaries due to the blocking effect. In contrast, Figure 3.2(b) shows the output of CS reconstruction algorithm that works with large blocks. Since it processes the whole image at once no blocking effect can be seen in the output image.

In the next chapter we compare the experimental results of MARX, hybrid TV-MARX and CS algorithms. The comparisons are conducted from both objective and subjective points of views.

Chapter 4

Experimental Results

In this chapter, we report experimental results of the proposed MARX and hybrid TV-MARX recovery algorithms and discuss our findings. To emphasize the importance of spatial adaptability of a CS recovery method and the effectiveness of MARX and hybrid TV-MARX, we conduct a comparative study between these algorithms and the total variation (TV) method. First we report the experimental results for the MARX algorithm. Considering that the CS method that uses Gaussian measurement matrix cannot handle large blocks due to high computational complexity, we divided the image into 15×15 blocks in our experiments. Table 4.1 is the comparison of the MARX algorithm with the TV method for the images in Figure 4.1. The number of measurements is 25% of the number of total pixels in each image ($M = 0.25 \times h \times w$ for an image of size $h \times w$). The curves of PSNR versus the number of CS measurements (presented as the percentage of the total number of pixels) are plotted in Figure 4.2 for the TV and MARX recovery methods. In the comparison, the MARX algorithm is clearly outperforming the TV method over all numbers of CS measurements as shown in Figure 4.2.

For an evaluation of tested methods in terms of visual quality Figures 4.3-4.6 are presented. These images confirm the superiority of the MARX algorithm in preserving the structures and details of edges and textures. The ability of the MARX algorithm in preserving the structure of edges is attributed to the underlying piecewise autoregressive model that adapts to spatially varying second-order statistics of image signal. However, the worm-like artifacts in the smooth areas of the images can be seen in Figures 4.3-4.6. The sensitivity of the MARX algorithm to the noises creates these artifacts in the recovered images. Also using small blocks during reconstruction decreases the quality of TV and MARX recovery methods.

Table 4.1: PSNR of TV and MARX recovery methods. The block size is 15×15 and measurement ratio is 25%. PSNR unit is dB.

Images	**TV**	**MARX**	**Improvement**
Boats	28.1446	28.9744	0.8298
House	31.0171	31.4712	0.4541
Old Plane	33.2489	35.2488	1.9999
Lena	28.7057	30.9070	2.2013
Barb	23.7171	26.4794	2.7623

To alleviate the blocking effect we need to increase the block size in TV reconstruction algorithm. For this we use Fourier measurement matrix instead of Gaussian measurement matrix, because the former has a far more computationally efficient implementation [64]. In our experiments we used the software provided in [64] to compare the results of TV and MARX recovery methods with small blocks versus the the ones with large blocks. The proposed hybrid TV-MARX algorithm is also compared with the other recovery methods. The block size for the recovery methods

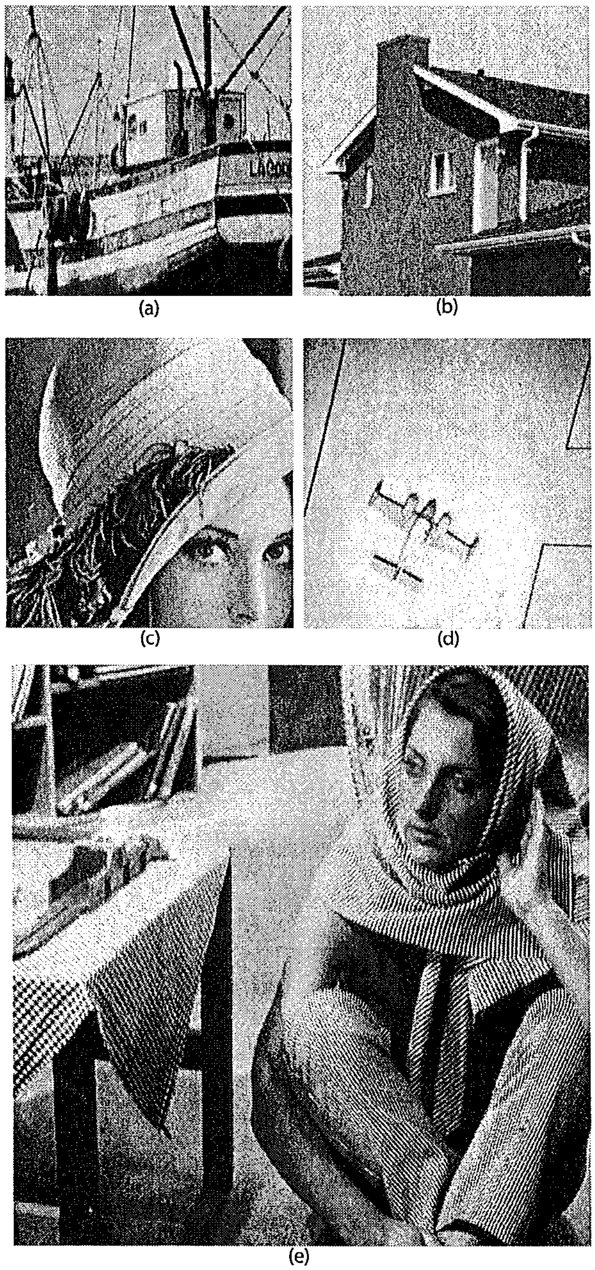

Figure 4.1: Images used to compare TV and MARX recovery algorithms. (a) *Boats* image of size 255 × 255. (b) *House* image of size 255 × 255. (c) *Old Plane* image of size 255 × 255. (d) *Lena* image of size 255 × 255. (e) *Barb* image of size 510 × 510.

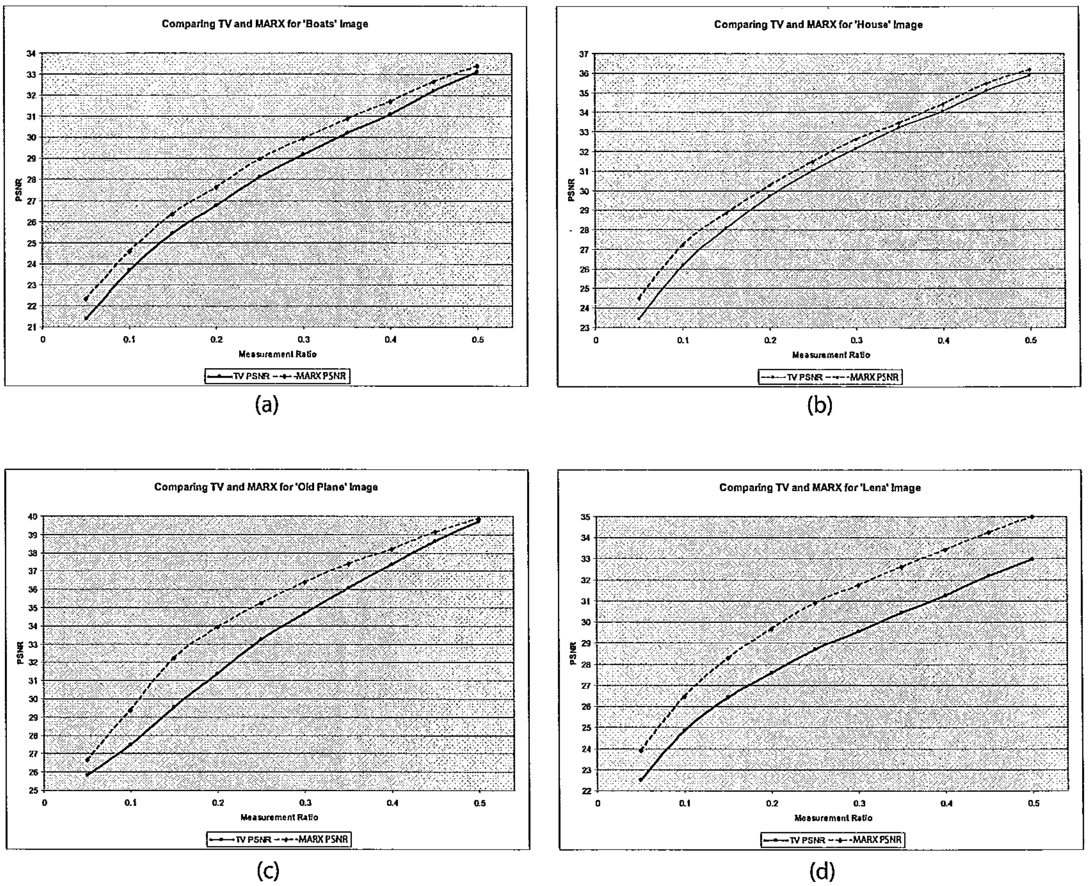

Figure 4.2: PSNR of TV and MARX recovery algorithms for four images with different measurement ratios.

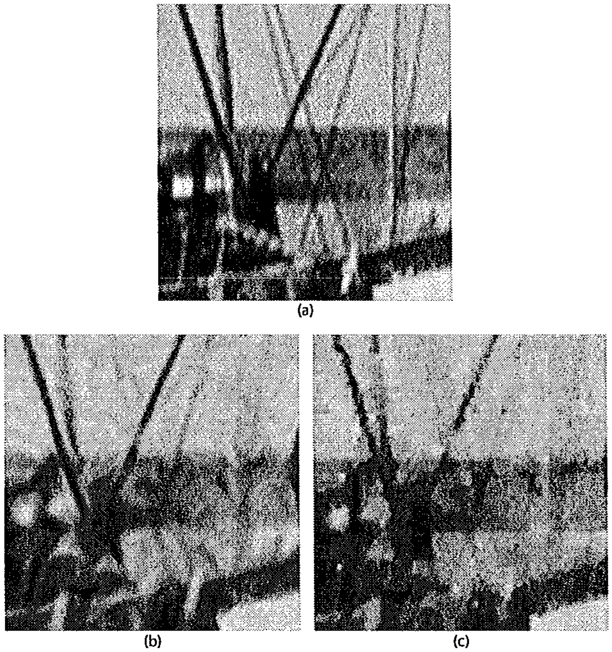

Figure 4.3: CS recovered *Boats* image (measurement ratio is 25%). (a) Original image. (b) MARX recovery. (c) TV recovery .

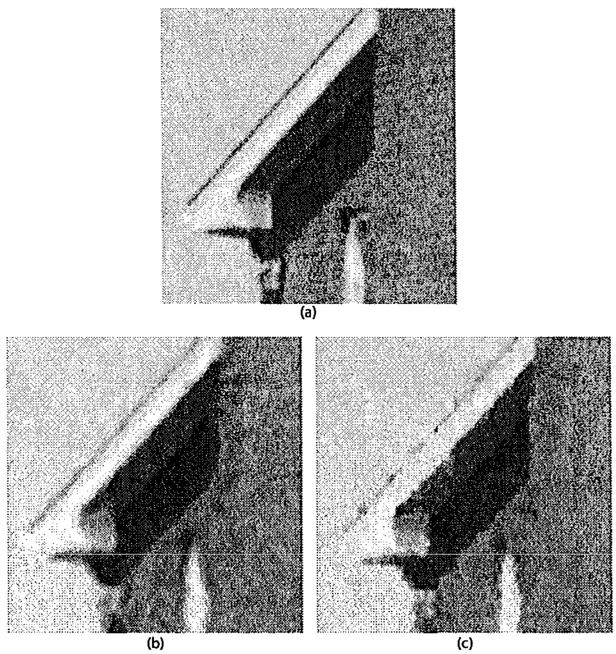

Figure 4.4: CS recovered *House* image (measurement ratio is 25%). (a) Original image. (b) MARX recovery. (c) TV recovery .

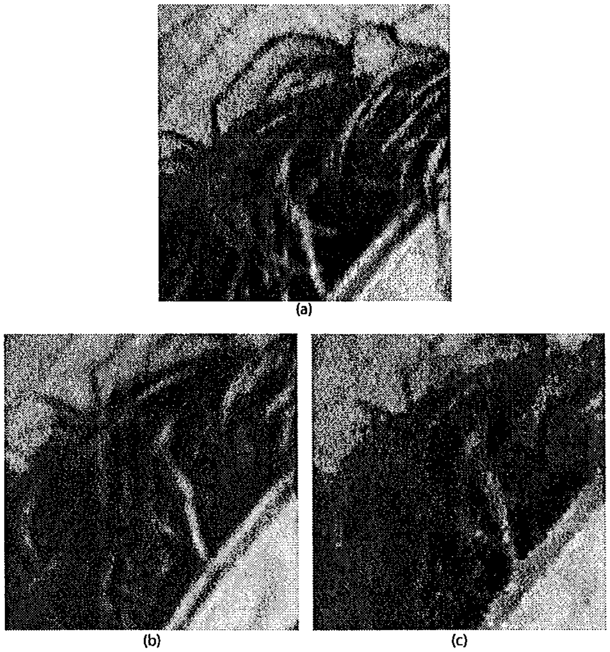

Figure 4.5: CS recovered *Lena* image (measurement ratio is 25%). (a) Original image. (b) MARX recovery. (c) TV recovery .

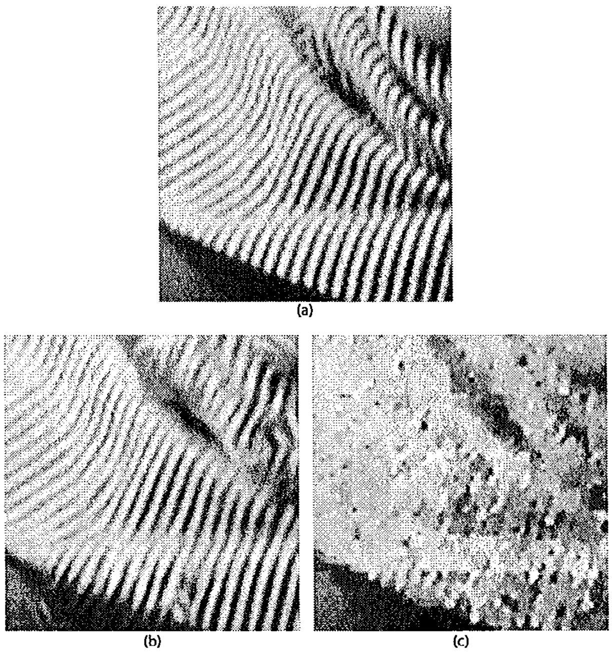

Figure 4.6: CS recovered *Barb* image (measurement ratio is 25%). (a) Original image. (b) MARX recovery. (c) TV recovery .

with small blocks is 15×15 and for the ones with large blocks, including hybrid TV-MARX method, is 128×128. Table 4.2 is the comparison of the five recovery methods (TV with small blocks, MARX with small blocks, TV with large blocks, MARX with large blocks and hybrid TV-MARX) for the images in Figure 4.1. The number of CS measurements is 25% of the number of total pixels in each image. The curves of PSNR versus the number of CS measurements are plotted in 4.8 for TV, MARX and hybrid TV-MARX recovery methods that uses large blocks. In comparison, the performance of TV recovery method is greatly improved when using large blocks nonetheless MARX algorithm performs better than TV in most cases (the PSNR of MARX is higher than the PSNR of TV by $1dB$ in average). However, it can be seen in Table 4.2 that the PSNR of MARX method is less than the PSNR of TV method by $0.15dB$ for *Old Plane* image. The large smooth areas in this image degrades the performance of MARX algorithm since MARX is sensitive to the noises in such areas. On the other hand, the performance of TV recovery method is increased since the blocking effect is removed due to the use of large block size.

Comparing the hybrid TV-MARX algorithm with other methods manifest the superiority of this hybrid algorithm. However, comparing the results of hybrid TV-MARX method with the results of MARX algorithm with large blocks shows that the improvement in the PSNR for the image *Lena* is not as good as the improvements for other images. The reason is that the edges in this image are not strong enough (look at the texture of the hat and furs). This makes it hard for the edge detection algorithm to find the edges. To solve this problem a more sophisticated edge detection algorithm should be used in the hybrid method. On the other hand, the superiority of the hybrid TV-MARX approach in preserving the structures and details of edges and textures is clear in *Barb* image. This is because in contrast to MARX algorithm, hybrid TV-MARX uses weighted PAR models that gives more weight to the PAR models in the dominant edge direction which preserves the edge structures better than equally weighted PAR models.

Figures 4.9-4.12 are presented for an evaluation of tested methods in terms of visual quality. The worm-like artifacts created by MARX algorithm are reduced in hybrid method (see Figures 4.9-4.11) and edge structures are preserved better in hybrid TV-MARX algorithm compared to other recovery methods(See Figure 4.12).

Table 4.2: PSNR of TV and MARX with small and large blocks and hybrid TV-MARX. Measurement ratio is 25%. PSNR unit is in dB.

Images	TV With Small Blocks	MARX With Small Blocks	TV With Large Blocks	MARX With Large Blocks	Hybrid TV-MARX
Boats	28.1446	28.9744	30.1539	31.0645	31.2975
House	31.0171	31.4712	34.4211	34.3413	34.7080
Old Plane	33.2489	35.2488	40.4268	40.2790	40.8630
Lena	28.7057	30.9070	30.3692	32.1493	32.0865
Barb	23.7171	26.4794	24.6116	27.5200	28.3940
Average	28.9667	30.6161	31.9965	33.0708	33.4698

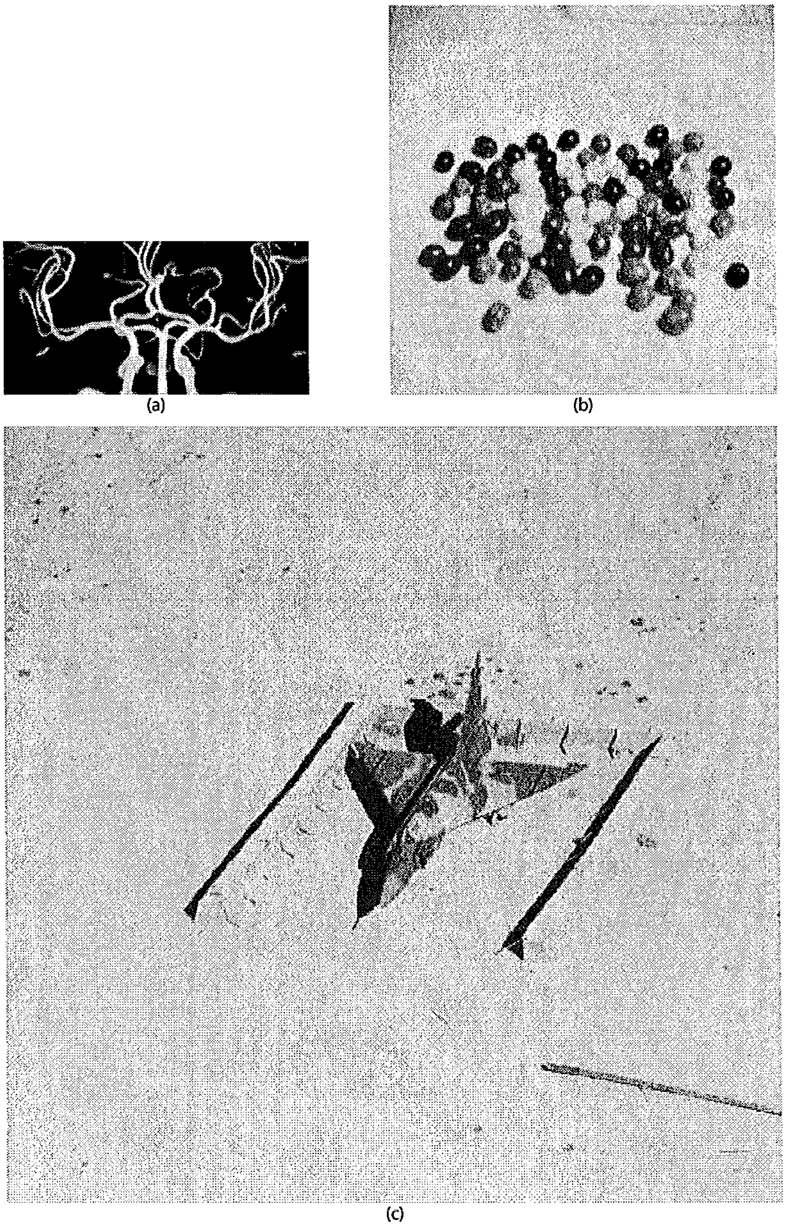

Figure 4.7: Original images used to Compare CS recovery methods. (a) *Inflow* image of size 224 × 112. (b) *Jelly Beans* image of size 256 × 256. (c) *Plane* image of size 512 × 512.

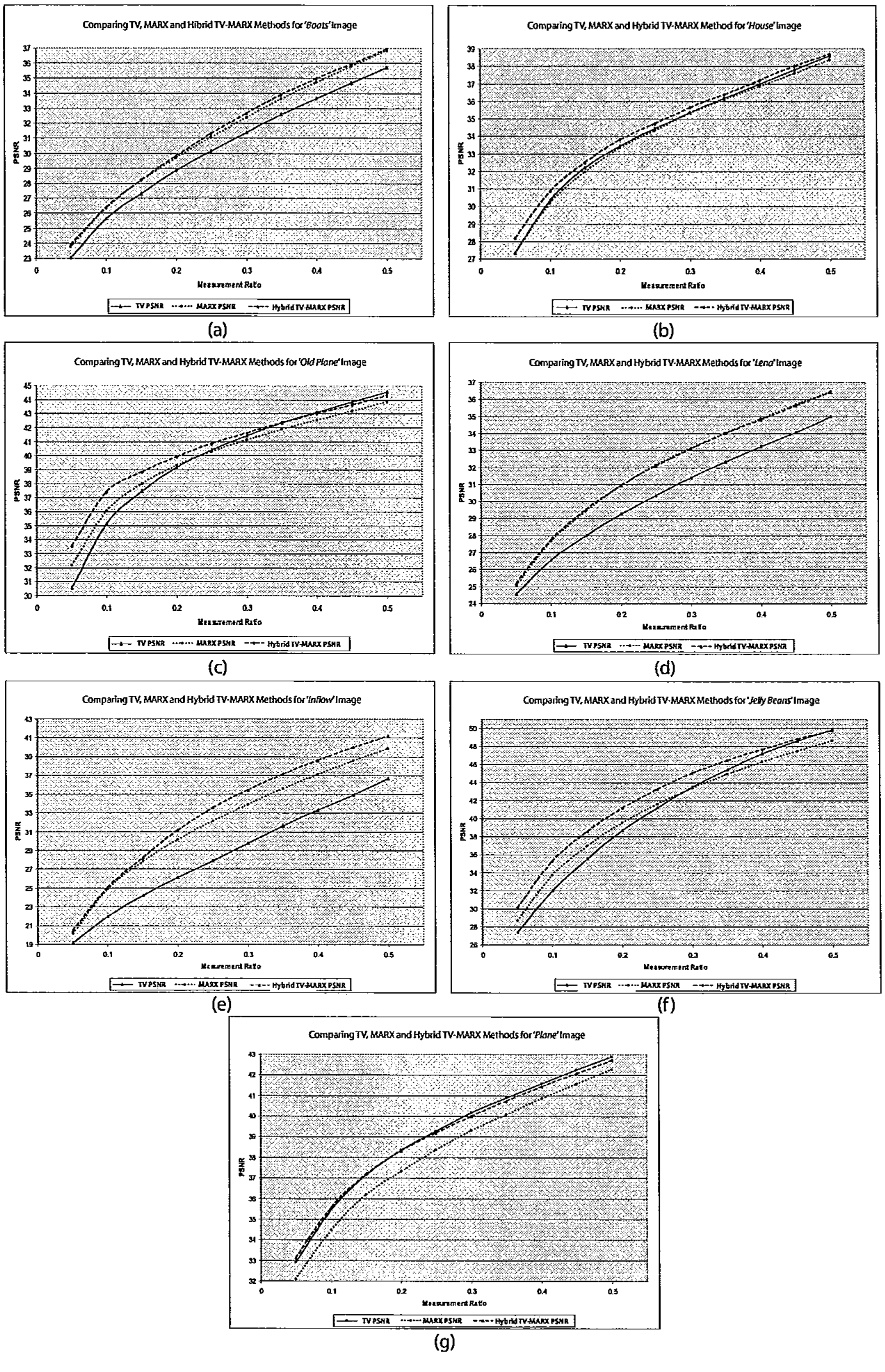

Figure 4.8: PSNR of TV, MARX and hybrid TV-MARX recovery methods with large blocks with different measurement ratios.

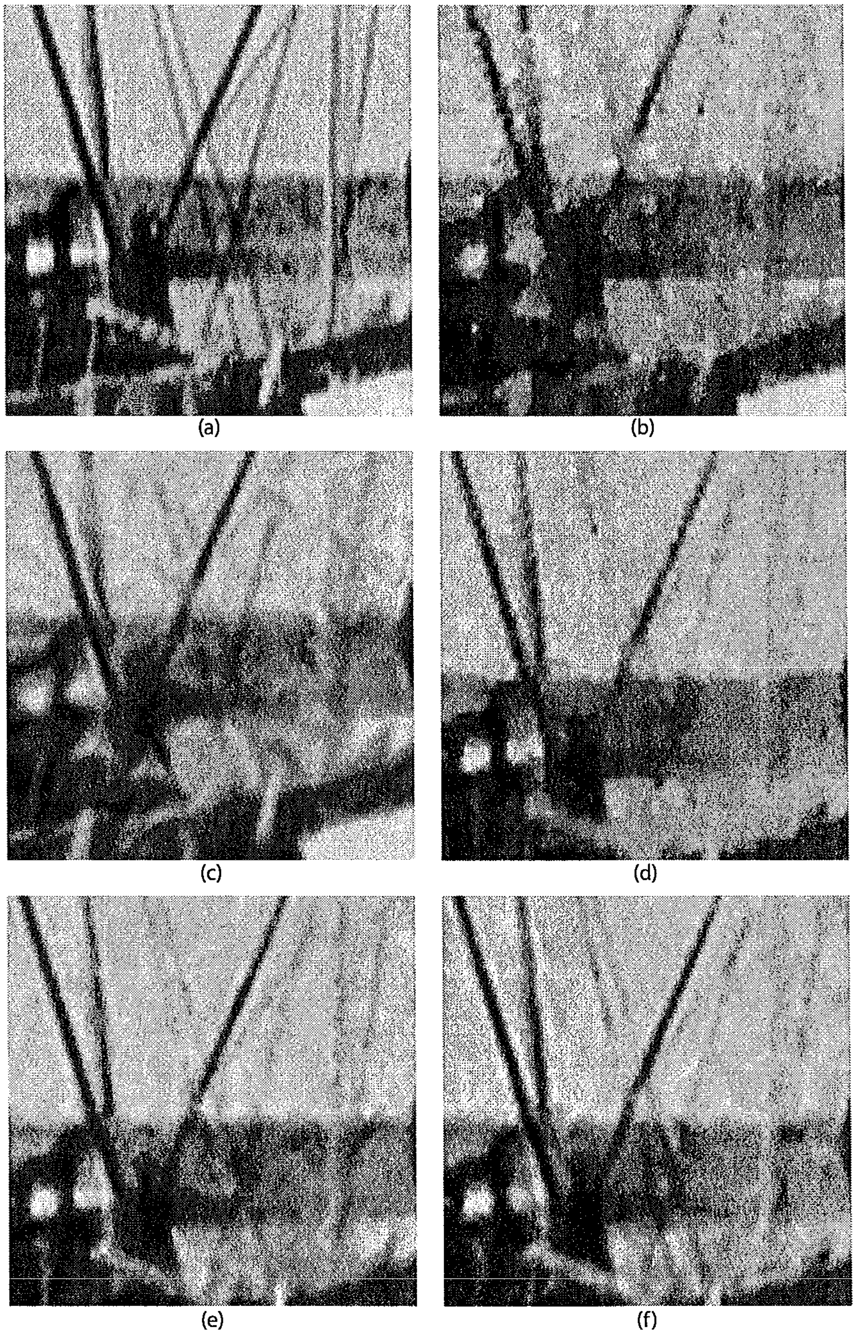

Figure 4.9: CS recovered *Boats* image (measurement ratio is 25%). (a) Original image. (b) TV recovery with small blocks. (c) MARX recovery with small blocks. (d) TV recovery with large blocks. (e) MARX recovery with large blocks. (f) Hybrid TV-MARX recovery

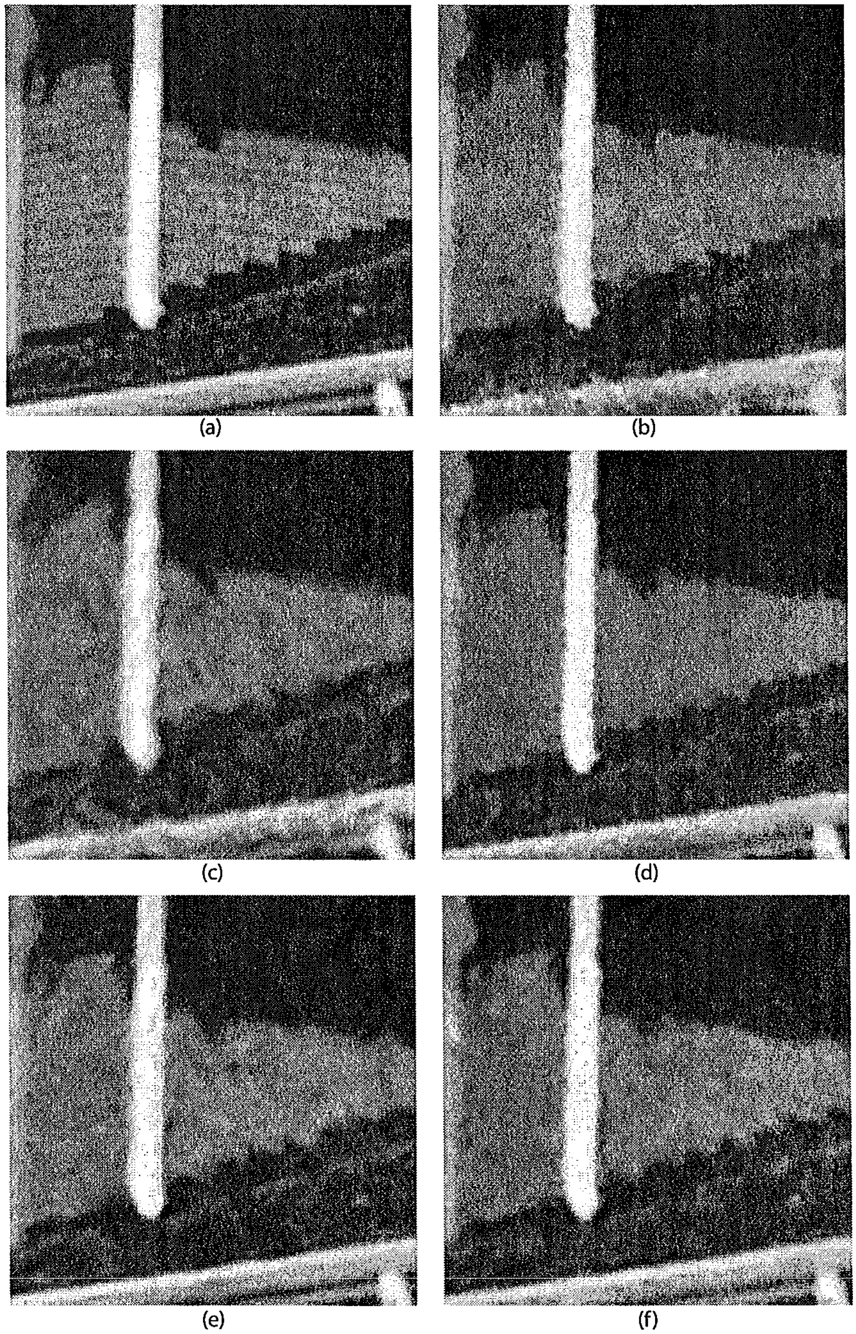

Figure 4.10: CS recovered *House* image (measurement ratio is 25%). (a) Original image. (b) TV recovery with small blocks. (c) MARX recovery with small blocks. (d) TV recovery with large blocks. (e) MARX recovery with large blocks. (f) Hybrid TV-MARX recovery

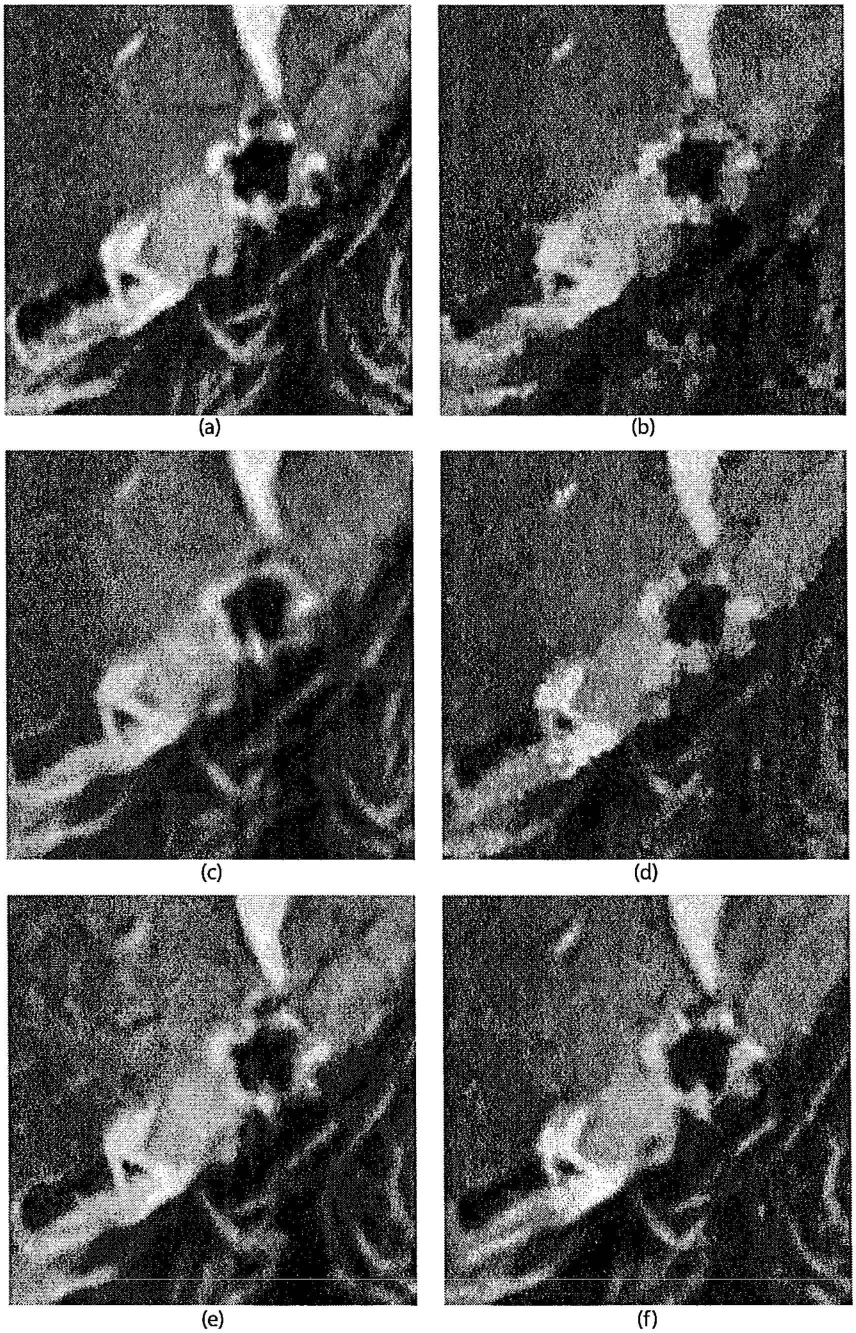

Figure 4.11: CS recovered *Lena* image (measurement ratio is 25%). (a) Original image. (b) TV recovery with small blocks. (c) MARX recovery with small blocks. (d) TV recovery with large blocks. (e) MARX recovery with large blocks. (f) Hybrid TV-MARX recovery

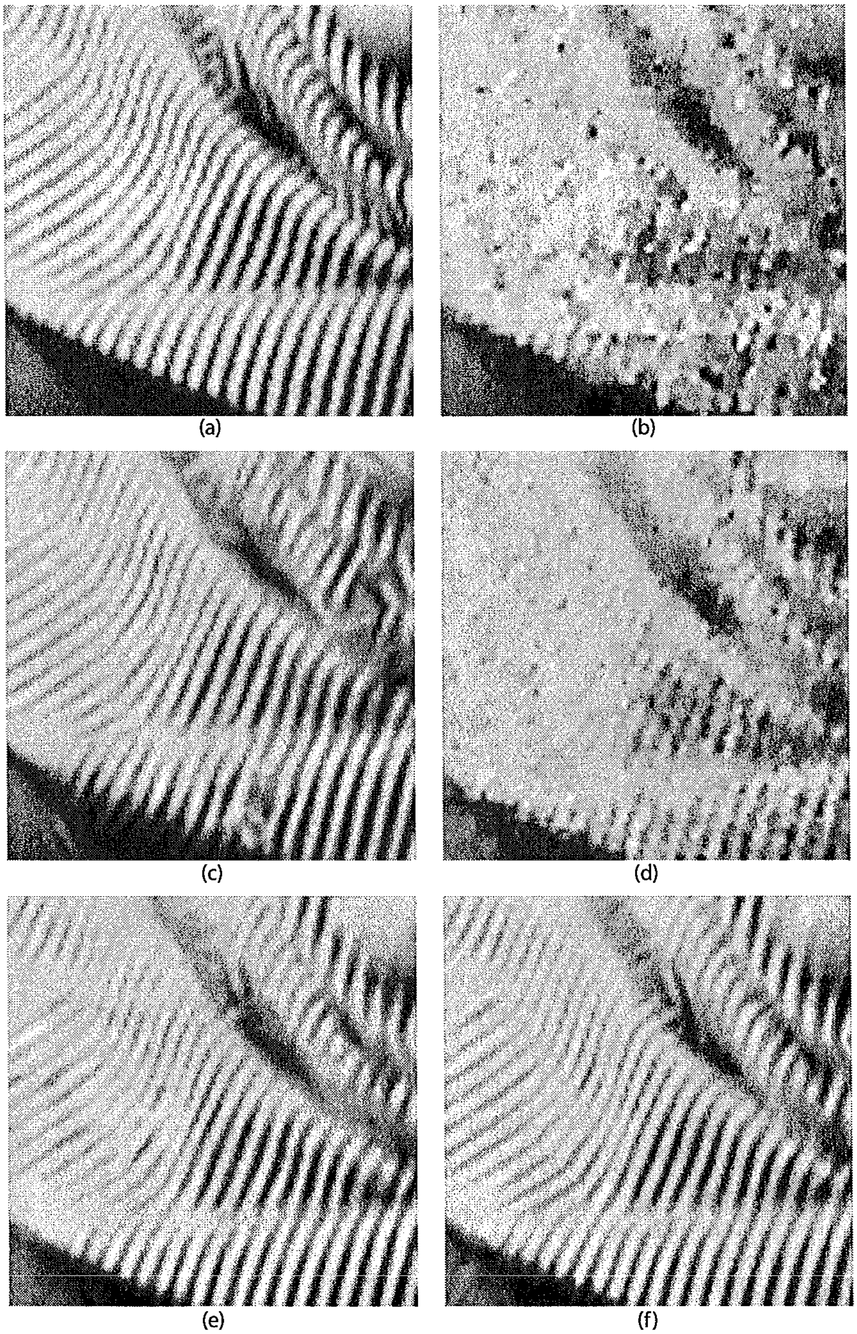

Figure 4.12: CS recovered *Barb* image (measurement ratio is 25%). (a) Original image. (b) TV recovery with small blocks. (c) MARX recovery with small blocks. (d) TV recovery with large blocks. (e) MARX recovery with large blocks. (f) Hybrid TV-MARX recovery

More experiments are conducted to compare the results of hybrid TV-MARX, TV and MARX recovery methods with large blocks. Table 4.3 is the comparison of the TV, MARX and hybrid TV-MARX recovery methods for the images in Figure 4.7. The number of CS measurements is 25% of the number of total pixels in each image. Figures 4.13 - 4.15 are presented for the comparison of the visual quality of these recovery methods. Again, these images recovered from CS sampling by the three methods manifest the superiority of the hybrid TV-MARX performance.

Even though the gain in the PSNR of the recovered *Plane* image for hybrid method is not extensive compared to the PSNR of TV recovery method, however the visual quality of recovered image for hybrid TV-MARX is better than the TV method (see Figure 4.15 (b) and (d)).

Table 4.3: PSNR of TV, MARX and hybrid TV-MARX with large blocks. Measurement ratio is 25%. PSNR unit is in dB.

Images	TV	MARX	hybrid TV-MARX	Improvement
Jelly Beans	41.2091	41.5937	43.2553	1.6616
Inflow	27.9172	32.1083	33.5621	1.4538
Plane	39.2904	38.3716	39.1874	0.8158

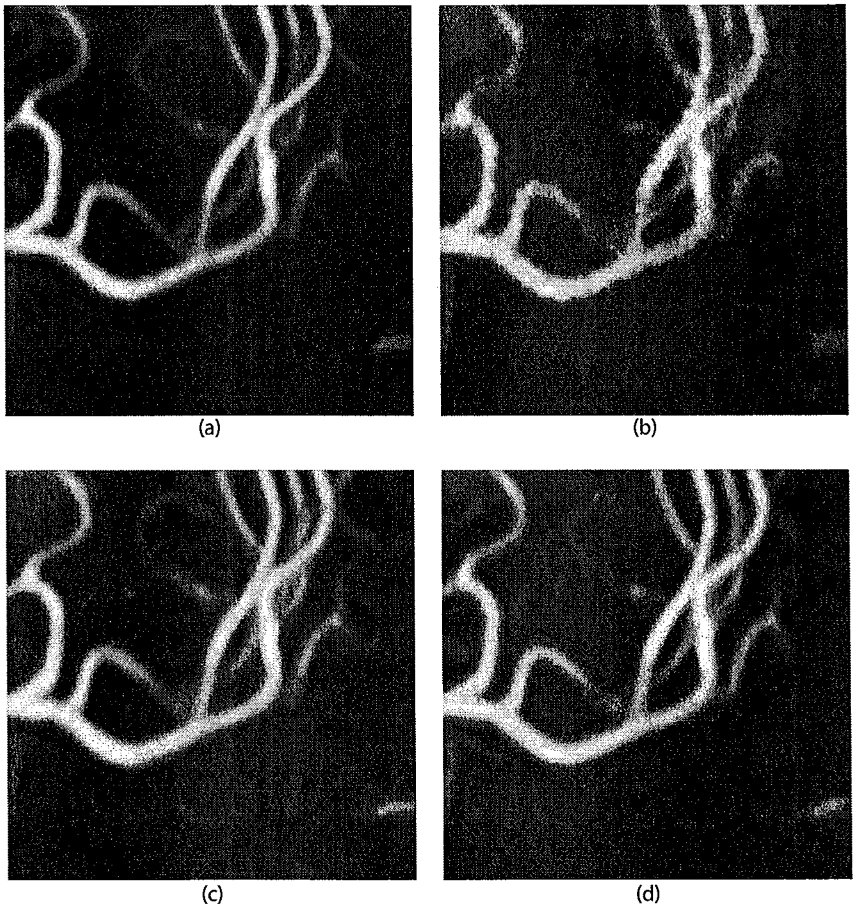

Figure 4.13: CS recovered *In Flow* image (measurement ratio is 25%). (a) Original image. (b) TV recovery. (c) MARX recovery. (d) Hybrid TV-MARX recovery

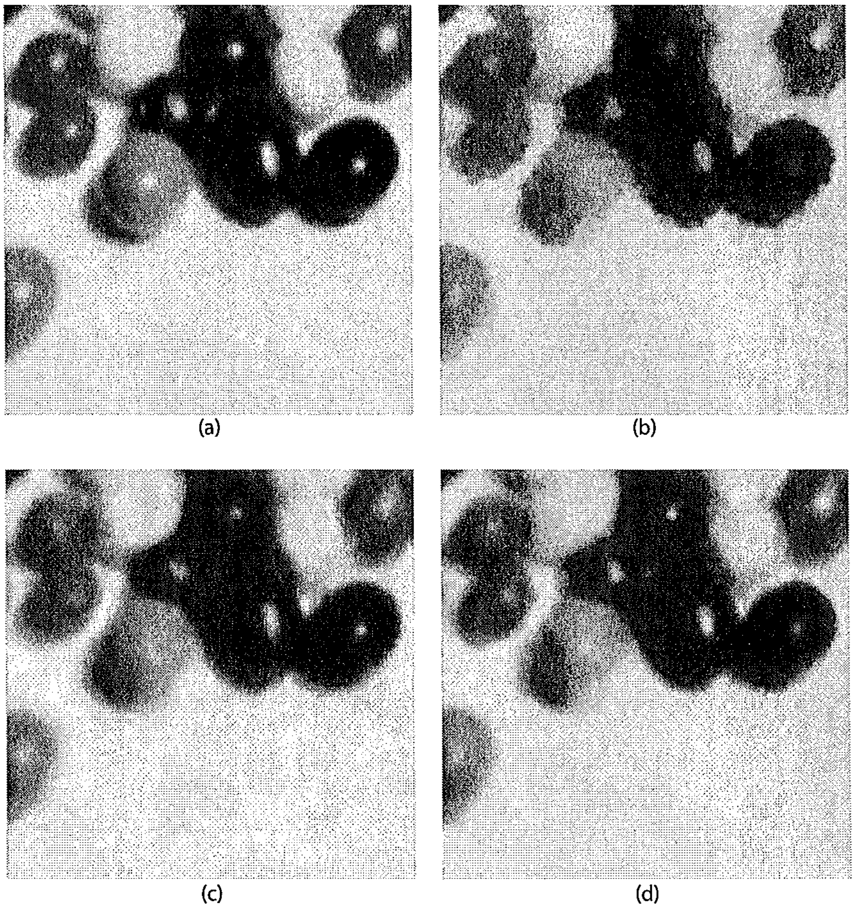

Figure 4.14: CS recovered *Jelly Beans* image (measurement ratio is 25%). (a) Original image. (b) TV recovery. (c) MARX recovery. (d) Hybrid TV-MARX recovery

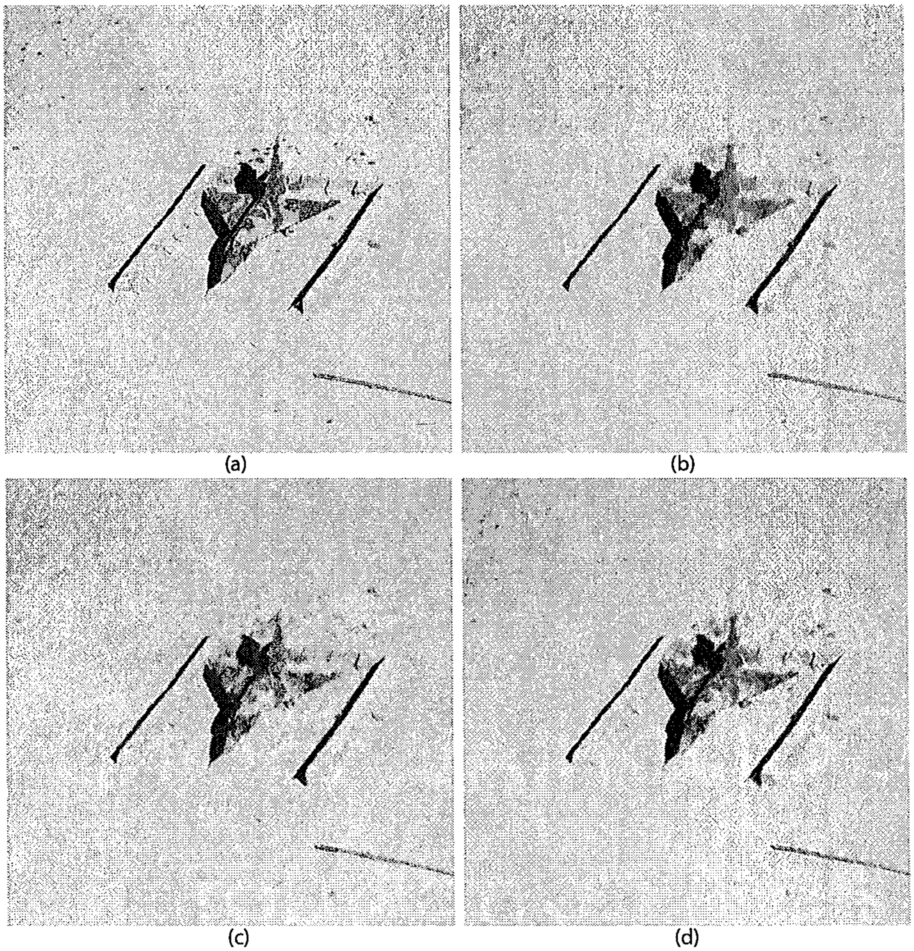

Figure 4.15: CS recovered *Plane* image (measurement ratio is 25%). (a) Original image. (b) TV recovery. (c) MARX recovery. (d) Hybrid TV-MARX recovery

Chapter 5

Conclusion

In this thesis we investigated the problem of recovering CS-acquired images and proposed two CS recovery methods: MARX and Hybrid TV-MARX. Both techniques belong to the class of edge-directed image restoration methods.

In the proposed MARX algorithm, a diagonal and an axial PAR models are used to constrain the image to be recovered. The model parameters are jointly estimated with the pixels in a constrained nonlinear least square process. This nonlinear estimation problem is solved by an iterative structured total least square (STLS) algorithm.

The hybrid TV-MARX algorithm improves the performance of MARX by adaptive weighting of multiple PAR models and by choosing between the TV and PAR constraints depending on the context of the current pixel. It can preserve edge structures better than MARX and eliminate worm-like artifacts that plague the MARX algorithm in smooth regions. The adaptive weights on different directional estimates are computed using Gabor functions.

The proposed model-based recovery algorithms are implemented and tested on a wide range of images. Experimental results show superior performance of these algorithms over the popular TV-based CS recovery method.

Bibliography

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

[2] E. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics*, vol. 6, no. 2, pp. 227–254, 2006.

[3] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[4] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2005.

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

[6] E. Gurses and O. Akan, "Multimedia communication in wireless sensor networks," *Annals of Telecommunications*, vol. 60, no. 7-8, pp. 799–827, 2005.

[7] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, R. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Proceedings of the 10th European Conference on Computer Vision*, 2008.

[8] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magezine*, vol. 24, no. 4, pp. 118–124, 2007.

[9] X. Wu, X. Zhang, and J. Wang, "Model-guided adaptive recovery of compressive sensing," in *Proceedings of the 2009 Data Compression Conference*, pp. 123–132, 2009.

[10] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall, 2000.

[11] A. Oppenheim, R. Schafer, and J. Buck, *Discrete-Time Signal Processing*, ch. 5. Prentice Hall, 1989.

[12] C. Shannon, "Communication in the presence of noise," *Proceedings of Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, 1949.

[13] J. M. Whittaker, "The fourier theory of the cardinal functions," *Proceedings of the Edinburgh Mathematical Society*, vol. 1, no. 3, pp. 169–176, 1929.

[14] V. Kotelnikov, "On the transmission capacity of "ether" and wire in electrocommunications," *Izd. Red. Upr. Svyazzi RKKA*, 1933.

[15] M. Unser, "Sampling-50 years after shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[16] M. Khan, R. Goodall, and R. Dixon, "Implementation of non-uniform sampling for alias-free processing in digital control," in *Proceedings of the UKACC International Conference on Control*, 2008.

[17] P. Albertos and A. Crespo, "Real-time control of non-uniformly sampled systems," *Control Engineering Practice*, vol. 7, no. 4, pp. 445–458, 1999.

[18] D. Carrica, M. Benedetti, and R. Petrocelli, "Random sampling applied to the measurement of a dc signal immersed in noise," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 5, pp. 1319–1323, 2001.

[19] F. Filicori, G. Iuculano, A. Menchetti, and D. Mirri, "Random asynchronous sampling strategy for measurement instruments based on nonlinear signal conversion," in *IEEE Proceedings Science, Measurement and Technology*, pp. 141–150, 1989.

[20] F. Marvasti, *Non-uniform Sampling: Theory and Practice.* Springer, 2001.

[21] I. Bilinskis, *Digital Alias-free Signal Processing.* John Wiley and Sons Ltd, 2007.

[22] I. Bilinskis and A. Mikelsons, *Randomized Signal Processing.* Prentice Hall, 1992.

[23] E. Candes and J. Romberg, "Encoding the l_p ball from limited measurements," in *Data Compression Conference*, 2006.

[24] E. Candes and J. Romberg, "Practical signal recovery from random projections," in *Proceedings of SPIE Conference*, 2005.

[25] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[26] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, pp. 4203–4215, 2005.

[27] E. Candes, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *Annual Symposium on Foundations of Computer Science*, 2005.

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

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

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

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

[32] Y. Tsaig and D. Donoho, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–571, 2006.

[33] D. Donoho and Y. Tsaig, "Fast solution of l_1-norm minimization problems when the solution may be sparse," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4789–4812, 2008.

[34] D. Donoho, I. Drori, Y. Tsaig, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," tech. rep., Department of Statistics, Stanford University, 2006.

[35] D. Donoho, "For most large underdetermined systems of linear equations, the minimal l_1 norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[36] D. Donoho and J. Tanner, "Sparse nonnegative solutions of underdetermined linear equations by linear programming," *Proceedings of the National Academy of Sciences*, vol. 102, no. 27, pp. 9446–9451, 2005.

[37] M. Lustig, D. Donoho, and J. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[38] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[39] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," in *Conference on Information Science and Systems*, 2008.

[40] P. Boufounos and R. Baraniuk, "Reconstructing sparse signals from their zero crossings," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[41] M. Duarte, M. Wakin, D. Baron, and R. Baraniuk, "Universal distributed sensing via random projections," in *Proceedings of the 5th International Conference on Information Processing in Sensor Networks*, 2006.

[42] M. Duarte, M. Wakin, and R. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," in *Proceedings of the Workshop on Signal Processing with Adaptative Sparse Structured Representations*, 2005.

[43] B. Hayes, "The best bits," *American Scientist*, vol. 97, no. 4, pp. 276–280, 2009.

[44] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *preprint*, 2007.

[45] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *oundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[46] S. Kim, K. Koh, M. Lustig, and S. Boyd, "An efficient method for compressed sensing," in *Proceedings of IEEE International Conference on Image Processing*, pp. 117–120, 2007.

[47] A. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, "Near-optimal sparse fourier representations via sampling," in *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pp. 152–161, 2002.

[48] A. Gilbert, S. Muthukrishnan, and M. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 243–252, 2003.

[49] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, "Algorithmic linear dimension reduction in the l_1 norm for sparse vectors," in *44th Annual Allerton Conference of Communication, Control, and Computing*, 2006.

[50] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, "One sketch for all: Fast algorithms for compressed sensing," in *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pp. 237–246, 2007.

[51] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2008.

[52] X. Wu and X. Zhang, "Model-based non-linear estimation for adaptive image restoration," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1185–1188, 2009.

[53] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1999.

[54] A. Purandare, "Jpeg2000 compression." http://www.geocities.com/ee00224/btp3.htm, September 2009.

[55] E. Candes, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, 2006.

[56] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[57] J. Romberg, "Imaging via compressive sampling [introduction to compressive sampling and recovery via convex programming]," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.

[58] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[59] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[60] E. Candes and M. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *IEEE Signal Processing Magazine*, pp. 21–30, 2008.

[61] I. Leonid, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[62] R. Gonzalez and R. Woods, *Digital Image Processing*, ch. 5. Prentice Hall, 2007.

[63] X. Wu, E. Barthel, and W.Zhang, "Piecewise 2d autoregression for predictive image coding," in *1998 International Conference on Image Processing*, pp. 901–904, 1998.

[64] J. Romberg, "L1-magic." http://www.acm.caltech.edu/l1magic/, June 2009.