Data-guided statistical sparse measurements modeling for compressive sensing

by

Tal Shimon Schwartz

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Digital image acquisition can be a time consuming process for situations where high spatial resolution is required. As such, optimizing the acquisition mechanism is of high importance for many measurement applications. Acquiring such data through a dynamically small subset of measurement locations can address this problem. In such a case, the measured information can be regarded as incomplete, which necessitates the application of special reconstruction tools to recover the original data set. The reconstruction can be performed based on the concept of sparse signal representation. Recovering signals and images from their sub-Nyquist measurements forms the core idea of compressive sensing (CS). In this work, a CS-based data-guided statistical sparse measurements method is presented, implemented and evaluated. This method significantly improves image reconstruction from sparse measurements. In the data-guided statistical sparse measurements approach, signal sampling distribution is optimized for improving image reconstruction performance. The sampling distribution is based on underlying data rather than the commonly used uniform random distribution. The optimal sampling pattern probability is accomplished by learning process through two methods - direct and indirect. The direct method is implemented for learning a nonparametric probability density function directly from the dataset. The indirect learning method is implemented for cases where a mapping between extracted features and the probability density function is required. The unified model is implemented for different representation domains, including frequency domain and spatial domain. Experiments were performed for multiple applications such as optical coherence tomography, bridge structure vibration, robotic vision, 3D laser range measurements and fluorescence microscopy. Results show that the data-guided statistical sparse measurements method significantly outperforms the conventional CS reconstruction performance. Data-guided statistical sparse measurements method achieves much higher reconstruction signal-to-noise ratio for the same compression rate as the conventional CS. Alternatively, Data-guided statistical sparse measurements method achieves similar reconstruction signal-to-noise ratio as the conventional CS with significantly fewer samples.

Acknowledgements

I wish to thank my supervisors Prof. David A. Clausi and Prof. Alexander Wong for their guidance and professional support throughout my time under their supervision. They have contributed tremendously through this research as well as to my growth as a researcher. I wish to thank my committee members who kindly agreed to act on the examining committee. Also I wish to thank Yeast Resource Center - University of Washington in Seattle for the test fluorescence microscopy data, National Research Council Canada (NRC), The Brown University pattern theory group for the test laser range data, Prof. S. Narasimhan and Mr. A. Sadhu from the department of civil engineering at the University of Waterloo for the bridge structure vibration data and Prof. K. Bizheva, department of Physics and Astronomy, University of Waterloo for the OCT data.

Dedication

I wish to dedicate this Doctoral dissertation to my parents, Haviva and Yehoshua Schwartz z"l, who are proudly looking after me from heaven. This work would not have been possible without the love and support of my family, Orit, Rotem, Ofir and Nitzan Schwartz.

Table of Contents

List of Figures

Glossary

$f(r,c)$: two-dimensional gray scale digital image, $f(r,c) \in \mathbb{R}^{R \times C}$

f : lexicographically concatenated version of image $f(r,c)$, $f \in \mathbb{R}^{N \times 1},\ N = RC$

y : M length data measurement vector of image f, $y \in \mathbb{R}^{M \times 1}$

Ψ : K-sparse representation domain, $\Psi \in \mathbb{R}^{N \times N}$

x : f is represented by coefficients vector x in domain Ψ

Φ : sampling basis, $\Phi \in \mathbb{R}^{N \times M}$ whose columns are $\{\phi_k\}_{k=1}^{M}$

φ_k : k^{th} column in sampling basis Φ

$\varphi_{T,k}$: k^{th} sparse column in sampling basis Φ with T nonzero elements

Φ^t : sampling basis transpose. $\Phi^t \in \mathbb{R}^{M \times N}$

A : measuring matrix. $A = \Phi^t \Psi,\ A \in \mathbb{R}^{M \times N}$

ϵ : denotes measurement additive noise

$\Omega_{R \times C}$: denotes $R \times C$ sampling locations organized in a finite, separable, rectangular lattice

Ω_K : denotes K sampling locations in the k space domain organized in a finite sequence

Ω_T : denotes sampled locations subset whose cardinality is equal to T

Ω_T^c : denotes unsampled locations subset whose cardinality is equal to $RC - T$

λ : denotes the regularization constant

$E\{.\}$: denotes average

$\Gamma(r,c)$: function that provides a quantitative measure of saliency at sampling location (r,c)

p_T : sampling basis probability density

$S(r,c)$: saliency map for all locations (r,c)

θ_i : saliency map is segmented into R clusters θ_i, $i = \{1, \ldots, R\}$ with common saliency properties

Chapter 1

Introduction

1.1 Compressive Sensing Domain

The last five decades have witnessed a digital revolution which set the foundation for numerous sensing systems with ever-increasing performance and resolution. Shannon's sampling theory has had a great impact on signal processing and communications [131, 132, 148]. The sampling theory claims that a band-limited function can be fully recovered from its samples if the sampling rate is greater then twice the highest frequency of the signal, the so called Nyquist rate [102]. The sampling theorem can be interpreted as coefficients of some basis functions, for example, through shifting and rescaling of multiplication of sinc-functions and the sampled signal in Euclidean space [148]. The reconstruction is exact when the sampled function is band-limited and sampled at the appropriate rate. The sampling function may be forced to be band-limited by multiplying the function with a box function in the frequency domain or alternatively convolving with a sinc function in the time domain. This can be an "engineering" view of the sampling theory.

Based on Shannon's sampling theory, analog signal processing moved almost entirely to the digital domain creating high fidelity, inexpensive and robust digital systems. With this incredible advance of digital technology signal processing needs to compute, store and communicate huge amounts of data. With bandwidth and resolution increases in many emerging applications, sampling rates have become extremely high, creating expensive and sometimes not feasible systems that cannot acquire samples at Nyquist rate [102]. Lossy data compression can be used for applications where some level of fidelity degradation is acceptable. The core idea of data compression involves only preserving the largest coefficients of a signal. This process is called sparse signal approximation and the concept

is in the core of common compression standards such as JPEG, JPEG2000, MPEG, and others [41]. Even though date compression can address challenges such as reducing memory space, computation time, and communication load, the acquisition must meet Nyquist rate requirements for appropriate sampling. It would be desirable if one could sample and compress at the same time rather than sample and only then compress.

An approach with great potential for reducing the sampling rate while maintaining high reconstruction quality is the utilization of compressive sensing (CS) theory, which allows for a greatly reduced sampling rate through the use of sparse measurements (samples). The theory of CS provides an unambiguous proof that discrete signals, which admit a sparse representation in the domain of a suitable linear transform, can be accurately recovered from their sub-Nyquist measurements [8, 19, 18, 42, 51]. Such reconstruction can be performed through solution of a convex optimization problem, which maximizes the sparseness of the signal representation coefficients subject to a set of measurement constraints. In such cases, the measurement constraints are derived from a signal/image acquisition model, which normally describes the measurements as a (noise-contaminated) sequence of inner products between the quantity of interest and the elements of a sampling basis.

Even though CS is a relatively new theory and is still growing, it has already had positive impact on multiple applications. CS implementations are highly related to the acquisition hardware. For example, medical imaging measured directly in a transform domain such as magnetic resonance imaging (MRI) [86, 147] and computed tomography (CT) [27] has a sparse representation in some basis, such as wavelet and sub-sampled in the frequency domain. This sub-sampling saves scanning time and in the case of CT reduces x-ray radiation dose due to fewer cross-section samples. Sensor networks [40, 68, 152] benefit from CS by improving sensor network efficiency by energy and communication traffic reduction. Other applications use dedicated measuring hardware system for vector data acquisition implementing compressive sensing method [61], compressive sensing analog to digital converter [96, 109], and wide-band compressive sensing signal acquisition [142, 144]. A single pixel camera was developed [48] to perform measurements according to a random sampling method using a digital micro-mirror device (DMD).
It is noted, that even though CS has great potential to improve multiple systems, CS is not the solution for every problem. In some systems, where signal acquisition is not the bottleneck, compression techniques might be appropriate. In some hardware configurations, random sampling cannot be implemented currently, therefore CS might be considered for next generation systems only.

1.2 Motivation and Challenges

The research on CS is very active and its core research can be grouped into three major research areas.

1. **Sparsity reconstruction basis**: A priori sparsity knowledge in some basis of the target image is required to select an appropriate reconstruction basis. This area is well researched, and based on early results [16, 19, 45, 56, 57, 64], it has been shown that a sufficiently sparse representation of a signal can give a unique solution for the signal reconstruction problem.

2. **Image reconstruction**: Reconstruction techniques based on convex optimization and greedy methods are commonly used to recover the original signal. Greedy methods [145, 146], L_1 minimization [16, 45, 56, 57], and other methods [21] are used to recover the original signal from sparsest measurements. These methods have been well studied.

3. **Generalization of the sampling procedure:** The common sampling procedure in the CS research community is random sensing matrices. Common sensing matrices are based on Gaussian distributions and discrete Bernoulli distributions [95] sampling the entire image uniformly. Those sensing distributions satisfy the restricted isometry property (RIP) with a controlled probability [8, 20, 22, 43, 122]. Improvements in this area have been minimal compared to the other two areas.

While less studied in recent years compared to reconstruction basis and image reconstruction methods, the design of the sampling procedure can have a significant impact on CS performance for practical applications such as robotic vision, fluorescence microscopy, optical coherence tomography and laser range measurement, where the objects of interest have structured characteristics, thus making the sampling procedure a worthwhile area to study. Existing CS-based systems employ a sampling scheme that sample the entire scene in the same manner regardless of the underlying data. However, such an approach is limited for many practical applications, which involve distinct regions of interest in some basis, with highly salient structural characteristics. Given that such regions are of greater interest for analysis purposes, one is motivated to obtain higher quality reconstructions for these regions than the background regions.

The rest of this thesis is organized as follows. Overview of compressive sensing and relevant literature is reviewed in section 2 providing the background and up to date research. The unified sparse measurements model and framework is provided in 3 describing the

model for constructing a sampling pattern, conventional CS concept as well as data-guided CS method concept, methods for learning data distribution to guide sampling and method to construct the sampling pattern. Construct a sampling pattern based on the nonparametric modeling - direct learning approach is provided in 4 describing direct learning model and realization of energy-guided CS modeling approach in frequency domain as well as presenting experimental results for optical coherence tomography (OCT) and bridge structure vibrations applications are presented and discussed. While indirect learning approach is presented in 5 describing indirect learning model in the spatial domain, realization and experimental results of saliency-guided modeling approach in spatial domain are shown and discussed for applications such as 3D laser measurements, robotic vision, fluorescence microscopy. Summary of contributions and future research are presented in 6

1.3 Thesis Contributions

The principle research objective that leads to a significant contribution to the research community, is a system engineering approach for improving CS reconstruction performance by reducing the number of samples and improving the image reconstruction signal to noise ratio.

The thesis makes the following main contributions -

- A novel model to robust sparse image measurements based on learning data statistical properties, which is described in detail in Chapter 3, is designed, implemented and tested.
- A novel method to construct optimized sampling patterns based on two learning methods:
  - direct method, for learning a nonparametric probability density function directly from the dataset, which is described in detail in Chapter 4; and
  - indirect learning method, for cases where a mapping between extracted features and the probability density function is required, which is described in detail in Chapter 5

Chapter 2

Overview of Compressive Sensing

In 2006, it was proven [18, 19, 42, 51] that through the theory of compressive sensing, discrete signals can be fully recovered from their sub-Nyquist measurements if the signal has a sparse representation in some domain (2.1). Such reconstruction can be performed by solving an optimization problem for maximizing the signal sparseness in terms of representative coefficients, subject to the measurement model constraint (2.2). The signal/image acquisition model is usually described as a sequence of inner products between the signal measurements and the sampling basis (2.3). Some standard choices of such sampling bases include pseudo-random sequences (e.g., Gaussian, Bernoulli, etc.) (2.5), which are often preferred for the strong theoretical guarantees they provide [8, 20].

2.1 Sparse Representations

The concept of sparse representations is an important method in signal processing and is in the core of CS theory. To find an efficient representation, one can express a signal as the linear combination of a few elements taken in a specific basis. For example, natural images have sparse representation in some basis [23, 120]. Wavelets [140] and total variation [9, 51] are widely used for representation and compression of images [51, 120]. Images can be represented in some basis by only a few non-zero coefficients. Despite its great power, the sparsity concept is simply and intuitive. Let f be an N-vector real signal. To analyze two-dimensional images, a vector can be chosen that contains all N pixels (column-wise). Suppose that the basis $\Psi = \{\psi_1, \psi_2, \dots \psi_N\}$, provides an N-dimensional representation of f:

$$f = \Psi x \tag{2.1}$$

This is a linear combination of N vectors from Ψ. where f is an $N \times 1$ vector, the basis matrix Ψ is $N \times N$ with the basis vectors ψ_n as columns in matrix Ψ, and where x is an $N \times 1$ coefficients vector. When a signal has K-sparse representation in Ψ, $K \ll N$, one can represent signal f by linear combination of K vectors only, chosen from Ψ, and discard the zero coefficients without loss [45, 56]. The K vectors are chosen from a large collection called a dictionary. In practical cases it is not realistic to require:

$$x_n = 0; \ \forall \ n > K \tag{2.2}$$

Therefore this assumption is replaced by a weaker notion of sparsity called "compressibility" where the coefficients decay rapidly to zero but not equal to zero [22, 64]:

$$|x_n| \leq \alpha n^{-\beta}; \ \forall \ n > K \tag{2.3}$$

where $\alpha > 0$ and $\beta > 0$

In this case one can approximate the K-term to $\widetilde{f_K}$:

$$\left\| f - \widetilde{f_K} \right\|_2 < \epsilon_K \tag{2.4}$$

for some approximation error ϵ_K.

Sparse representations are commonly used as a method for data compression by minimizing the number of vectors (K) in the representation domain. In a broader view, signals can be modeled efficiently based on sparsity properties [42]. The hope is that a very sparse representation exists among all possible representations. In other words, a representation with just a few nonzero coefficients. The search for sparse representation can be performed using norm ℓ_p where $0 < p < \infty$ [30, 42, 52]. It was shown [42, 52] that as we move from convenient ℓ_2 regularization towards $0 < p \leq 1$ we promote sparser solutions. The tendency of ℓ_p norm to drive sparser results can be illustrated by considering the following problem [19, 52]:

$$\arg\min_x \|x\|_p \ \ s.t. \ f = \Psi x \tag{2.5}$$

Intuitively, one can solve the problem by "blowing" an $\|x\|_p$ balloon centered around the origin until reaching the constraint $f = \Psi x$ [52]. The ℓ_2 norm does not promote sparsity while ℓ_p with the range $0 < p \leq 1$ does promote sparsity [52, 64]. Moreover, the extreme ℓ_p norm, where $p \to 0$, provides the ultimate sparsest solution. This norm is denoted as ℓ_0 [38, 52]:

$$\|x\|_0 = \lim_{p \to 0} \|x\|_p^p = \lim_{p \to 0} \sum_{n=1}^{N} |x_n|^p = \#\left[i \forall x_i \neq 0\right] \tag{2.6}$$

In words, 2.6 is an intuitive measure of vector x sparsity containing the number of nonzero entries in the vector [52]. Even though the ℓ_0 norm is the common term used, pseudo-ℓ_0 is more appropriate as the ℓ_0 norm cannot be calculated according to norm definition [52]. The first problematic point is taking p^{th} root; it is impossible to take the 0^{th} root. In addition, the homogeneity property of a norm is not met:

$$\|tx\|_0 = \|x\|_0 \neq t \|x\|_0 \tag{2.7}$$

2.2 Compressive Sensing Signal Reconstruction Model

As concluded in (2.1), ℓ_0 norm is a good sparsify operator for the following named exact sparse optimization:

$$\arg\min_x \|x\|_0 \;\; s.t. \; y = Ax \tag{2.8}$$

where x is an $N \times 1$ vector and A is $M \times N$ matrix. In addition to high sparsity property, the exact sparse (2.8) optimization leads to a unique solution for x if A is an over-complete orthogonal basis which is selected properly [45]. Unfortunately, there are some significant challenges solving exact sparse optimization (2.8) using the ℓ_0 norm. The exact sparse problem leads to a combinatorial optimization problem with exponential complexity [100]. The solution would require searching among all possible combinations of columns of A. It was proposed [30] to consider a modified optimization using ℓ_1 norm called basis pursuit (BP):

$$\arg\min_x \|x\|_1 \;\; s.t. \; y = Ax \tag{2.9}$$

BP optimization (2.9) offers a significant advantage. The ℓ_1 norm is a convex function that supports a global minimum, unlike $\ell_p; 0 < p < 1$ that is non-convex. BP is a much simpler problem compared to exact sparse which can be solved by various classical optimization techniques. A symmetry between exact sparse (ℓ_0 norm based) and BP (ℓ_1 norm based) was shown [45, 56, 64], making an important contribution. If a signal has a sufficiently sparse representation then it has a unique solution for both the exact sparse and BP problems. Before addressing the condition for this symmetry, an important principle needs to be introduced - the uncertainty principle. The classical uncertainty principle states that two conjugate variables such as position and momentum cannot both be known with arbitrary precision. Therefore, a signal cannot be tightly concentrated both in time and frequency domains and actually there is lower boundary on product of the variance in time and frequency [52]. A similar condition exists between discrete uncertainty principles and sparse approximation [44, 45, 64, 145]. A signal cannot be sparsely represented in

both time and frequency domain, and actually this applies to any two arbitrary orthogonal domains [56]. Suppose we have a non-zero vector b and two orthogonal bases Ψ and Φ. Assume b can be represented as a linear combination of vectors of Ψ or Φ:

$$b = \Psi\alpha = \Phi\beta \tag{2.10}$$

For any arbitrary basis Ψ or Φ, either α or β can be sparse but not both at the same time [53]. This fact depends on the mutual-coherence between Ψ or Φ. The mutual-coherence $\mu(A)$, where A is constructed from an arbitrary pair of orthogonal basis Ψ or Φ such as $A = \Phi^t\Psi$, is defined as the maximal inner product between columns from those two bases [45]. Let Φ be the sampling basis: $\Phi = \{\phi_1, \dots \phi_N\}$ of function f. The inner product representation of the sampled signal y is given by $y_k =< f, \phi_k >$ and the representation basis is Ψ: $\Psi = \{\psi_1, \dots \psi_N\}$. The coherence between the sensing basis Φ and the representation basis Ψ is defined as:

$$proximity(\Phi, \Psi) = \mu(A) = \max_{1 \le i,j \le n} \left|\left\langle \phi_i^t, \psi_j \right\rangle\right| \tag{2.11}$$

where ϕ_i^t is column i transpose of Φ and ψ_j is column j of Ψ and where $< \cdot\ ,\ \cdot >$ is inner product. The mutual-coherence of such two matrices satisfies $\frac{1}{\sqrt{n}} \le \mu(A) \le 1$. In compressive sensing, it is desired to select high coherence pair (low incoherence) of sampling basis and representation bases. In the case of classical time-frequency sampling domains, Φ is "impulse" basis (Dirac comb [46]) and Ψ is Fourier basis. Apparently this pair has minimal coherence and is therefore not appropriate for compressive sensing. Based on mutual-coherence property and previous work [45] it was shown [56] that if a signal has a sufficiently sparse representation which satisfies (2.12):

$$\|x\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(A)}) \tag{2.12}$$

then the optimization solution for both the exact sparse (2.8) and BP (2.9) problems is unique. Therefore, in this case, the non-convex program (2.8) can be replaced with the much simpler and convex (2.9) problem.

In practical cases where we intend to recover a vector from incomplete and contaminated measurements than:

$$y = Ax + \epsilon \tag{2.13}$$

where A is a matrix with fewer rows than columns ($M \ll N$) and ϵ represents additive noise. x can be recovered by solving the following convex quadratic optimization known as basis pursuit denoising (BPDN) [19]:

$$\arg\min_x \|x\|_1 \ \ s.t.\ \ \|y - Ax\|_2 \le \epsilon \tag{2.14}$$

Efficient iterative methods for solving the optimization has been proposed , such as matching pursuit (MP) [90], iterative thresholding [39], orthogonal matching pursuit (OMP) [104, 143] and others [54]. Many of those are based on the concept of greedy algorithms [145]. The greedy algorithms search for a sparse vector x that optimally represent the measured data in the sparse domain. The iterative process starts by setting an initial value to x (usually 0) and adding a residue parameter r_s to the measured data y. The algorithm searches for a column in matrix A that best correlates to the data with residue r_s. The value is set according to the selected column and the residue parameter r_S is updated accordingly. This process repeats iteratively for minimizing the energy of the residual value [54].

Soft-thresholding is a popular approach for solving BPDN (2.14). The regularization constant λ is used as a signal threshold for noise filtering. Therefore, λ is been set according to the noise level.

One approach to solve the BPDN problem is using the augmented Lagrangian form [54]:

$$L = \arg\min_x \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1 \tag{2.15}$$

where the first term is squared to be differentiable while not affecting the minimization and the multiplication by $\frac{1}{2}$ is absorbed within λ. Since A is an orthonormal basis, one can use the Parseval equality to express the minimization by coefficients only:

$$L = \arg\min_x \frac{1}{2} \|\beta - x\|_2^2 + \lambda \|x\|_1 \tag{2.16}$$

where $\beta = A^{-1}y$ or in a summation form:

$$L = \arg\min_x \left[\frac{1}{2} \sum_{i=1}^{N} (\beta_i - x_i)^2 + \lambda \sum_{i=1}^{N} |x_i| \right] \tag{2.17}$$

The minimization can be performed for each i coefficient separately by solving $\frac{\partial L}{\partial x_i} = 0$ The solution to this problem, x_i^* is given by soft-thresholding [54]

$$x_i^* = sign(\beta_i)(|\beta_i| - \lambda)_+ \tag{2.18}$$

where

$$(t)_+ = \begin{cases} t, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases} \tag{2.19}$$

The total variation (TV) minimization approach has been shown to be an efficient regularizing criterion for piecewise smoothness in the spatial domain [19, 86]. Therefore, BPDN (2.14) can be improved:

$$\arg\min_{x} \|x\|_1 + \lambda \|x\|_{TV} \;\; s.t. \;\; \|y - Ax\|_2 \leq \epsilon \tag{2.20}$$

where the first term represents sparsity of coefficients in the selected domain and the second term represents piecewise smoothness in the spatial domain subject to the data fidelity term.

2.3 CS Sampling Model for Images

Consider the signal being measured as an image contain $R \times C$ sampling locations organized in a finite, separable, rectangular lattice $\Omega_{R\times C}$, with the measured value at each sampling location representing the gray-scale value at that location:

$$\Omega_{R\times C} = \{(r, c) \mid r = 0, \ldots, R-1, c = 0, \ldots, C-1\}. \tag{2.21}$$

The measurements $f(r, c) : \Omega_{R\times C} \to \mathbb{R}$ are assumed to be bounded, i.e., $\max|f(r, c)| < \infty$ $\forall r, c$, given that they represent discrete quantized values of gray-scale. Let $\mathcal{H}$ denote the class of such signals and, given a collection of $M \leq RC$ discrete sampling functions $\{\varphi_k\}_{k=1}^{M}$ (with $\varphi_k \in \mathcal{H}, \forall k$), the linear measurements of f can be generally written as:

$$y_k = \langle f, \varphi_k \rangle = \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} f(r, c)\, \varphi_k(r, c), \tag{2.22}$$

where $k = 1, 2, \ldots, M$ and where M is the amount of measurements ((2.8)

The data acquisition is done according to (2.22) as a sparse measurements model. This model can be equivalently expressed in the form of vector matrix multiplications. Specifically, let $N = RC$ and let Φ be an $N \times M$ matrix whose columns are $\{\phi_k\}_{k=1}^{M}$. Also, let $f \in \mathbb{R}^N$ be a lexicographically concatenated version of the original image (reorder the $R \times C$ 2-D image as one $N = RC$ vector). Then, the measurement model becomes

$$y = \Phi^t f \tag{2.23}$$

where $y \in \mathbb{R}^M$ denotes M length data measurement vector and where Φ^t denotes Φ transpose.

As defined in (2.1), signal f which has K-sparse representation, in a $N \times N$ matrix Ψ, can be represented by linear combination of K vectors only, chosen from Ψ.

$$f = \Psi x \tag{2.24}$$

Using matrix algebra, the sensing model is thus:

$$y = \Phi^t \Psi x + \epsilon = Ax + \epsilon \tag{2.25}$$

where y is a $M \times 1$ measurements vector and where the sparse representation and the sensing basis merged into the global $M \times N$ sensing basis $A = \Phi^t \Psi$. And where in a more practical case, ϵ, which denotes measurement additive noise. In most cases, the noise is assumed to be Gaussian [19, 18, 51, 42].

The signal f can be recovered by solving the BPDN problem [19] (2.14). In order to guarantee a reliable signal reconstruction, the required property of the sensing matrix A meets $RIP(2K, \delta)$ where RIP - Restricted Isometry Property [8, 20] is defined by :

$$(1 - \delta) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta) \|x\|_2^2 ; \ 0 < \delta < 1 \tag{2.26}$$

The Restricted Isometry Property replaces the process of verifying linear independence of all K combinations of A rows [20]. In other words, every combination of K rows of A behaves as an orthogonal transformation and is $(1 - \delta)$ far from singularity [55] and the singular values $\kappa(\cdot)$ are bounded such as $(1 - \delta) \leq C\kappa(A^t A) \leq (1 + \delta), \ C > 0$ [8, 55]. It was shown [20] that 2.8 has a unique solution if $\delta_{2K} < 1$ and 2.9 is equivalent to 2.8 if $\delta_{2K} < \sqrt{2} - 1$.

Stochastic constructions of RIP matrices exist and promote sparsity [8, 81]. Therefore, matrix construction based on random sensing can meet the RIP criterion as well as promoting sparsity. A very common RIP random construction is Gaussian white noise $\sim \mathcal{N}(0, 1)$ and a sequence of Bernoulli random variables taking values ± 1 with equal probability [20, 95]. The construction of a sampling matrix is done by selecting M basis functions $(\phi_i; \ i = 1 \dots M)$ in order to build the sensing orthonormal bases $\Phi = \{\phi_1, \phi_2, \dots \phi_M\}$ such as $\phi_{i,j} \sim \mathcal{N}(0, 1)$. The sampling matrix is used to construct the sensing matrix: $A = \Phi^t \Psi$. It was shown that for any orthonormal sparsity basis Ψ, A which its entries are also random variable which are drawn from Gaussian distributed and the RIP properties of the Gaussian distributed matrix Φ are maintained for A. In this case, it was shown [80] that the number of basis functions M is:

$$M = \Theta(\delta^{-2} K ln(N/K)), \ \delta > 0 \tag{2.27}$$

Random matrix construction is challenging to implement because of practical implementation reasons. For example, a random sensing matrix that is used for data sampling is needed for signal reconstruction. This matrix needs to be stored and in many cases transmitted via a communication line to the decoding side. For large images, the sensing matrix is significantly large. Therefore, randomizing functions have been used for creating a pseudo-random variable operating on a "seed" and constructing pseudo-random sensing matrices. Other then Gaussian distribution have been used for constructing the sensing matrices such as Noiselet transform [113] that has high mutual-coherence with the wavelet basis [23].

2.4 Sensing Strategy

The conventional CS method is illustrated in a conceptual flow chart (Fig-2.1) including the three major steps in CS process: Step #1 image sparse representation (2.1), step #2 sub-sampling (2.3) through measurement matrix Φ, and step #3 image reconstruction from a sub-set of samples (2.2). The image acquisition is performed according to the CS sparse measurement model (2.22, 2.23) using sampling basis $\{\phi_k\}_{k=1}^{M}$. The measurement matrix Φ (2.23) entries are random variable drawn from a given distribution, the most common distributions used for CS measurement matrix entries are Gaussian or Bernoulli distributions. According to the required compression rate ρ, only a certain percentage of sampling location are sampled. ρ is defined as one minus the ratio between the number of measured sampling locations and the total number of sampling locations.

The common constructed sensing matrices in the CS research community are random and uniformly distributed. In other words, the sampled locations are selected randomly, drawn from a uniform distribution. A simplistic explanation of sensing matrix construction can be viewed as setting zero columns in the sampling matrix transpose Φ^t at random and uniformly distributed locations. The non-zero entries in the matrix are random variables drawn from Gaussian or Bernoulli distributions. For illustration purposes, a sampling matrix construction is presented (Fig-2.2). In this illustration (2.2), the matrix's zero elements are marked by X in a square box and a non sampled pixel in the image vector is marked by X in a circle. If an entire column in the matrix is zero, the multiplication of sampling matrix and image, equal zero for the corresponding image location, regardless the pixel value. Therefore, this pixel can be ignored and not sampled at all.

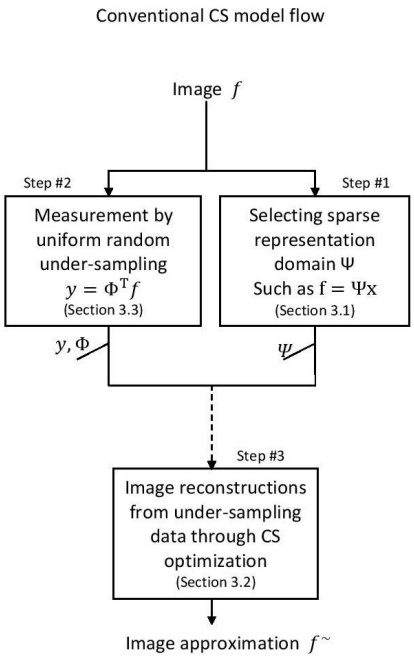


Figure 2.1: CS implementation flow chart for image reconstruction from sub-sampling

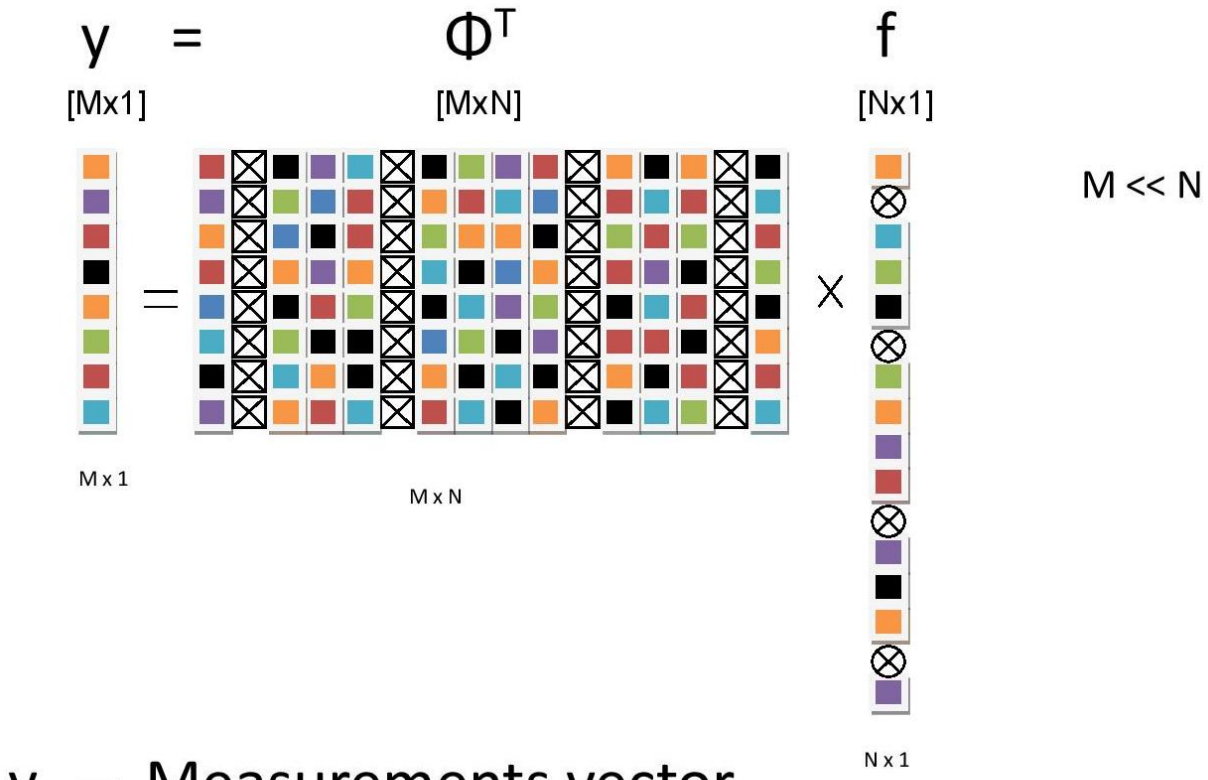


Figure 2.2: Sampling matrix illustration

2.5 Sensing Strategy - Recent Developments

In recent years there has been interest in optimizing the CS sensing probability density for improving reconstruction performance in imaging applications. Representations of MRI images have nonrandom structures since most of the image energy is concentrated close to the representation domain origin [86]. Therefore, it was proposed [150, 86, 112] to consider a variable density random under-sampling which would sample more near the origin and less in the periphery of the representation domain. The proposed sampling function was to adjust the probability density according to the power of distance from the origin [86] in the Fourier domain. Based on the same concept, it was proposed [150] to apply a similar method of variable density random under-sampling for the images domain. Image reconstruction was evaluated [150] with different sampling patterns (radial, logarithmic and random). Similarly [86], the sampling function was to adjust according to the concept of dense sampling near the origin and sparse sampling in the periphery. The coherence between the sparsity and variable density sensing bases was evaluated [112] confirming that the reconstruction performance was maintained.

In some hardware implementation scenarios, random access to each pixel in the 2D grid can be challenging [49, 118]. Structured compressed sensing was proposed [49, 118] where separable (rows and columns) matrices are appropriate. The structured sensing matrix is constructed from random patterns of rows and columns. The non-uniform sensing approaches are static, considering general representation domain properties and hardware constraints in order to improve reconstruction performance. In another study [156], there was an attempt to reduce the number of measurements by dividing the scene to blocks and sample each block by uniform distribution random sampling with different number of samples based on average block saliency level. This approach is missing important saliency information due to very low resolution through averaging block saliency as well as uniformly sampling based on an arbitrary block size reduces sampling effectiveness by missing region of interest within the block.

Sequential adaptive compressive sensing techniques have been recently proposed for optimizing the support of the sensing matrix [69]. It has been proposed [69] to construct the measurement matrix adaptively through an iterative process in order to select appropriate rows of the sensing matrix A for emphasizing the non-zero coefficients of vector x (2.13). This sequential process iteratively searches for non-zero coefficients by multiple sensing matrices starting with a full ranked matrix which measures and considers the entire coefficient vector and iteratively converges to a low support matrix. This sequential approach is very time consuming repeating the CS reconstruction process every iteration. Moreover, this method uses acquired data from the entire image in the iteration process.

In another study of this adaptive method [24], an assumption of an"infinite number of observations is available" [24] is taken for constructing a sampling matrix with minimum support for noisy signals while ignores an important goal of image reconstruction based on small subset of sampled locations. The adaptive CS method was compared mathematically to uniform distributed sampling [5] showing that there is not a significant advantage to adaptive CS universally, where the entire image information is considered equally important. Even though the analysis seems to be accurate, this hypothetical case does not represent real-world practical situations where region or interest exist in the image rather non-practical case where all regions in the image are equally important.

Several novel contributions made in this thesis have been published by the thesis author recently. A novel saliency-guided sparse measurement method [130] was developed that significantly improves CS performance for situations where regions of interest have structured characteristics (5.3.1). This model [130] optimizes the sampling probability density function (pdf) according to salient regions in the spatial domain in a binary way i.e., high sampling probability at high saliency and low sampling probability at low saliency. Based on this concept, a saliency-guided sparse reconstruction ensemble fusion method was implemented [127, 129] for improving SNR in compressive fluorescence microscopy (5.3.3). Using the previous developed model [130] the sampling pdf is adaptively optimized based on the regions of interest rather than sampling the entire area uniformly. Furthermore, ten saliency-guided sparse reconstructions are fused together into a much improved reconstruction.

An improved multi-scale saliency-guided compressive sensing approach [128] was developed as well (5.3), offering an efficient robotic laser range measurements method. This model optimizes the sampling pdf gradually and smoothly between high and low salient regions. It was shown that this approach achieves greater performance in comparison to the recently published saliency-guided sparse measurement model [130] where images are contaminated with high noise levels.

A learning process is required to adjust the sampling pdf. It was proposed [127, 128, 129, 130] to detect rough saliency in the spatial domain by an initial stage. In this initial stage, a conventional CS procedure is used where the image is sampled by sparse and uniform probability density that requires the acquisition of only a limited number of samples. An energy-guided learning approach to compressive OCT has been proposed recently [126] as well. In this method (4.1), a direct learning approach is proposed for off-line learning the data pdf in the sampled frequency domain. The learned pdf which is based on underlying data then guides the CS sampling probability. A significant improvement in CS reconstruction performance was archived using this method.

Chapter 3

Unified Sparse Measurements Model and Framework

3.1 Model for Constructing a Sampling Pattern

There is an increased interest in improving the sensing strategy of compressive sensing method (2.5). This can be achieved by optimizing the sensing probability density. In conventional CS systems, the standard choice of such sampling bases are pseudo-random sequences (e.g., Gaussian, Bernoulli, etc.) which sample the entire scene uniformly. However, in practical situations, the object of interest in the sampling domain is typically characterized by high probability salient characteristics which can be identified by the pdf. Since we are interested primarily in preserving such salient characteristics when aiming to achieve high reconstruction performance, the use of conventional pseudo-random sequences is limited in their suitability for such scenarios. A method to improve compressive sensing performance is adjusting the sensing distribution based on underlying data rather than the commonly used uniform distribution sampling for the entire scene. A learning approach to construct a sampling pattern based on the nonparametric modeling (3.4.3) can improve significantly compressive sensing performance. This method draws samples from the distribution to create a sampling pattern which optimizes compressive sensing reconstruction performance. The core of this method is learning the nonparametric model based on underlying data and optimizing the sensing probability according to data importance.

Using a measurement system which contains N sampling locations Ω_N across a scene, with the measured value at each sampling location across the entire scene define by:

$$\Omega_N = \{n \mid n = 0, \ldots, N-1\}, \tag{3.1}$$

Where $\Omega_N \in \mathbb{R}^\beta$ and where β is defined according to the sampling domain. For example, if the sampling domain is the frequency domain, Ω_N is a one dimension domain ($\beta = 1,\ \Omega_N \in \mathbb{R}^1$) while in the case of sampling in the spatial domain, Ω_N is two dimensional ($\beta = 2,\ \Omega_N \in \mathbb{R}^2$). Subsequently, Ω_N is partitioned into two complementary sets Ω_T and Ω_T^c such that

$$\Omega_N = \Omega_T \cup \Omega_T^c, \text{ with } \Omega_T \cap \Omega_T^c = \emptyset, \tag{3.2}$$

whose cardinalities are equal to $\#\Omega_T = T$ which means that number of sampling location across the scene is T and $\#\Omega_T^c = N - T$. The subset Ω_T denotes sampling location across the entire scene and Ω_T^c denote unsampled locations. Subset Ω_T is generated based on a function $\Gamma(n)$ that provides a quantitative measure of sampling probability at sampling location n

$$\Gamma(n) \rightarrow [0, 1] \ \forall\ n \tag{3.3}$$

where $\Gamma(n)$ returns a sampling probability level value between 0 and 1 for every location n:

$$\Omega_N \in \mathbb{R}^\beta \rightarrow \Gamma(n),\ \forall \beta. \tag{3.4}$$

Consider that the signal of interest at the sampling domain $f : \Omega_N \rightarrow \mathbb{R}$ is assumed to be bounded, i.e., $\max_k |f(k)| < \infty$, given that it represents discrete quantized signal values. Let $\{\varphi_k\}_{k=1}^M$ denote a collection of $M \leq N$ discrete sampling functions. The linear measurements of f can be generally written as:

$$y_k = \varphi_k^t\ f + \epsilon_k \tag{3.5}$$

Where $f \in \mathbb{R}^{N\times 1}$ is a concatenated version of the original signal and where φ_k^t is the sampling function φ_k transpose and ϵ_k denotes measurement noise.

This model can be described in a matrix form (2.3). Let Φ be an $N \times M$ matrix whose columns are $\{\phi_k\}_{k=1}^M$. Then, the measurement model becomes:

$$y = \Phi^t f + \epsilon \tag{3.6}$$

Where $y \in \mathbb{R}^M$ denotes M length data measurement vector and where Φ^t denotes Φ transpose.

The sampling bases $\{\varphi_k\}_{k=1}^M$ are chosen to be random with optimized pdf Γ (3.3) based on underlying data rather than the uniform or static distributions (2.5) which are used for conventional CS. Since this optimized pdf Γ is unknown, it should be constructed by learning and optimization methods (3.4.1, 3.4.2, 3.4.3). In general case, this pdf (Γ)will be approximated through a nonparametric method since no known parametric structure can be identified that represents all cases. This model represents under sampling where

$M < N$, the sampling basis definition is modified to account for the lack of observations at locations in Ω_T^c:

$$\varphi_{T,k}(n) = \begin{cases} \varphi_k(n), & \text{if } n \in \Omega_T, \\ 0, & \text{if } n \in \Omega_T^c. \end{cases} \tag{3.7}$$

Consequently, the sparse measurements model can be expressed as:

$$y_k = \langle f, \varphi_{T,k} \rangle + \epsilon_k, \quad k = 1, 2, \ldots, M \tag{3.8}$$

where $\varphi_{T,k}$ values are realizations of a random variable z whose probability density is defined in (3.9):

$$p_T(z \mid \Gamma(n), \pi) = (1 - \pi\Gamma(n))\, \delta(z) + \pi\Gamma(n)\, p_S(z)) \tag{3.9}$$

The sampling basis entries are random variables drawn from distribution $p_S(z)$ where the most common probability functions are Bernoulli or normal ($\mathcal{N}(z|0, 1)$. Sampling location is a random variable drawn from a probability which is defined with a weight at location n that is guided by underlying data. Underlying data is represented by its pdf: $\Gamma(n) \rightarrow [0, 1]$. Here the pdf of $\delta(z)$ has zero mean with zero variance and where π is the required under sample level $\pi \rightarrow [0, 1]$. Thus, each entry to the optimally constructed sampling matrix Φ, is random variable drawn from probability p_T (3.9) is zero with probability $(1 - \pi\Gamma(n))$ and is p_S distributed with probability $\pi\Gamma(n)$.

3.2 Conventional CS - General Concept

The conventional CS method is illustrated in a conceptual flow chart (Fig-2.1) in section 2.4. The conventional CS sampling basis used in existing literature is defined as:

$$\varphi_{T,k}^{CS}(n) = \begin{cases} \varphi_k(n), & \text{if } (n) \in \Omega_{CS} \\ 0, & \text{if } (n) \in \Omega_{CS}^c \end{cases} \tag{3.10}$$

The measuring matrix Φ is constructed according to distribution p_T (3.9) where each entry to the sampling matrix is a random variable drawn from $\varphi_k(n)$ (3.10) for each location n. The most common probability functions used for φ_k are Gaussian distribution with zero mean and unit variance or Bernoulli while 0 and 1 are equally distributed.

In the conventional CS, the non-sampled locations are selected randomly drawn from a uniform distribution. Therefore, for the conventional CS, $\Gamma(n)$ (3.9) represents a uniform distribution for sampling the entire scene uniformly. Ω_{CS} (3.10) represents sparse sampling locations and Ω_{CS}^c (3.10) represents locations that are not sampled. The cardinalities of data-guided statistical sparse CS measurements (3.2) and conventional CS are equal therefore $\#(\Omega_T) = \#(\Omega_{CS})$ and $\#(\Omega_T^c) = \#(\Omega_{CS}^c)$, respectively.

3.3 Data-guided CS Method - General Concept

For illustration purpose, conceptual flow charts (Fig-3.1 and Fig-3.2) describes the proposed data-guided statistical sparse method.

The proposed data-guided statistical sparse method optimizes the sampling pattern according to underlying data while the conventional CS method sample the entire scene uniformly. In order for optimizing CS reconstruction performance, the proposed model (3.1) is constructing an optimized sampling matrix Φ (3.6) which is determined through learning. Sampling locations are random variables drawn from a learned pdf $\Gamma(n)$ (3.3) which provides a quantitative measure of sampling probability at sampling location n.

The proposed data-guided statistical sparse method contains density function learning process through two methods - direct and indirect. The direct method is used for learning a nonparametric pdf directly from the dataset where no mediator operation is required. The indirect learning method is for cases where a mapping between extracted features and the pdf is required. In both direct and indirect learning methods, a nonparametric pdf is approximated. The optimized sampling matrix Φ is constructed by sampling patterns which are created by the approximated nonparametric pdf.

The flow chart (Fig-3.1) refers to the proposed data-guided statistical sparse method steps (3.4.1 and 3.4.2). The flow chart indicates the proposed method innovation as well as the common steps that are similar to the conventional CS approach. The innovation focuses on constructing an optimized sampling pattern based on underlying data through learning. After the optimized sampling matrix Φ is constructed, the proposed as well as the conventional CS methods use the same steps, such as sparse representation and image reconstruction. All steps which are common to the conventional CS and the data-guided CS method (such as reconstruction algorithms) use identical implementation and parameters in order to compare the two methods reliably.

3.4 Learning Data Distribution to Guide Sampling

Learning a data distribution is a key process for optimizing sampling patterns and constructing the sampling matrix. In this section, learning methods for constructing an optimized sampling pattern based on nonparametric modeling are discussed. In sections 3.4.1 and 3.4.2 two learning options are discussed - direct and indirect.

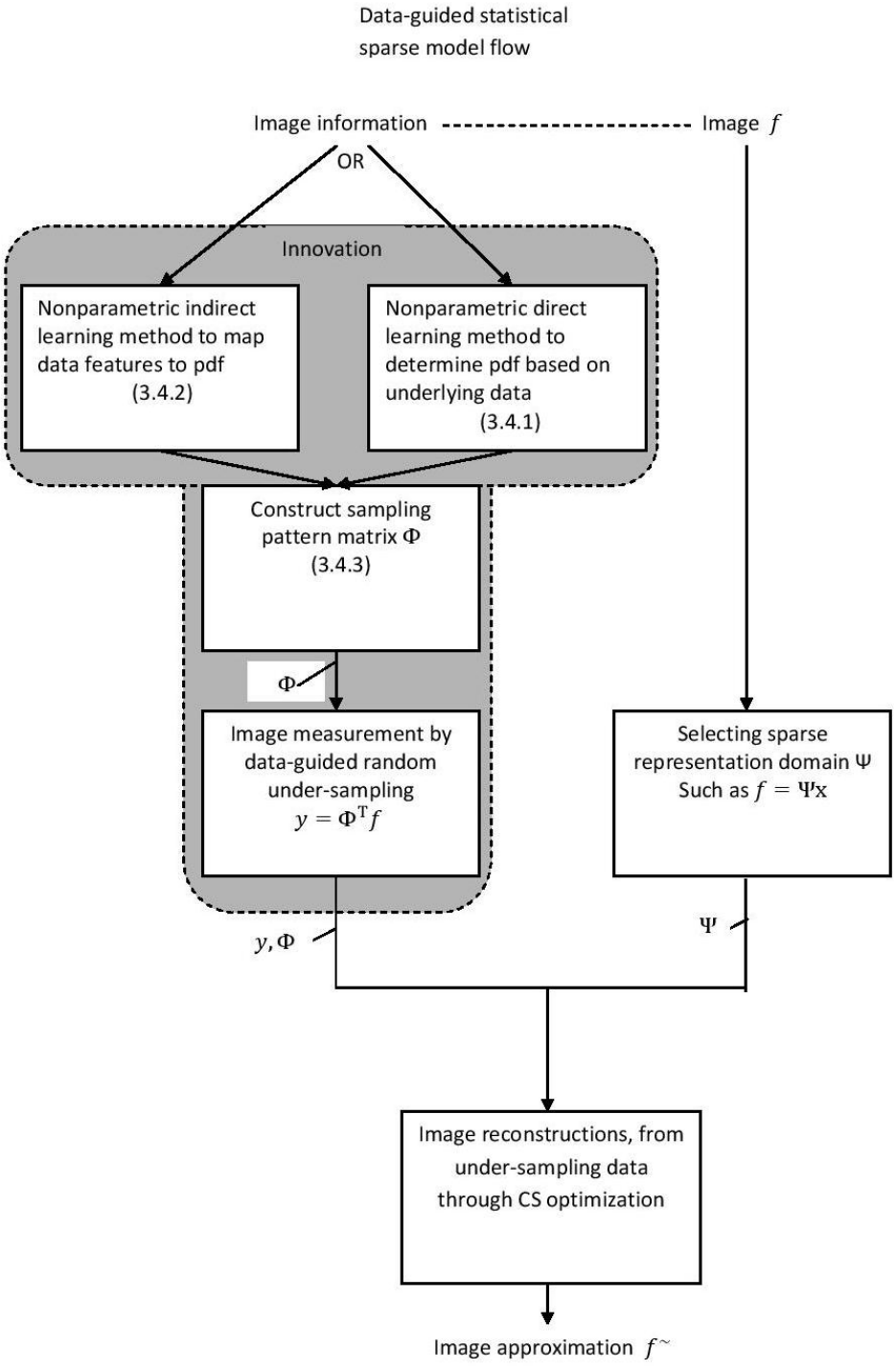


Figure 3.1: Data-guided CS implementation flow chart for image reconstruction from sub-sampling.

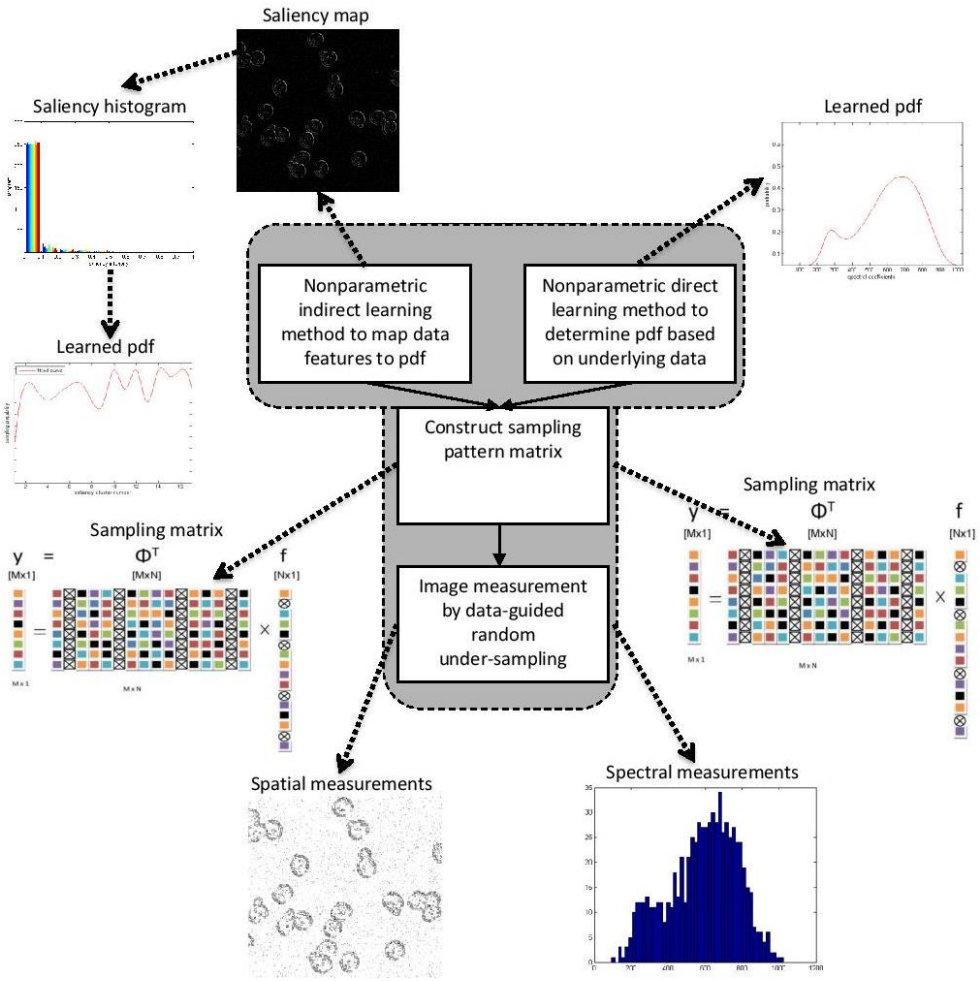


Figure 3.2: Data-guided CS direct and indirect implementation flow chart. The direct method is used for learning a nonparametric pdf directly from the dataset, for example, spectral coefficients of OCT data. The indirect learning method is for cases where a mapping between extracted features and the pdf is required. for example, saliency map of fluorescence microscopy data is calculated. Based on the histogram of the saliency information, clusters are formed and pdf is determined. Based on the learning pdf, a sampling matrix is contracted which creates optimized under-sampling patterns for improving CS performance

3.4.1 Learning Nonparametric pdf Based on Underlying Data Directly

In this work, CS sensing distribution is adjusted based on underlying data rather than the uniform distribution which is commonly used. Therefore, there is a great interest in pdf estimation which is guiding sampling pattern construction $\varphi_k^T(n)$ (3.7) based on underlying data. In this section, methods for learning pdf Γ (3.3) are described for cases where Γ can be learned based on the dataset directly where no mediator operation, such as feature extraction for example, is required. For example, if signal energy is considered as high importance, and sampling is given in the frequency domain, then a direct learning approach of the frequency domain pdf would be appropriate and constructing the sampling patterns accordingly is efficient.

A pdf is a fundamental concept in the field of statistics. Let k be a random variable that has pdf Γ which provides the probability of variable k in the range $[a, b]$:

$$P(a < k < b) = \int_a^b \Gamma(k)dk; \ \forall \ a < b \tag{3.11}$$

Assume that a set of measurements is given which is sampled from an unknown pdf. This pdf can be reconstructed through density estimation analysis [134]. One option is to perform parametric density estimation. In parametric density estimation analysis, the samples are drawn from a known distribution class that contains a set of parameters. One common example is the normal distribution where the mean and variance are the parameters. The measurement's parametric density function can be determined by approximating the parameters only and set their approximated values in the known pdf. Since we are interested in a generic model for constructing a sampling pattern (3.1), it would be preferred to consider a more general concept, where no formal parametric structure is assigned. In this practical situation one cannot assume that the measurements belong to any known pdf, therefore a nonparametric approach is required. In this general case, the pdf belongs to a large family of densities which cannot be represented by a finite number of parameters [75]. The field of nonparametric density estimation is well researched and broadly used in statistical analysis [75]. In the last six decades many nonparametric density estimation methods were proposed and efficiently implemented for univariate and multivariate data. One of the oldest density estimation methods, and still very common, is the histogram. In the histogram density estimation method, the origin and bin width have to be selected, where the bin width effects and controls the function smoothness level.

Given the signal X, the multi dimensional pdf approximation is given by:

$$\widetilde{\Gamma} = \frac{\sum_{k=1}^{K} N^k \cdot \delta(X_k - U_k)}{N \cdot h} \tag{3.12}$$

where $\delta(X_k - U_k)$ describes bin k, N^k represents the number of times that the level X_k appears at bin k, U_k is the location of the k_{th} bin location, $N = \sum_{k=1}^{K} N^k$ and h is the bin width.

One of histogram disadvantages is the need to define the bin width. In some applications, such as this research, the bin width is set to one pixel for the image analysis. The naive density estimator [11] is a summation over all measurements of a rectangular of width $2h$ (where h is a small number) and height $\frac{1}{2hn}$ on each measurement. Similarly to the histogram pdf estimation method, the bin width still applies (parameter h), however, in the naive estimation, each point is considered to be at the center of the sampling interval. A generalized presentation of the naive density estimator is the kernel estimator [121, 151] which replaces the weight function with a kernel function The result of kernel estimation is smoother than the previous methods since a smooth function replaces the rectangular one in the kernel estimation case. The main disadvantage of the kernel estimation method is noise that appears at the long tail of the distribution function. The nearest neighbor density estimator method [85] improves the performance compared to kernel estimation regarding noisy long tail distribution. The variable kernel method [15] also controls the smoothness to local data density but unlike the kernel estimator the scale parameter changes from one data point to another.

Different approaches and methods to nonparametric density estimation were proposed. The orthogonal series estimators [25] estimate a pdf by applying a low pass filter in Fourier domain. This method has been used for multivariate densities and has been implemented in applications in different areas such as pattern recognition and classification [75]. Statistical approaches such as maximum penalized likelihood estimators [63] have been developed and used broadly. This method penalizes the likelihood function for producing "not accurate" estimations through optimization of two elements, data fidelity and smoothness. The optimization maximizes the logarithm of the likelihood function underlying a set of independent distributed observations subjected to regularized roughness [134].

3.4.2 Learning Approach for Optimized Mapping to Actual Probability

In the previous section (3.4.1), a direct learning approach for modeling based on underlying data through learning a nonparametric pdf is described. This direct approach is applicable

in cases where the region of interest appears from the data directly and no mediator operation, such as feature extraction, is required. However, in many cases, the object of interest within the scene is characterized by structured characteristics with distinct features. In those cases, an indirect approach is more appropriate since an initial operation, such as feature extraction is performed, in order to identify structured characteristics. Therefore, a mapping between extracted features and an approximated pdf Γ (3.3) is required. This pdf is used for construction of sampling pattern $\varphi_k(n)$ (3.7) based on underlying data indirectly.

The indirect learning approach is based on learning an optimized mapping between sampling probability and derived features. Since we are primarily interested in preserving such important characteristics when aiming to achieve high reconstruction performance, one can learn the model indirectly via extracting derived features and then learn a non-parametric model to optimize reconstructed image quality. For example, if patterns and structures in some domain are considered to be of high importance, then a feature extraction operation for identifying those regions of interest is required as an initial process. After identifying regions of interest through features, a mapping between those features and an optimized pdf is realized through learning. This optimized pdf is used for constructing CS sampling patterns for maximizing reconstruction performance.

Assume the image has structured characteristics with distinct features and that those features S were extracted and are represented in a multi dimensional space, $S \in \mathbb{R}^{\beta}$, where β represents the feature space dimension. In order to make the feature to pdf mapping more practical, group the features S in clusters with common importance level. Since this model is generic and can be implemented in a wide range of applications, the definition of importance level will be abstractive and will need to be selected appropriately for a specific practical implementation as will be described in chapter 5. Group the features in R clusters denoted by θ_i, $i = \{1, \ldots, R\}$ that have common properties. One efficient way to perform this clustering process is by learning the features pdf through a multi-dimensional histogram [1] and cluster features with common probability [1, 72].

G is defined as a mapping operator (3.13):

$$G : S \in \mathbb{R}^{\beta} \rightarrow \Gamma \in \mathbb{R}^{\beta} \tag{3.13}$$

where β is the dimension of S and Γ (3.4). Operator G maps features from S to an optimized pdf Γ. Since G operates on the clusters, $G(\theta_i^{\mu})$ is a function of cluster θ_i, $i = \{1, \ldots, R\}$ and where the clusters are sorted by the cluster mean θ_i^{μ}, which is the average importance of region i.

In order to complete the process, one needs to define the mapping function G (3.13) which will provide the desired pdf Γ leading to improved CS reconstruction performance.

One can select a parametric or nonparametric approach as was discussed above (3.4.1). Through the evolution of developing this model, an initial and simpler attempt was made by selecting a parametric approach. Intuitively, a good mapping function would emphasize high important features more then lower important ones therefore one can select the mapping function to be a parametric logarithmic-based function.

An optimized approach which is not depend on an arbitrary parametric selection, is a nonparametric approach that will find an optimized mapping function G (3.13) which is generic having no assumptions on parametric function of signal of interest. In this case, constructing a sampling pattern based on this nonparametric modeling approach will maximize CS reconstruction performance through learning an optimized pdf. This optimization approach is evaluated in Section 5.3.4. Monte Carlo can be used as an efficient optimization method [117] for learning the pdf Γ by an optimized mapping function G.

Since our optimization problem is multidimensional problem $\overline{z}$ with R variables, as the number of clusters, a multivariate method will be appropriate. While the problem dimensionality is unknown since the number of clusters is affected by the specific image as well as noise level. In many optimization problems, there is a risk of converging to a local extremum rather than to the desired global one. This situation is even more critical where the optimization method deals with multidimensional problems. One good candidate for addressing this optimization issue is the Markov chain Monte Carlo (MCMC) method. The MCMC [3, 67, 59] method optimizes the function $G(\overline{z})$ using a Markov chain. Markov chain principle is that the probability of a random variable in a sequence $\overline{z}_1, \overline{z}_2, \ldots, \overline{z}_n$ depends on the previous variable only. In other words, there is no "memory" beyond more than one variable back in the sequence when generating the next variable: $P(\overline{z}_k|\overline{z}_{k-1}, \overline{z}_{k-2}, \overline{z}_1) = P(\overline{z}_k|\overline{z}_{k-1}),\ \forall k$ [31] which is called transition probability distribution at step k.

The mapping function G will optimize the pdf Γ which will maximize reconstruction performance:

$$\arg\min_{\overline{z}} \|\widetilde{f} - f\|_2 \ s.t.\ \{G(\overline{z}) : S \rightarrow \Gamma\} \tag{3.14}$$

where $\overline{z}$ is a Markov chain $\overline{z}_1, \overline{z}_2, \ldots, \overline{z}_n$ where each sequence $\overline{z}_i \in \mathbb{R}^R$, function G (3.13) is mapping features (saliency in our case) from S to an optimized probability function Γ to maximize CS reconstruction where f is the original image and $\widetilde{f}$ is the approximation image.

A common method for performing this optimization is Metropolis-Hastings method [31, 125]. In Metropolis-Hastings, one draws a sequence of samples $\overline{z}^*$ to assign to sequence $\overline{z}_k$ by a conditional distribution through an accept-reject process. Sequence $\overline{z}^*$ will be

accepted to be $\overline{z}_k$ with probability $\alpha(\overline{z}^*, \overline{z}_{k-1})$:

$$\alpha(\overline{z}^*, \overline{z}_{k-1}) = min\{1, \Delta(f(\overline{z}^*), f(\overline{z}_{k-1}))\} \tag{3.15}$$

where function Δ is selected according to the application to optimize the process (such as exponential function for example (see 5.14 in implementation 5.3.4)). The method compares uniform distributed random number u with the probability criterion (3.15). If $u < \alpha(\overline{z}^*, \overline{z}_{k-1})$ then we accept and assign $\overline{z}_k = \overline{z}^*$ otherwise it is rejected and unlike acceptance-rejection method where the rejected variable is discard, in our case we assign the previous variable in the sequence $\overline{z}_k = \overline{z}_{k-1}$. A random walk chain has been commonly implemented for creating the Markov chain [31] where $\overline{z}^* = \overline{z}_{k-1} + \overline{w}$ where $\overline{w}$ could come from normal, uniform or other distribution.

3.4.3 Method to Construct the Sampling Pattern

The model for constructing a sampling pattern (3.1) starts with learning data distribution Γ (3.3) to guide sampling (3.4), followed by constructing the sampling matrix according to the learned nonparametric pdf, in order to improve CS reconstruction performance. Unlike the conventional CS which sample the entire scene uniformly, the optimized sampling matrix is constructed according to underlying data through this learned pdf (Γ (3.3)) which guides sampling sparseness. For example, if the pdf Γ is higher, the sampling is denser. The model (3.1) provides the framework for constructing the sampling basis $\varphi_{T,k}$ (3.7). The basis values are realizations of a random variable z whose distribution p_T is defined in (3.9). For illustration purposes and to understand better the main difference between sampling matrix constructing in the conventional CS method versus the proposed data-guided statistical sparse method, the simplistic explanation for the conventional CS (2.2), will be reviewed in Section 2.4. In the conventional CS method, the zero columns in the sampling matrix transpose Φ^t are selected at random locations drawn from a uniform distribution. In the proposed data-guided statistical sparse method, the zero columns of Φ^t are selected randomly with probability p_T (3.9) which is guided by the learned pdf (Γ (3.3)). The non-zero entries in the matrix are random variables drawn from normal distribution, identical to the conventional CS method.

We can use the illustration (2.2) from the conventional CS discussion CS (2.4) to illustrate the sampling matrix reconstruction for the data-guided statistical sparse method as well. In this illustration (2.2), the non-sampled location (non sampled pixel in the image vector is marked by X in a circle) belongs to Ω_T^c set. In the data-guided statistical sparse method, those zero column locations are been selected randomly with distribution p_T (3.9) which is guided by the learned pdf (Γ).

Inverse transform sampling method is used to construct an optimized sampling pattern from the direct learning probability (3.4.1). Through inverse transform sampling one can draw samples from the desired pdf using a uniform distribution. The inverse transform sampling method is defined [28, 117]: Let $C_\Gamma(w)$ be the cumulative distribution function (cdf) of random variable w over the learned pdf Γ:

$$C_\Gamma(w) = P(\Gamma < w) \tag{3.16}$$

In other words, the cdf can be calculated based on the learned pdf Γ. Let C_Γ^{-1} be the inverse cdf defined [28, 117]:

$$C_\Gamma^{-1}(v) = \inf\{v \mid C_\Gamma(w) \geq v; \;\; \forall \;\; 0 < v < 1\} \tag{3.17}$$

An inverse operator C_Γ^{-1} is used to draw samples for the direct approach as following: if random variable v comes from uniform $(0, 1)$ distribution, than random variable $w = C_\Gamma^{-1}(v)$ comes from the distribution of w. In other words, if the inverse C_Γ is available through learning, one can draw samples from it using uniform distribution in order to create the sampling pattern according to the learned pdf Γ.

For creating the sampling pattern in the indirect approach case, pdf Γ will guide the sampling pattern (3.7) sparsity probability where $\varphi_{T,k}$ values are realizations of a random variable drawn from distribution p_T (3.9). The learned (3.4.2) mapping operator G (3.13) maps each sample location within cluster θ_i, $i = \{1, \ldots, R\}$ to Γ for constructing the sampling patterns. The construction of an optimized sampling pattern from an indirect learning probability (3.4.2), is performed by using Acceptance-Rejection sampling method.

Acceptance-Rejection sampling [28, 31, 117] is another popular sampling method to draw samples from desired distribution. Similarly to the inverse transform sampling case, an easy to access and available distribution is used to draw samples from the targeted distribution. The function $G(\theta_i)$ maps cluster sampling locations to their appropriate sparsity probability level in the range $(0, 1)$ according to pdf Γ. Acceptance-Rejection method is used for this allocation as follows (3.18): draw a sample n_i from this cluster θ_i. The sample will be accepted with probability which is mapped for this cluster $G(\theta_i^\mu)$ where θ_i^μ is the i^{th} cluster mean (3.4.2).

The implementation algorithm uses random number u with uniform distribution in the range $(0, 1)$, in order to implement the Acceptance-Rejection method. This random number is compared with the probability criterion $\Gamma(n_i)$. In other words, n_i is accepted if $u < \Gamma(n_i)$. The higher $\Gamma(n_i)$ is (for cluster i), the higher the acceptance probability is. If n_i is accepted, this location will be sampled, therefore $n_i \in \Omega_T$ (3.7). If n_i is rejected this location will not be sampled and $n_i \in \Omega_T^c$ (3.7). The excepted location value

in the sampling matrix will be a random number, with normal distribution according to probability p_T (3.9), and the rejected location value will be zero. In this case, the sampling pattern (3.7) construction is modified according to the learned mapping function (3.18):

$$\varphi_{T,k}(n_i) = \begin{cases} \mathcal{N}(z|0,1), & \text{if } u < \Gamma(n_i); \;\; n_i \in \theta_i \; \forall i, \; i = \{1, \ldots, R\} \\ 0, & \text{if } u \geq \Gamma(n_i); \;\; n_i \in \theta_i \; \forall i, \; i = \{1, \ldots, R\}. \end{cases} \tag{3.18}$$

Where a random variable is been drawn from normal distribution $\mathcal{N}(z|0,1)$ with 0 mean and variance 1, $\Gamma(n_i)$ is the acceptance rejection criterion for location n_i that is belong to cluster θ_i.

Chapter 4

Nonparametric Direct Learning Approach to Sampling Pattern Generation

The proposed sensing model (3.1) allows for the creation of an improved CS sampling pattern according to a learned pdf. The model is generic, providing the foundation for improved CS sampling of images in multiple applications and wide range of sparse representation domains. The optimized CS sampling pattern is based on underlying data and require learning. A direct learning approach to construct a sampling pattern based on the nonparametric modeling is proposed. The nonparametric direct learning method optimizes the sampling probability according to data importance in the domain of interest. One important example as a domain of interest in many systems and applications (such as spectral-domain optical coherence tomography (SD-OCT), Magnetic Resonance Imaging (MRI), vibrational sensors) is the frequency domain. In many cases, high energy in the frequency domain represents data importance. In this chapter, a frequency domain energy guided CS (EGCS) model is presented based on the energy distribution of the measured signal in the frequency domain, with applications to SD-OCT and bridge vibration data collection.

4.1 Frequency Domain Energy-guided CS Model

A frequency domain EGCS model is proposed which learns the energy spectral density to improve CS reconstruction performance. In applications such as SD-OCT and MRI, the

measurements are made in what is known as k-space [86], which is a formalism referring to the Fourier domain. Consider the spatial/time domain signal d being measured in k-space/frequency domain D. The relationship between d and D is formulated as

$$d = \mathbb{F}^{-1}\{D\} \tag{4.1}$$

where $\mathbb{F}^{-1}$ is the inverse Fourier operator. To reduce the number of samples, one wishes to reconstruct the signal d, from only a small subset of measured signal in the k-space domain. Let an undersampled reconstruction d_u be expressed as

$$d_u = \mathbb{F}^{-1}\left\{\Phi^t D\right\} \tag{4.2}$$

where D is a vector and where Φ^t is a measurement operator defining which spectral locations would be considered. The goal is to reconstruct d from a sparse sampling of D. According to the theory of compressive sensing (2) one can reconstruct d by maximizing the sparsity of the signal in the transformed domain and enforcing data fidelity in the k-space domain (2).

The sampling method can have a significant impact on data acquisition efficiency in terms of the number of required samples in the frequency domain while maintaining high reconstruction quality in the spatial/time domain. In the conventional CS approach, the entire scene is sampled randomly and uniformly distributed, where all sampling locations are considered with equal probability [8, 20].

A static non-uniform sampling approach was implemented in the frequency domain for MRI measurements [86]. In this case, a static function (one over power distance from the origin) is used, not considering the underlying data. However, in practical situations, the energy of coefficients in the frequency domain are usually concentrated in some frequency bands and sparse in others. In other words, the energy spectral density (ESD) has structural characteristics. The conventional sampling approach is limited in the capability of preserving high ESD.

The proposed frequency domain energy guided CS model addresses this important aspect, preserving high energy spectral density to improve reconstruction performance. Consider the scene being measured using a measurement system to contain N sampling frequency locations (3.1). Subset Ω_T is generated based on a probability function $\Gamma(n)$ (3.3) that provides a quantitative measure of ESD at frequency sampling location n. The linear measurements of D can be generally written as:

$$y_k = \langle D, \varphi_k \rangle + \epsilon_k \tag{4.3}$$

where $k = 1, 2, \ldots, M$ and ϵ_k denotes measurement noise vector. In a more compact matrix form:

$$y = \Phi_T^t \, D + \epsilon \tag{4.4}$$

where y and ϵ are $M \times 1$ vectors, D is $N \times 1$ vector and Φ_T^t is $M \times N$ sampling matrix with T non-zero elements. Since $M < N$ in the EGCS learning model, the sampling basis is modified to account for the lack of observations at frequency locations in Ω_T^c (3.7) Consequently, the frequency domain energy guided CS model can be expressed as

$$y_k = \langle D, \varphi_{T,k} \rangle + \epsilon_k, \quad k = 1, 2, \ldots, M \tag{4.5}$$

where $\varphi_{T,k}$ value is a realizations of a random variable whose probability density is defined in (3.9) and where each sampling basis entry is random variable drawn from distribution p_S (3.9) where its probability here is defined as Bernoulli with equal probability 0 and 1. Sampling location is random variable drawn from probability p_S (3.9) which is defined with weights at sampling location n that is guided by the learned spectral measured coefficients represented by its pdf: $\Gamma(n) \rightarrow [0, 1]$ (3.9).

In this model, the measurement sparsity is ESD-guided for each n frequency location of $D(n)$ where $\Gamma(n)$ is in the range of $[0, 1]$ and is proportional to the energy spectral density.

4.2 Direct Learning Model in the Frequency Domain

The energy spectral density is quantified based on the signal measurements in the frequency domain. Given that high energy measurements within the frequency domain scene can have structural characteristics, it would be useful to quantify such characteristics using the measurements pdf. Therefore, in the proposed implementation, spectral measurements pdf are learned in step one (3.4.1). Let Γ be the measurements pdf (4.6) that will be used to implement the sparse sampling probability p_T (3.9):

$$\Gamma(n) = \frac{\sum_i |D_i(n)|}{\sum_n \sum_i |D_i(n)|} \tag{4.6}$$

where D_i are k-space measurements at spatial location i (4.1) which are the Fourier transform of the spatial signal d_i. In the EGCS learning approach, the measurement sparsity is guided by the measurement pdf. Since Γ is an univariate distribution, inverse transform sampling method can be easily implemented [28, 117] (3.17).

4.3 Realization of Energy-guided Modeling Approach in Frequency Domain

The implementation of the proposed EGCS modeling approach in frequency domain includes two steps: learning step and EGCS step. For this practical realization, based on the learned pdf, the measurement sparsity is ESD-guided for each n frequency location of $D(n)$. The first learning step is concluded once the sampling probability (3.9) as well as subset Ω_T are determined. In the second step, the EGCS step, D is sampled by Φ^t (4.4) with probability p_T (3.9) in order to measure higher energy coefficients at higher accuracy. The samples from subset Ω_T are used for creating the sampling basis $\varphi_{T,n}$ that create sampling matrix Φ^t (4.4). The acquired samples are used to reconstruct the signal at a higher quality.

4.3.1 SD-OCT Application

Given the aforementioned EGCS model, one can illustrate the effectiveness of such a model by applying it to the application of SD-OCT. Optical coherence tomography (OCT) is a non-invasive optical imaging modality that can provide cross-sectional and volumetric images of biological tissue with cellular level resolution and at depths of up to 2mm in biological tissues [47, 73]. As such, OCT is well suited for non-invasive imaging of ocular tissue (retina and cornea) and over the past 15 years has emerged as one of the dominant ophthalmic diagnostic modalities [58, 62, 149].

Imaging large volumes of retinal or corneal tissue, while keeping the image quality high requires high density sampling of the imaged volume, which results in very large data sets and increases significantly the image acquisition time. As a result, natural eye motion that occur with frequency of 1Hz [93] such as fixational micro-saccade, can introduce motion artifacts in the imaged data, which can render partial or whole data sets unusable, and indirectly increase the patient examination time significantly.

Different hardware or software approaches to dealing with eye motion have been proposed in the past. Low density scanning protocols are typical for commercial OCT system however, in this approach the shorter imaging time comes at the expense of reduced image quality. Recently, a compressive sparse sampling algorithm was proposed for optimal OCT scanning in XY direction, that preserves high OCT image quality [155]. Use of FDML lasers has increased the OCT scanning rate from kilohertz to megahertz [78], however, currently OCT systems based on FDML technology are very complex, expensive and require

high speed electronics. Combining eye tracking with OCT can correct most of the motion artifacts, however, this solution requires very complex hardware redesign and software algorithms.

Furthermore, some clinical applications of OCT, such as, whole eye imaging, require a very large scanning range, while maintaining the image quality (spatial resolution and SNR). In SD-OCT, the scanning range and the depth variation of the system SNR are dependent on the sampling density of the interferometric signal [32, 82].

In SD-OCT, the number of sampling points is determined by the number of illuminated pixels in the camera interfaced to the spectrometer [82]. In swept source OCT (SS-OCT), the number of sampling points is determined by the sweep rate of the tunable laser in relation to the digitizer rate [32]. In both cases, the cost of the detection technology, CCD and CMOS cameras in the case of SD-OCT, and digitizers in the case of SS-OCT, increases monotonically and significantly with increasing the number of camera pixels or digitizing rate. Furthermore, the increased number of sampling points inevitably leads to longer data and image processing time.

Recently, the use of compressive sensing for reconstruction in high resolution OCT imagery has been reported by a number of different researchers such as Mohan et al. [97], and Young et al. [154]. Those publications show promising efficient reconstruction of OCT images from highly under-sampled k-space data. However, these studies utilized either simulated OCT signals or actual images of onions, which have very different and much less complex morphology as compared to living biological tissues such as the human retina, cornea, skin, etc. In a recent work [83], a non-local strategy for sparse OCT reconstruction algorithm was evaluated on living tissues, which showed better results compared with the L_1 minimization approach. However, the experimental results showed that for tissues with complex morphology, such as human retina, cornea and skin, a high percentage of the originally sampled data is required to generate a reconstructed image of sufficiently good quality, which is still a very large data volume to process in clinical applications where real time visualization of volumetric images is becoming a requirement.

While the main focus of the CS research community is image reconstruction, an optimized and adaptive sampling procedure has been less studied for practical applications such as compressive as SD-OCT. The design of a data adaptive sampling procedure can have a significant impact on CS performance for practical applications such SD-OCT, where the objects of interest have structured characteristics in the frequency domain, thus making the sampling procedure worth investigating. Existing CS based systems employ a sampling scheme that samples the entire scene in the same manner regardless of the underlying data. However, such an approach is limiting for many practical applications, which involve dis-

tinct regions of interest in some basis, since it does not consider data importance. In many cases such region of interest are of greater interest for analysis purposes, one is motivated to obtain higher quality reconstructions for these regions than the background regions.

In a recent study, a saliency-guided sparse measurement model [127, 128] has been proposed for a significant CS reconstruction improvement. This method optimizes the sampling pdf according to salient regions in the spatial domain, where high saliency is sampled with high probability and low saliency is sampled at lower probability. It has been shown that this approach achieves greater reconstruction performance in comparison to the common uniform sampling distribution or matching reconstruction performance with significantly fewer samples.

While signal acquisition in most current SD-OCT systems is based on static pixel array, studies [6, 84, 136, 139] have explored in detail the development of CCD cameras with randomly addressable pixel sensors that are being designed for high speed imaging. As with all existing literature in compressive SD-OCT, the novel methodology being developed here acts as the theoretical foundation for advanced research in high speed SD-OCT systems that are optimized for the integration of such cameras once they become available for SD-OCT systems, with the aim to acquire significantly fewer samples while maintaining image quality.

In this section, a direct learning approach to construct a sampling pattern based on nonparametric modeling is presented. This approach, improves compressive SD-OCT reconstruction performance by dynamically adapting the sampling model based on the underlying data in the frequency domain. Consider the spatial domain signal d being measured in k-space domain as one dimension D (4.1), where d represents one OCT reflectivity profile (A-scan) vector, which contains spatial information of the OCT depth dimension. An OCT 2D slice (B-scan), is represented by image $f(r, c)$ and is created by combining all depth spatial data d in this slice. The 3D volume OCT data is created by the collection of all 2D B-scan slices. The motivation for CS in general and for EGCS more specifically, is measuring the data by much fewer A-scan spectral coefficients. In CS, the sub-sampling is selected randomly and uniformly distributed while in EGCS, the spectral coefficients are selected based randomly according to learned pdf. The proposed learning method can be based on a reference B-scan from another tissue which has similar properties (such as retina from another sample for example). Another learning option can be based on the OCT light source energy spectral distribution which is selected according to the tested tissue therefore can be used as previous knowledge.

For illustration purposes, the learned distribution of different types of tissues are demonstrated in Fig. 4.1. Retinal and corneal pdf are presented in Fig. 4.1(a) and Fig. 4.1(b).

The pdf of a different type of tissue, fingertip, is examined as well in Fig. 4.1(c). In the case of the sampling pdf learned from retina data acquisitions, the energy is concentrated at two ranges, highest energy range at spectral coefficients 600-800 and secondary energy peak at spectral coefficients 250-350. For the sampling pdf learned from cornea data acquisitions, the energy is concentrated at two ranges, with highest energy range at spectral coefficients 550-650 and secondary energy peak at spectral coefficients 250-300. The fingertip pdf is more uniformly distributed in comparison to retinal and corneal pdfs. The energy is concentrated at two ranges, with the highest energy at spectral coefficients 350-750 and secondary energy peak at spectral coefficients 100-150.

The three learned pdfs (retina, cornea, and fingertip) were obtained from OCT volume datasets which are separated from the tested datasets. The energy spectral densities have structural characteristics which are different for the three tissues (Fig. 4.1). The learned pdf is obtained according to 4.6 using one B-scan for each tissue.

To demonstrate the sampling distribution, histogram of spectral coefficients for 70% sampling rate is presented (Fig. 4.2) for each tissue. The sampling pattern is created for optimizing CS reconstruction performance. In the conventional CS, the sampling distribution is uniform across the entire scene.

In the implementation of the conventional CS as well as the EGCS, the OCT non-local regularization reconstruction method [83] is used for efficient OCT image reconstruction

4.3.2 SD-OCT - Experimental Results and Discussions

To evaluate the performance of the proposed EGCS learning approach to compressive SD-OCT, a number of experiments were conducted. For comparison purposes, the conventional CS approach was also evaluated as a baseline reference where sparse random sampling locations are distributed uniformly. In this case, Γ (4.6) is distributed uniformly. To evaluate the effectiveness of the proposed method, a series of OCT images acquired in-vivo from the human i) retinal fovea (Fig. 4.3(a)), ii) cornea (Fig. 4.4(a)), and iii) fingertip (Fig. 4.5(a)).

The images were acquired with a research grade, high-speed, SD-OCT system [111], operating at 1060nm wavelength, that utilizes a super-luminescent diode (λ_c = 1020nm, $\delta\lambda$ = 110nm, P_{out} = 10mW) and a 47kHz InGaAs linear array, 1024 pixel camera (SUI, Goodrich) interfaced with a high performance spectrometer (P&P Optica). The SD-OCT system provides 3μm axial and 15μm lateral resolution in the human corneal and fingertip tissue, and 6μm axial resolution in the human retina. The OCT images were acquired from healthy subjects using an imaging procedure carried out in accordance with the University

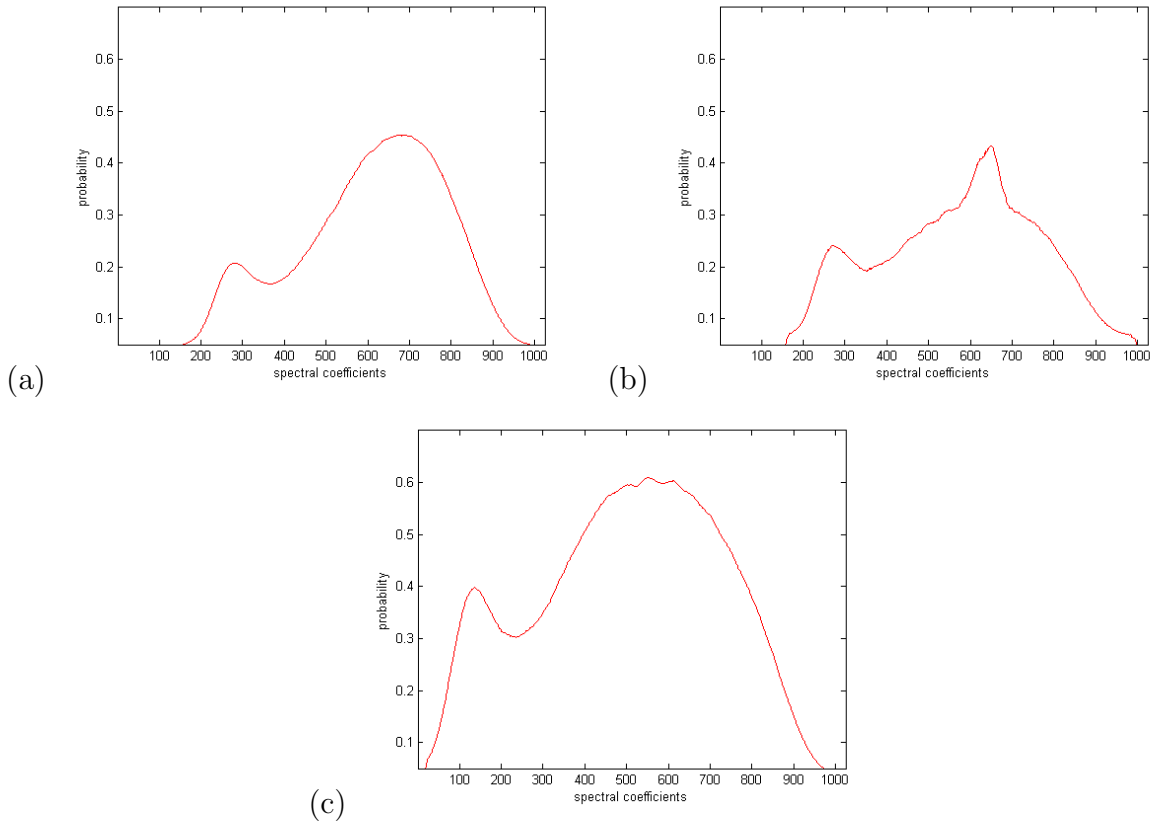


Figure 4.1: Sampling data pdf Γ (4.6) obtained from EGCS learning approach based on different type of tissues. (a) Retina, (b) Cornea, and (c) Fingertip

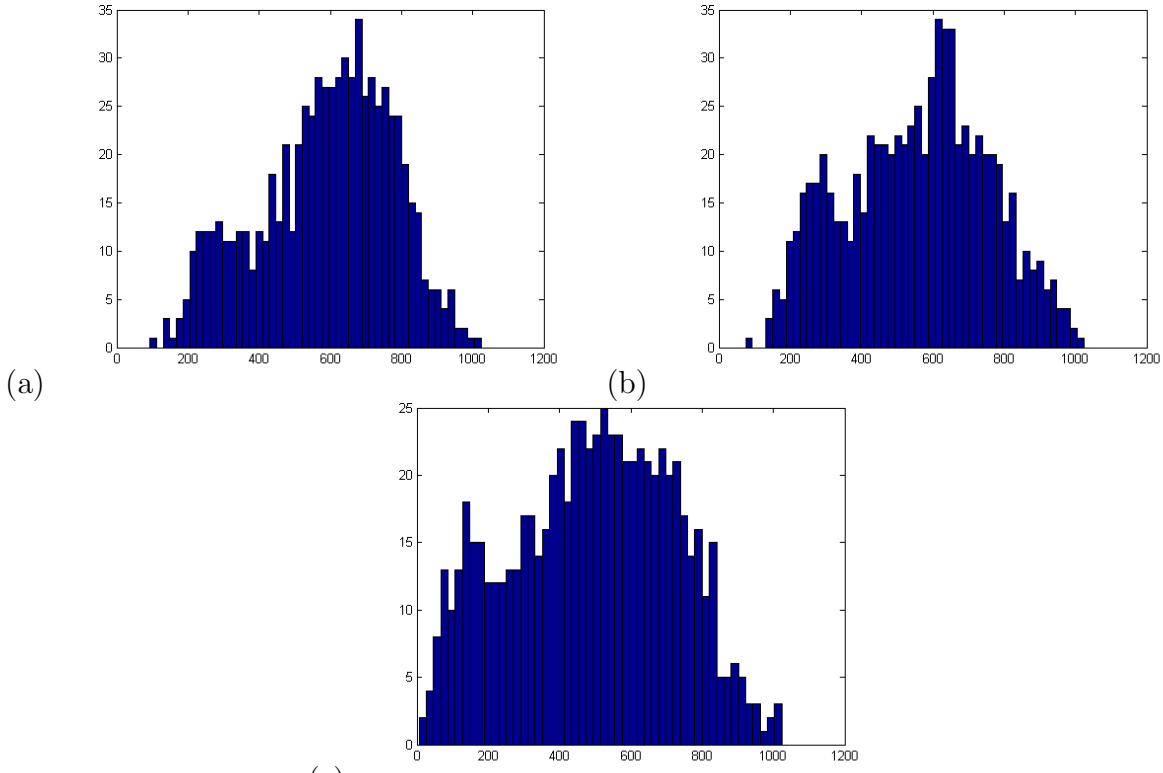
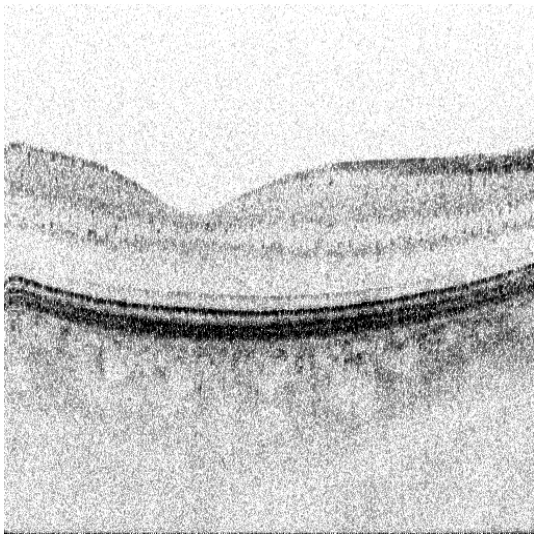
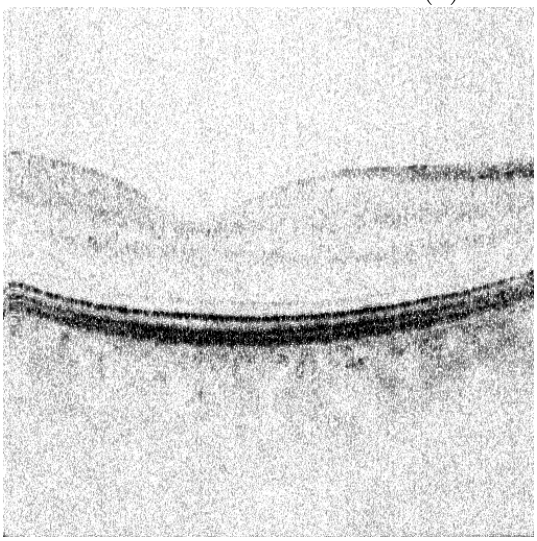


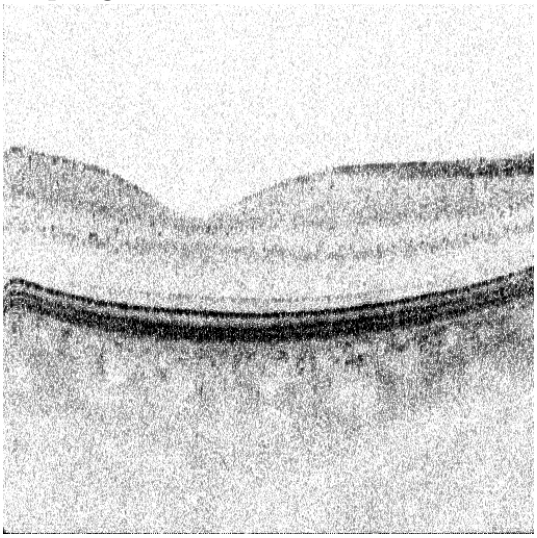
Figure 4.2: Example of EGCS sampling histogram at 70% sampling rate guided by learned pdfs of different tissues. The EGCS learned sampling distribution follows the data energy spectral densities (Fig. 4.1) to optimize CS reconstruction. In the conventional CS case, the sampling is uniformly distributed, not considering underlying data: (a) Retina, (b) Cornea, and (c) Fingertip



(a) 100% sampling

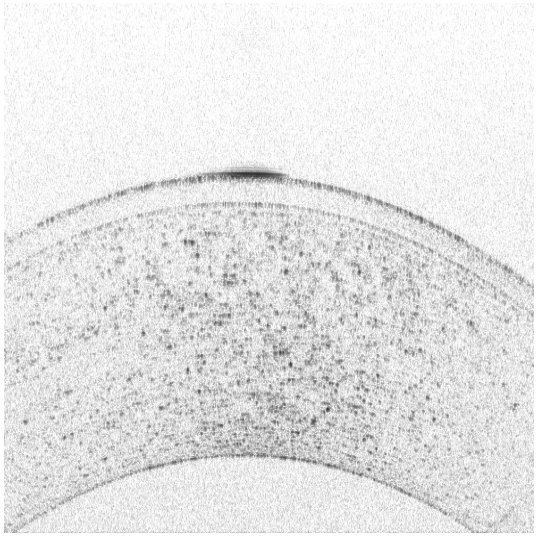


(b) CS

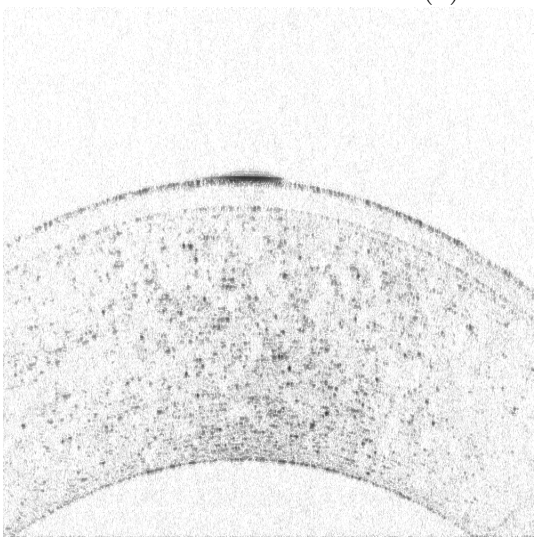


(c) EGCS

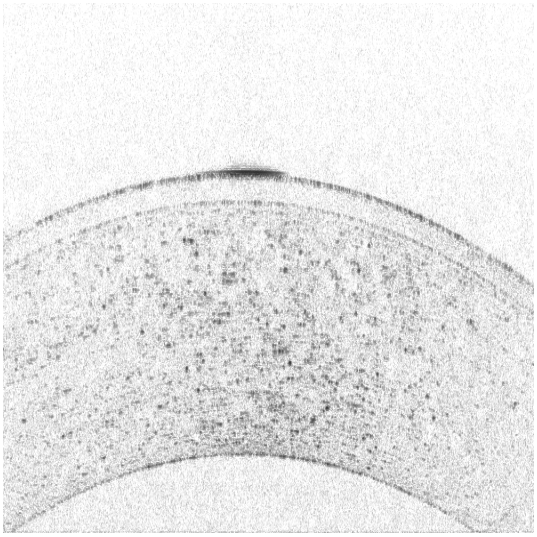
Figure 4.3: Reconstruction results from 50% of the acquired human retinal fovea data. The reconstruction results using 100% of the samples are provided as a reference.



(a) 100% sampling

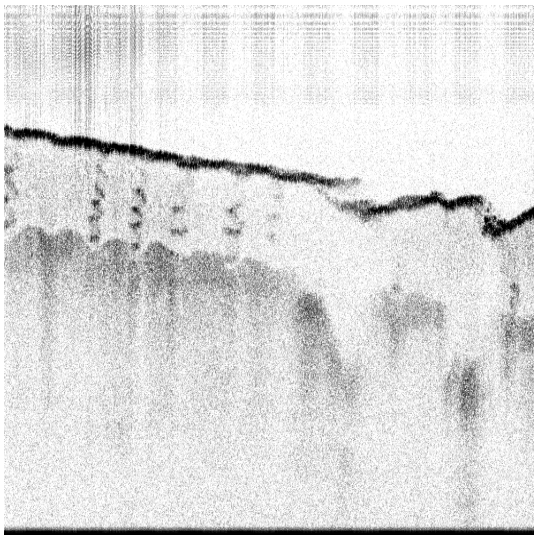


(b) CS

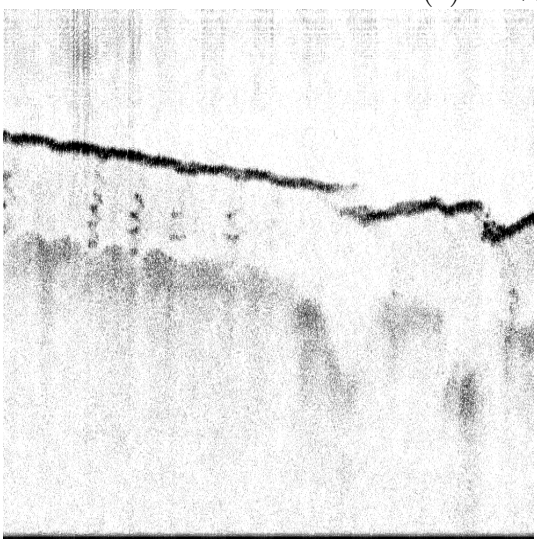


(c) EGCS

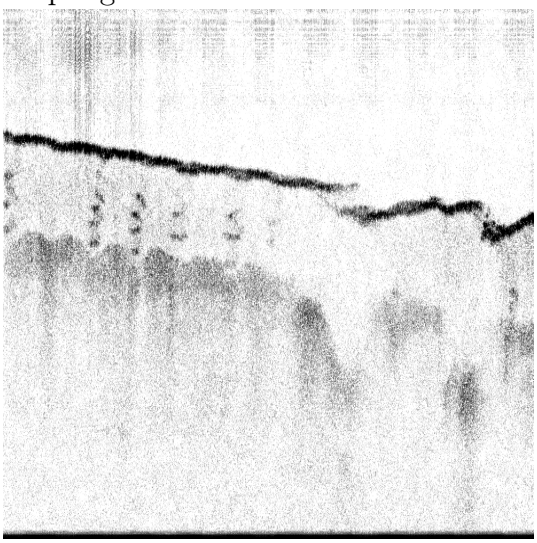
Figure 4.4: Reconstruction results from 50% of the acquired human corneal data. The reconstruction results using 100% of the samples are provided as a reference.



(a) 100% sampling



(b) CS



(c) EGCS

Figure 4.5: Reconstruction results from 50% of the acquired human fingertip data. The reconstruction results using 100% of the samples are provided as a reference.

of Waterloo research ethics regulations. To reconstruct OCT images from only a percentage of the camera pixels, the original data is sampled in the spectral domain using a uniformly distributed pseudo-random mask in the CS case. The obtained spectral data was then used to populate the k-space grid according to the known functional dependency of wavenumber on the pixel index [84].

A systematic assessment of the reconstruction performance of the different methods is performed. Reconstruction performance is evaluated by computing the peak signal to noise ratio (PSNR) for a wide range of sampling rates. The PSNR metric was computed as follows:

$$PSNR = 10 \cdot log_{10}\left(\frac{max(f)^2}{MSE}\right) \tag{4.7}$$

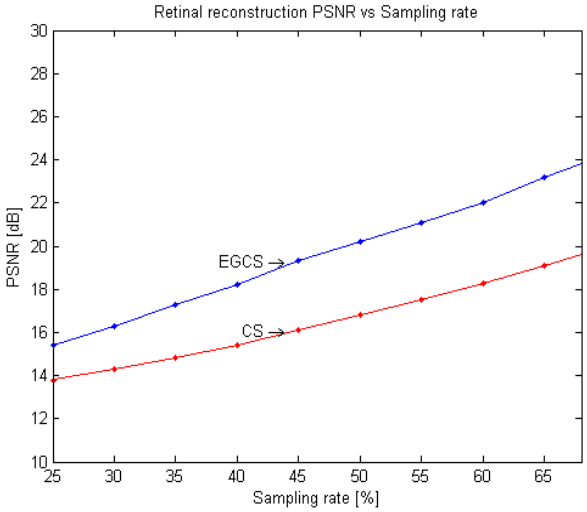
and (MSE) was defined as mean squared error between original image and reconstructed image:

$$MSE = \frac{1}{N}\sum_{n\in\Omega_N}\left(f(n) - \widetilde{f}(n)\right)^2 \tag{4.8}$$

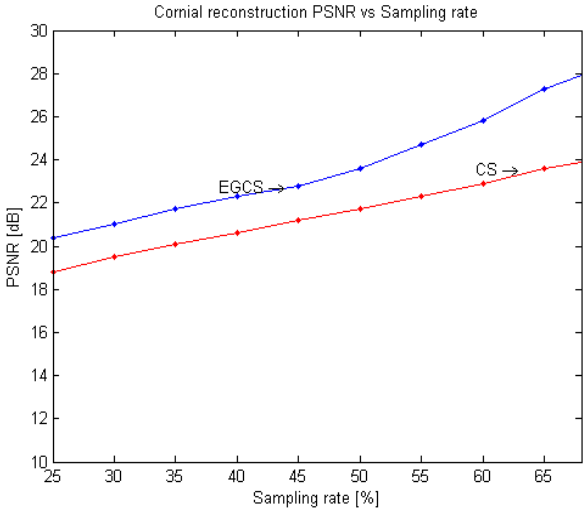
where $f(n)$ is original image, $\widetilde{f}(n)$ is reconstructed image, and N is the number of pixels in each image.

For illustrative purposes, the PSNR was measured for retinal, corneal and fingertip measurements reconstructed across the range of 25% and 70% of the camera pixels

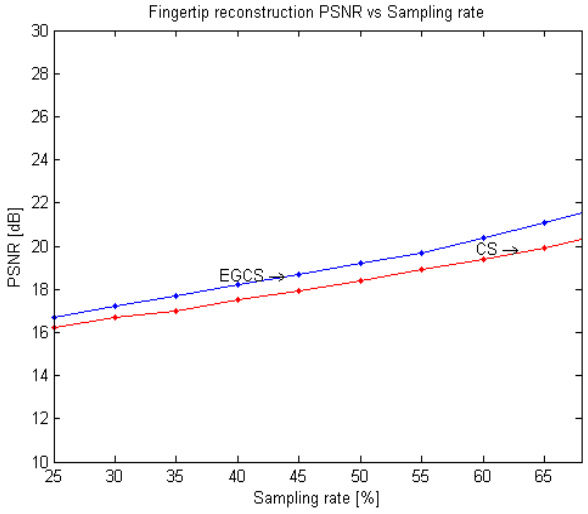
From the PSNR vs. sampling rate plot shown in Fig. 4.6, the PSNR achieved using the EGCS outperforms conventional CS through the entire tested sampling rate and for the different OCT images. The average difference between the EGCS and conventional CS for the tested sampling rate of 25%-70% is corneal: 2.3dB, fingertip: 0.8dB and retinal: 3.1dB. Through the experiments, the training datasets were separated from the testing dataset and all data were obtained from the same volumetric data. More specifically, one plane of the volumetric dataset is dedicated for training and other planes are used for testing. To validate the concept even further, additional processing was performed, where the learning dataset is based on a different retinal eye tissue which means one volumetric dataset was used for training and another volumetric dataset for testing. The average PSNR for this case is 3.0dB, which is very similar to the experiment where the training dataset is obtained from the same volumetric data with average PSNR of 3.1dB. The OCT light source is selected appropriately according to the tissues type, therefore the light source energy spectral distribution can be considered as previous knowledge that can guide EGCS. The PSNR vs. sampling rate plot shown in Fig. 4.7 compares reconstruction performance



(a) retina



(b) cornea



(c) fingertip

Figure 4.6: PSNR vs. sampling rate for cornea, retina and fingertip measurements.EGCS outperforms conventional CS through the entire tested sampling rate and for the different OCT images. The average difference between the EGCS and conventional CS for the tested sampling rate of 25%-70% is corneal: 2.3dB, fingertip: 0.8dB and retinal: 3.1dB.

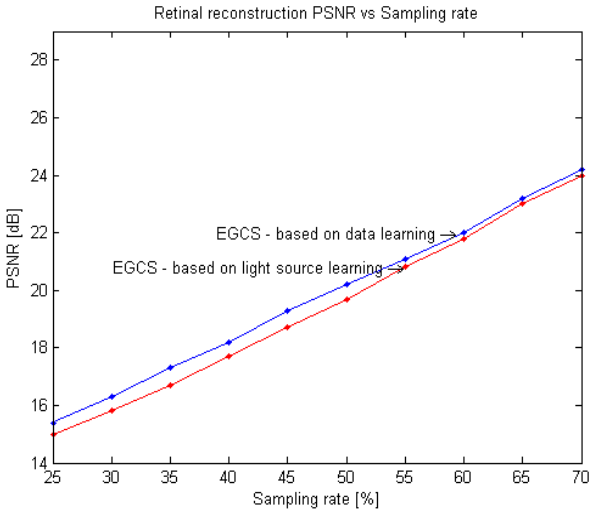
by learning process based on data energy spectral distribution and learning based on light source energy spectral distribution. Even though learning based on data provides better performance, the difference is small. As transparent the tissue is as small the difference. For the cornea case, which is transparent tissue, the performance is almost identical with difference average of 0.3dB. In the case of retina the difference is slightly larger (average of 0.4dB) and the largest difference is in the fingertip case with 0.7dB in average. In this case there is a larger shift in the spectrum since the tissue less transparent compared to the eye tissue (Fig. 4.7).

Furthermore, a qualitative visual assessment was performed on the reconstructed data to investigate the reconstruction performance and the preservation of details achieved using the tested methods at 50% sampling rate.

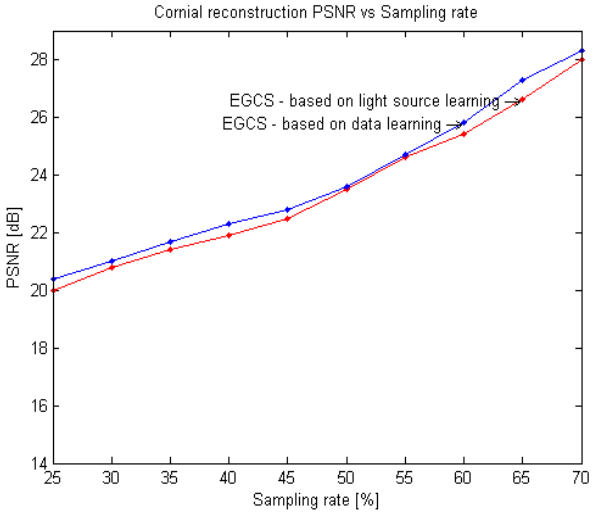
Figures 4.3, 4.4, and 4.5 show examples for each of the three types of in-vivo human OCT imaging data, each reconstructed using the two reconstruction methods from spectral data acquired using 50% of the camera pixels. The human retina (Fig. 4.3) contains a number of morphological details. Cross-sections of retinal capillaries (circular black features in the retina), as well as blood vessels. The EGCS method results in noticeably better image quality as compared to that produced using the conventional CS method, although the contrast of the individual retinal layers is not as good as in the original image. The image reconstructed using the EGCS approach is closer to the image reconstructed from 100% of the acquired samples.

The human cornea (Fig. 4.4) contains a number of morphological features of different size and optical properties. As observed in Fig. 4.4, the conventional CS method results in an image where most of the layers are still visible, however, the overall contrast of the image is drastically lower as compared to the image reconstructed from 100% of the acquired samples. The EGCS approach result in significantly better reconstruction of the corneal morphological details, as well as higher image contrast as compared to the conventional CS method. The image reconstructed using the EGCS approach is closer to the image reconstructed from 100% of the acquired samples.

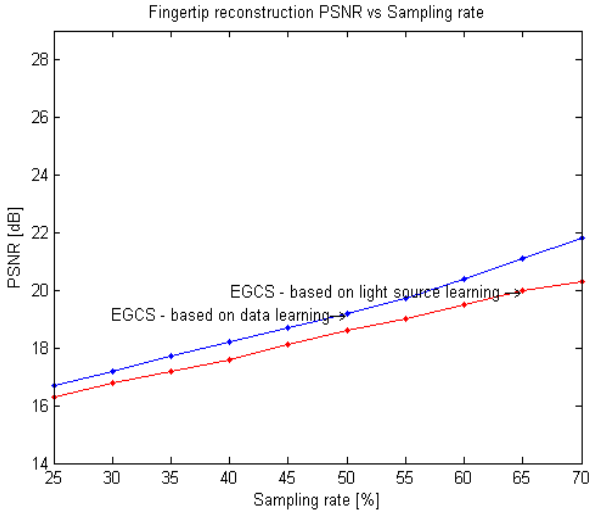
The human fingertip contains spiral shaped sweat glands in the skin epithelial region, which are clearly visible in Fig. 4.5. The conventional CS method and EGCS method have almost similar reconstruction performance (Fig. 4.5) with only 1 dB PSNR difference (Fig. 4.6(d)). The reason is the fingertip energy spectral distribution. The energy is more concentrated in certain frequencies at the retinal and corneal datasets (Fig. 4.1(a) and Fig. 4.1(b)) while the energy is more spread in the fingertip case (Fig. 4.1(c)). The EGCS method optimizes the sampling probability according to underlying data. Since the energy is spread in the fingertip case, an optimized sampling probability is closer



(a) retina



(b) cornea



(c) fingertip

Figure 4.7: PSNR vs. sampling rate for cornea, retina and fingertip measurements based on two learning cases:learning based on data energy spectral distribution and learning based on light source energy spectral distribution. The OCT light source is selected appropriately according to the tissues type, therefore the light source energy spectral distribution can be considered as previous knowledge that can guide EGCS. Even though learning based on data provides better performance the difference is small. As transparent the tissue is as small the difference.

to uniform probability in comparison to retinal or corneal data. The proposed EGCS approach produces reconstructed OCT data with higher PSNR values for all levels of camera pixel under-sampling when compared to the conventional CS method. The EGCS method outperformed the conventional CS even for fingertip data even though not as much as for retinal and corneal data.

4.3.3 Bridge Structure Vibration Data Compression Application

Structural monitoring through a large set of uncorrelated measured signals is common. In the last decade, blind source separation (BSS) methods [4] have been used for processing those large sets of signals towards modal identification of various civil and mechanical structures. The application of BSS methods to flexible structures such as bridges have been investigated for a large number of numerical and experimental cases [4, 70, 119]. These results clearly manifest the potential of using BSS for a broad range of structural modal identification problems.

Structural monitoring can be performed using sensor technology for measuring vibrations in multiple locations and then been received at remote location. Popular ambient modal identification methods published in the literature utilize measurements simultaneously collected from a large array of sensors [14] to obtain the modal characteristics such as natural frequency, damping and mode shape matrix of the system. The process involves a centralized processing unit performing the necessary processing tasks [89].

Recent advances in micro-electro-mechanical systems (MEMS) and wireless smart sensor networks (WSSNs) have provided an affordable hardware environment that can be deployed for large scale structures [14, 60, 87, 159]. The main advantage of WSSNs is that limited processing can be undertaken at the sensor level, which allows performing identification and health monitoring tasks in a de-centralized fashion, thereby reducing the need for a centralized processing unit. [124, 159, 135]. In structure monitoring applications, large volume of data is collected and processed. In those applications, reducing the volume of transmitted data still remains a significant issue to be addressed.

The objective of data compression methods is to find a concise representation of a signal that is possible with an acceptable level of distortion upon reconstruction [124]. Several forms of lossless and lossy data compression techniques have been adopted in seismic data and in vibration response data [7, 88, 158].

Compressive sensing can be a good candidate to address this challenge within the framework of BSS modal identification [124]. In this applications [124], the volume of data is large and the node hardware resources are limited. Each sensor measures the requiered

signal and produces spectral coefficients. Due to hardware limitations, the node can handle only a small subset of spectral coefficients while maintaining high data fidelity for bridge resonance monitoring is required.

Similar to the direct learning approach for construct a sampling pattern based on the nonparametric modeling used for SD-OCT application (4.3.1), a direct learning approach in the frequency domain is implemented to perform an efficient signal compression for bridge resonance monitoring data. The sub-set data selection is done according to energy-guided statistical learning approach based on the underlying data in the frequency domain. By doing so, the proposed method optimizes the compressed sub-set data to minimize the number of samples needed while maintaining high reconstruction quality. The learning is based bridge structure resonance frequencies.

Consider the time domain signal d and its frequency domain representation as one dimension D (4.1), where d represents vibrations monitoring signal which is measured by each sensor on the bridge. To reduce data volume, the system selects a sub-set of spectral coefficients D. EGCS has great potential for reducing the data volume by representing the data by much fewer spectral coefficients. EGCS selects the spectral coefficients randomly according to learned pdf while CS selects the coefficients randomly and uniformly distributed. The learning method is based on previous knowledge regarding the bridge structure resonance spectrum.

The proposed method is illustrated using a full-scale ambient vibration data obtained from UCLA structure [79, 124]. For illustration purposes, the learned bridge structure resonance pdf from different type of bridges are demonstrated in Fig. 4.8. pdf of synthetic bridge signals are presented at Fig. 4.8(a), Fig. 4.8(b), and Fig. 4.8(d) where Fig. 4.8(a) is the pdf of noiseless synthetic bridge signal with two resonance frequencies, Fig. 4.8(d) is pdf of noisy synthetic bridge signal with two resonance frequencies and Fig. 4.8(b) is pdf of noiseless synthetic bridge signal with four resonance frequencies. In addition an example of pdf of real bridge signal is demonstrated at Fig. 4.8(c). In the case of sampling pdf learned from the synthetic signals, the resonance frequencies can be seen clearly. The real signal contain multiple resonance frequencies at different energy levels and is contaminated by noise.

The four bridges under test have different resonance frequencies therefore represented by different energy spectral densities (Fig. 4.8). The learned pdf is obtained by 4.6 according the bridge structure data. The sampling distribution is illustrated using histogram of spectral coefficients for 40% sampling rate (Fig. 4.9) for each bridge. The sampling pattern is implemented for optimizing CS reconstruction performance. In the conventional CS, the sampling distribution is uniform across the entire scene.

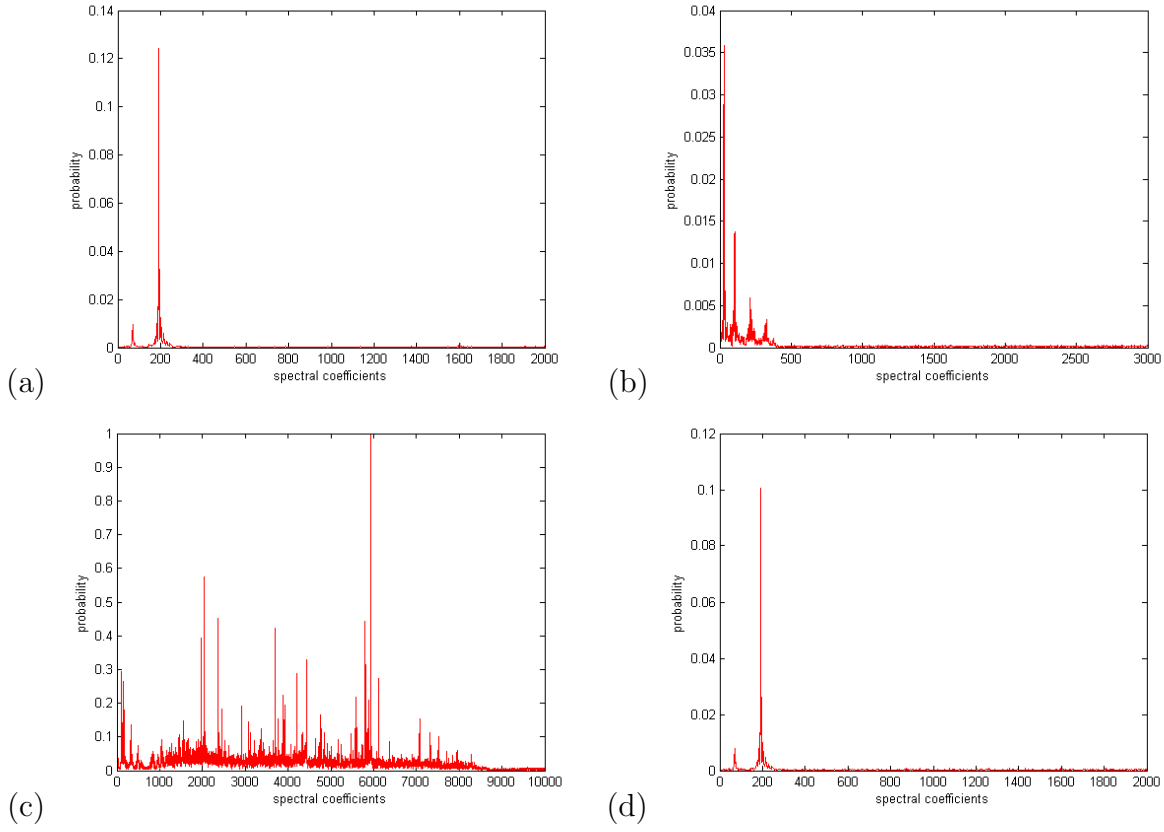


Figure 4.8: Sampling data pdf (Γ 4.6) obtained from EGCS learning approach based on different type of bridge resonance measurement signals:
(a) Noiseless synthetic bridge signal with two resonance frequencies ("Bridge01")
(b) Noiseless synthetic bridge signal with four resonance frequencies ("Bridge02")
(c) Real and noisy bridge signal ("Bridge03")
(d) Noisy synthetic bridge signal with two resonance frequencies ("Bridge04")

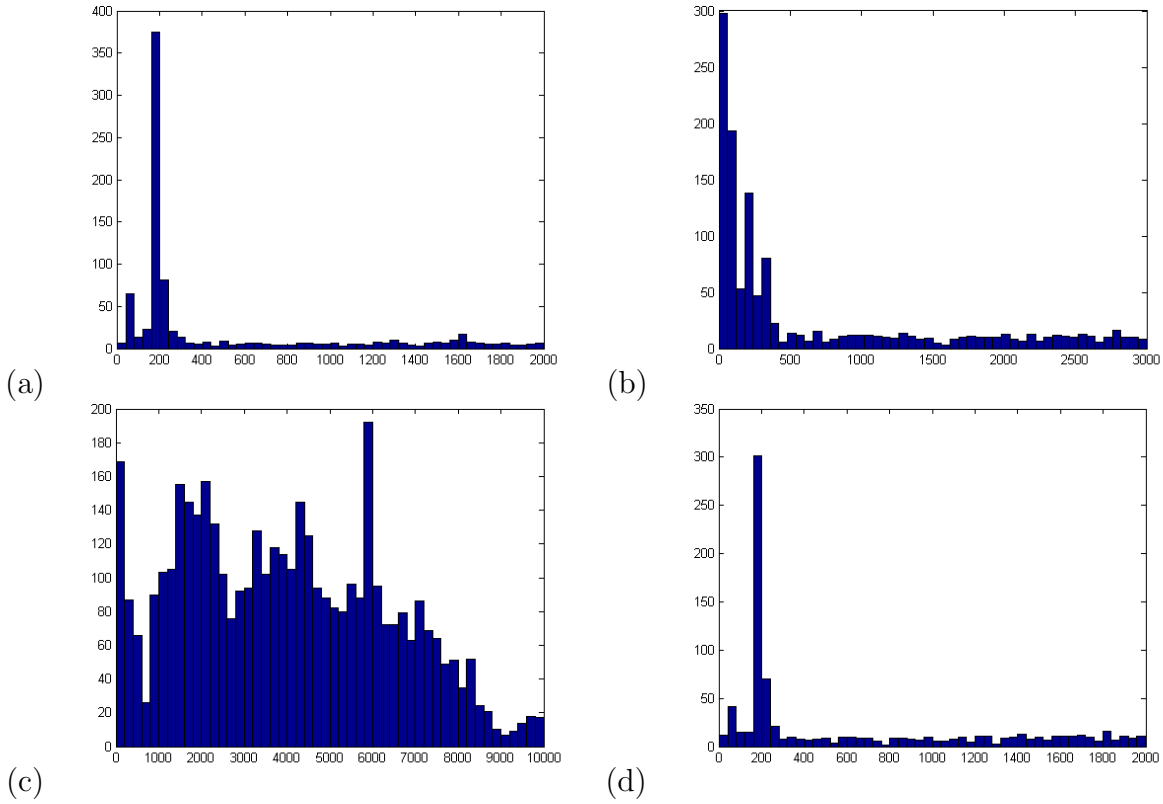


Figure 4.9: Example of EGCS sampling histogram at 40% sampling rate guided by learned pdfs of different bridges. The EGCS learned sampling distribution follows the data energy spectral densities (Fig. 4.8) to optimize CS reconstruction. In the conventional CS case, the sampling is uniformly distributed, not considering underlying data.
(a) Noiseless synthetic bridge signal with two resonance frequencies ("Bridge01")
(b) Noiseless synthetic bridge signal with four resonance frequencies ("Bridge02")
(c) Real and noisy bridge signal ("Bridge03")
(d) Noisy synthetic bridge signal with two resonance frequencies ("Bridge04")

4.3.4 Bridge Structure - Experimental Results and Discussions

In the first set of experiments, a reconstruction performance evaluation is performed at different compression rates. The analysis is performed with four signals:

- noiseless synthetic bridge signal with two resonance frequencies ("Bridge01")
- noiseless synthetic bridge signal with four resonance frequencies ("Bridge02")
- real and noisy bridge signal ("Bridge03")
- noisy synthetic bridge signal with two resonance frequencies ("Bridge04")

Examples of fully sampled data in the time domain used for the experiments, are shown in Fig. 4.10.

The reconstruction performance is evaluated by computing the signal-to-noise ratio of each signal at a wide range of compression rates (75% - 99%). For comparison, reconstruction of conventional CS at the same conditions is provided as well.

Based on the SNR vs. compression rate plot shown in Fig. 4.11, the SNR achieved using the learning process based on data energy spectral distribution method outperforms significantly conventional CS through the entire tested sampling rate and for the different signals. The average difference between the EGCS and conventional CS for the tested sampling rate of 75%-99% is: "Bridge01" 38.4 dB, "Bridge02" 28 dB, "Bridge03" 8 dB and "Bridge04" 20 dB.

For visualization purposes, several examples of signal reconstruction in high compression rate (85% and 95%) are shown in Fig. 4.12. To emphasize the reconstruction performance and since the signal contains many samples, a zoom-in presentation is presented in Fig. 4.13.

The reconstruction examples demonstrated that signal reconstruction based on learning the data energy spectral distribution method provides very high reconstruction quality. The synthetic bridge signal with two resonance frequencies ("Bridge01") is almost perfectly reconstructed at 95% compression rate which means only 5% of data is used. The noise contaminated reconstruction of the same signal ("Bridge04") shows slightly lower reconstruction performance but still excellent performance at very high compression rate. The more complicated signal - synthetic bridge signal with four resonance frequencies ("Bridge02") shows also very high reconstructed quality at 95%. The proposed method achieved very high signal reconstruction performance even for the real noisy bridge signal ("Bridge03") which contains multiple resonance frequencies and is contaminated by noise.

Differently from the conventional CS which distribute the samples uniformly across the spectrum, the EGCS sampling distribution follows the data energy spectral densities (Fig. 4.10) to optimize CS reconstruction.

Bridge Structure - summary

The main objective in bridge structure application, is reducing data volume significantly while maintaining data integrity. The system selects a sub-set of spectral coefficients to have a compressed representation of the data. The data is structured in the frequency domain with distinct resonance frequencies which are typical to a specific bridge structure. While CS method under-sample the entire spectrum randomly and uniformly distributed, EGCS under-sample spectral coefficients randomly and guided by a learned distribution. Therefore, and supported by test results (4.3.4), EGCS achieves much better compression rates compared to conventional CS, at similar reconstruction performance.

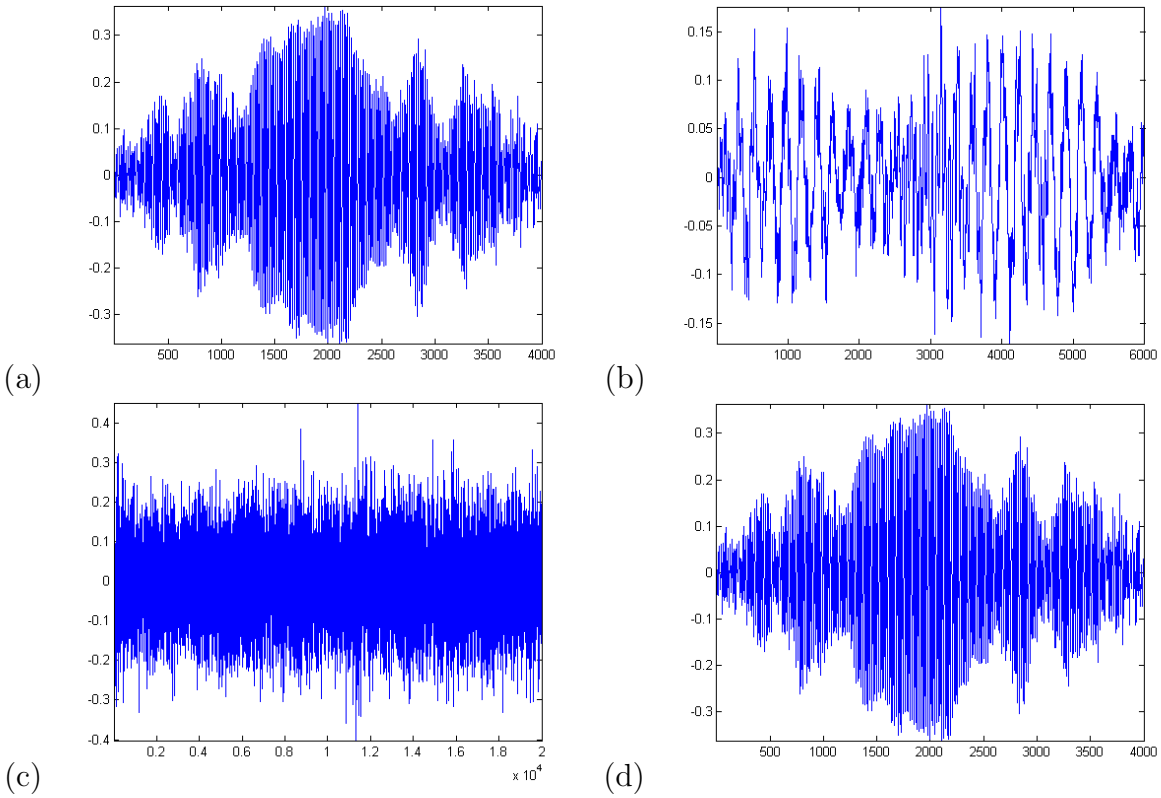


Figure 4.10: Example bridge resonance measurement signals used for testing:
(a) Original fully sampled synthetic bridge signal with two resonance frequencies ("Bridge01"),
(b) Original fully sampled synthetic bridge signal with four resonance frequencies ("Bridge02"),
(c) Original fully sampled real noisy bridge signal ("Bridge03"),
(d) Original fully sampled synthetic noisy bridge signal with two resonance frequencies ("Bridge04")

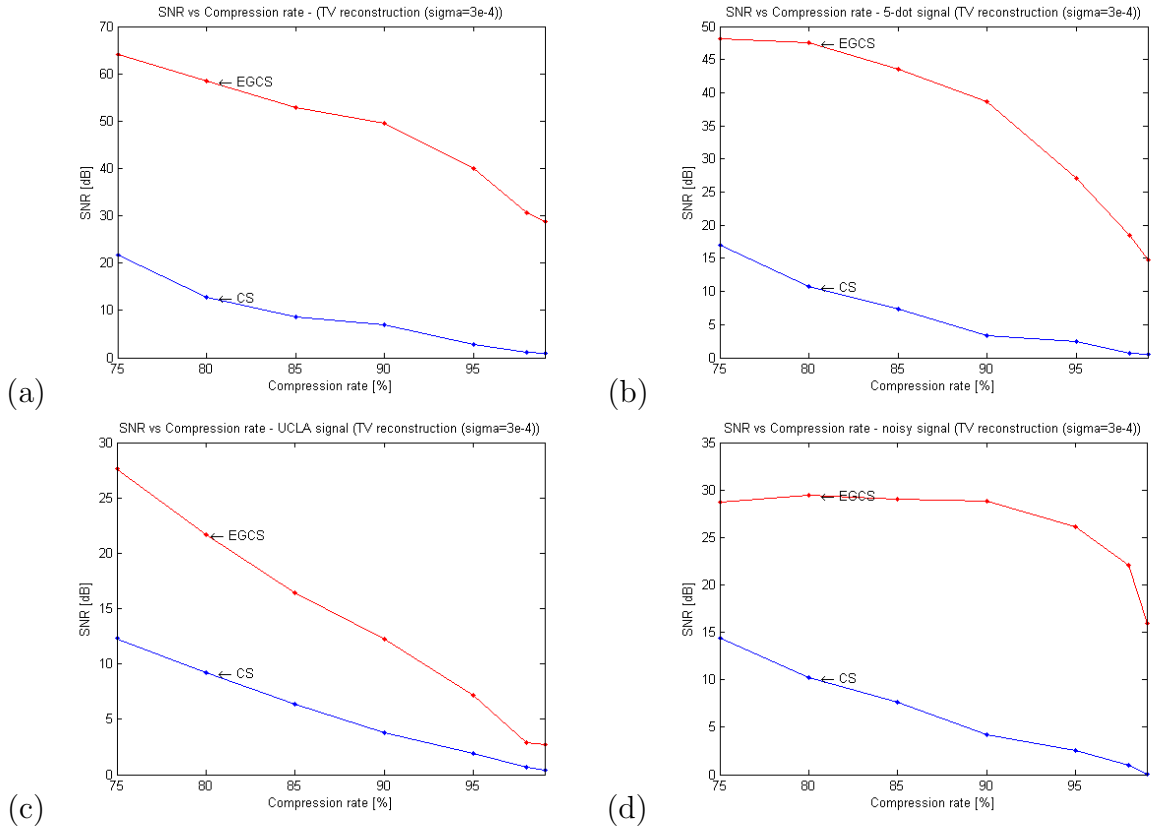


Figure 4.11: SNR vs. compression rate for bridge resonance measurements signals obtained from EGCS learning approach based on different type of bridge resonance data sets compared to conventional CS. SNR achieved using the learning process based on data energy spectral distribution method outperforms significantly conventional CS through the entire tested sampling rate and for the different signals. The average difference between the EGCS and conventional CS for the tested sampling rate of 75%-99% is: "Bridge01" 38.4 dB, "Bridge02" 28 dB, "Bridge03" 8 dB and "Bridge04" 20 dB:
(a) Noiseless synthetic bridge signal with two resonance frequencies ("Bridge01"),
(b) Noiseless synthetic bridge signal with four resonance frequencies ("Bridge02"),
(c) Real and noisy bridge signal ("Bridge03"),
(d) Noisy synthetic bridge signal with two resonance frequencies ("Bridge04")

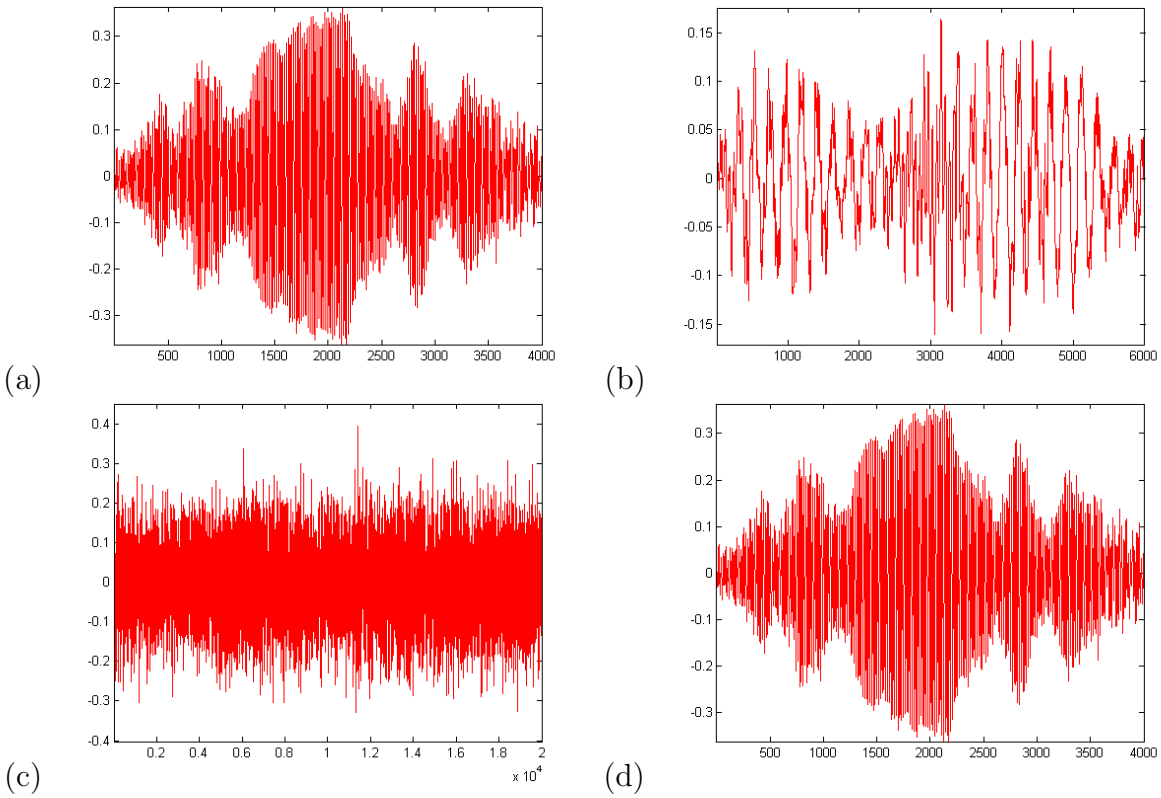


Figure 4.12: Signal reconstruction of bridge resonance measurement data:
(a) Synthetic bridge signal reconstruction at 95% compression rate with two resonance frequencies ("Bridge01"),
(b) Synthetic bridge signal reconstruction at 95% compression rate with four resonance frequencies ("Bridge02"),
(c) Real noisy bridge signal reconstruction at 85% compression rate ("Bridge03"),
(d) Synthetic bridge noisy signal reconstruction at 95% compression rate with two resonance frequencies ("Bridge04")

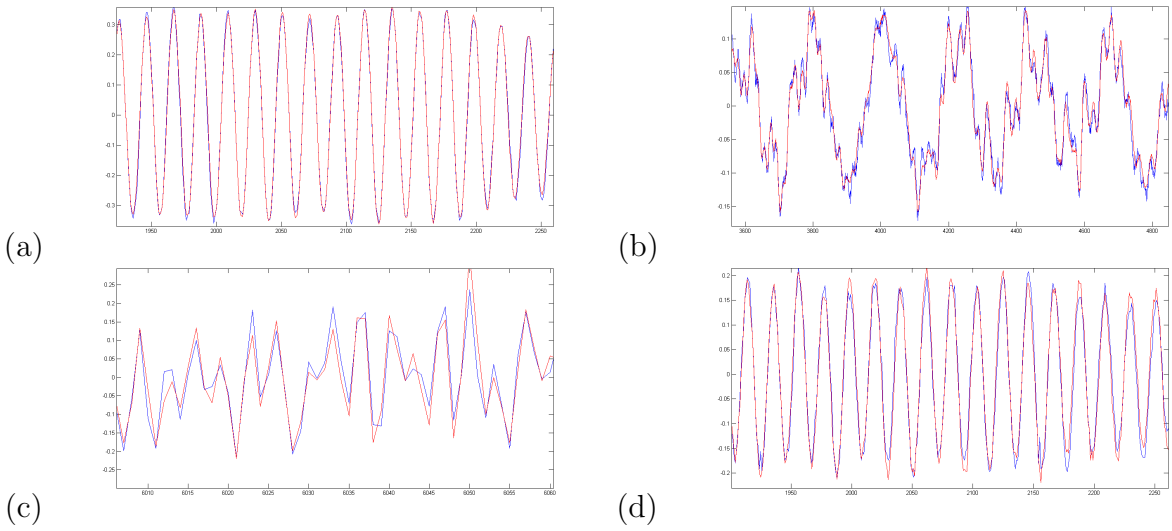


Figure 4.13: Zoom-in signal reconstruction of bridge resonance measurement data. The BLUE color is the original signal while RED is the reconstructed signal.
The synthetic bridge signal ("Bridge01") is almost perfectly reconstructed where only 5% of data is used. The noise contaminated reconstruction of the same signal ("Bridge04") shows slightly lower reconstruction performance. The synthetic bridge signal with four resonance frequencies ("Bridge02") shows also very high reconstructed quality at 95%. EGCS achieves very high signal reconstruction performance even for the real noisy bridge signal ("Bridge03") which contains multiple resonance frequencies and is contaminated by noise:
(a) Synthetic bridge signal reconstruction at 95% compression rate with two resonance frequencies ("Bridge01"),
(b) Synthetic bridge signal reconstruction at 95% compression rate with four resonance frequencies ("Bridge02"),
(c) Real noisy bridge signal reconstruction at 85% compression rate ("Bridge03")
(d) Noisy synthetic bridge signal reconstruction at 95% compression rate with two resonance frequencies ("Bridge04")

Chapter 5

Nonparametric Indirect Learning Approach to Sampling Pattern Generation

As previously shown (4) the direct learning approach is efficient for optimizing sensing probability density for improving compressive sensing reconstruction performance. As previously described, the direct learning approach relays on learning the nonparametric model directly based on underlying data. However, in many situations, the areas of interest within the signal are not characterized directly by the data itself, but by an indirect derived product of the data. For example, objects of interest within an image may be characterized by the structural characteristics with highly salient features. In this chapter, an indirect learning approach based on saliency in the spatial domain, with applications to robotic vision and fluorescence microscopy, will be explored to illustrate the usefulness of such data-adaptive approaches. The core of this indirect learning approach is based on learning an optimized mapping between sampling probability and derived features. Since we are primarily interested in preserving such salient characteristics when aiming to achieve high reconstruction performance, one can learn the model indirectly via extracting derived features and then learn a nonparametric model for the features in order to optimize reconstructed image quality.

5.1 Spatial Domain Saliency-guided Sparse Measurements Model

To address this important issue, a saliency-guided sparse measurements model is proposed that makes use of these saliency characteristics of objects of interest to improve the performance of CS systems for data reconstruction under such practical scenarios.

Consider the scene being measured using a measurement system to contain $R \times C$ sampling locations organized in a finite, separable, rectangular lattice $\Omega_{R\times C}$, with the measured value at each sampling location (2.21), where the sampling domain is the spatial domain $\Omega_{R\times C} \in \mathbb{R}^2$. Subsequently, $\Omega_{R\times C}$ is partitioned into two complementary sets Ω_T and Ω_T^c such that

$$\Omega_{R\times C} = \Omega_T \cup \Omega_T^c, \text{ with } \Omega_T \cap \Omega_T^c = \emptyset, \tag{5.1}$$

whose cardinalities are equal to $\#\Omega_T = T$ and $\#\Omega_T^c = RC - T$, respectively. The subset Ω_T denotes sampled locations and Ω_T^c denote unsampled locations. Subset Ω_T is generated based on a function $\Gamma(r, c)$ that provides a quantitative measure of saliency at sampling location (r, c)

$$\Gamma(r, c) \to [0, 1] \; \forall (r, c) \tag{5.2}$$

where $\Gamma(r, c)$ returns a sparsity level value between 0 and 1 for every location (r, c).

The image function $f(r, c) : \Omega_{R\times C} \to \mathbb{R}$ is assumed to be bounded, i.e., $\max_{r,c} |f(r, c)| < \infty$, given that it represents discrete quantized pixel values. Let $\{\varphi_k\}_{k=1}^M$ denote a collection of $M \leq RC$ discrete sampling functions. The linear measurements of f can be described in a general form (2.22) or in a more compact vector form

$$y_k = \varphi_k^t \; f + \epsilon_k \tag{5.3}$$

where $f \in \mathbb{R}^{RC\times 1}$ is a concatenated version of the original image and where φ_k^t is the sampling function φ_k transpose. In the case of a saliency-guided sparse measurements model, when $M < RC$, the sampling basis is modified to account for the lack of observations at locations in Ω_T^c the sparse sampling function is:

$$\varphi_{T,k}(r, c) = \begin{cases} \varphi_k(r, c), & \text{if } (r, c) \in \Omega_T, \\ 0, & \text{if } (r, c) \in \Omega_T^c. \end{cases} \tag{5.4}$$

Consequently, the saliency-guided sparse measurements model is defined in (3.8), where $\varphi_{T,k}$ values are realizations of a random variable z whose probability density p_T which is defined in (3.9) and where $p_S(z) = \mathcal{N}(z|0, 1)$.

5.2 Indirect Learning Model in the Spatial Domain

Assume that saliency can be quantified based on structural variations. Given that salient objects of interest within a scene can have structural characteristics at different scales, it would be useful to quantify such variations using a multi-scale approach. Therefore, in the proposed realization, a saliency map $S(r,c)$ [71] is constructed where high values reside at locations (r,c) for situations characterized by large variations at different scales based on a multi-scale extension of the spectral residual approach [71]. In this extended approach, the log-spectrum is analyzed in the spectral domain at different scales to extract the spectral residual at that scale. Therefore, the saliency map $S_{sc}(r,c)$ is created according the spectral residual approach [71] for every scale $sc = 1, 2, \cdots q$.

Finally, saliency maps from all scales are fused to obtain the final saliency map $S(r,c)$:

$$S(r,c) = E\left\{S_{sc}(r,c)\right\}, sc = 1, 2, \cdots q \tag{5.5}$$

where $E\{.\}$ denotes the fusion function (the average in this case) and q is the number of scales. In this practical realization, $q = 3$ provides strong results.

The saliency map is then clustered into R clusters θ_i, $i = \{1, \ldots, R\}$ with common saliency properties [1] implemented by using the Hill-Climbing algorithm [1, 72].

For this practical realization, based on the obtained saliency map, let us define the continuous saliency function $\Gamma(r,c)$ (5.2):

$$\Gamma(r,c) = G(\theta_i^{\mu}),\ G(\theta_i^{\mu}) \in [0,1]\ \ \forall\, i = \{1, \ldots, R\} \tag{5.6}$$

where $G(\theta_i^{\mu})$ returns a sparsity level value between 0 and 1 and is a function of the cluster θ_i index sorted by the cluster mean, which is the average saliency of region i. $G(\theta_i^{\mu})$ can be selected to be a function that emphasizes high saliency more then lower salient levels. All measurement locations that are not sampled due to sparse sampling are grouped into Ω_T^c. The collection of all sampled locations are included in Ω_T.

Through this saliency learning stage, a conventional CS procedure is used where $f(r,c)$ is sampled by sparse $\varphi_k^t(r,c)$ with the probability density p_T (3.9) where Γ represents uniform distribution that results in very few samples being acquired. The probability density of $\varphi_k^t(r,c)$, denoted by p_T, is defined as a Gauss-Bernoulli distribution, where $p_T(z) = 0$ with probability $(1-\pi)$ and $p_T(z)$ is Gaussian distributed with probability π. In this saliency learning stage, π is selected to be 0.1 (90% compression rate) at the first stage. Therefore, p_T is defined accordingly (5.7) for this example:

$$p_T = 0.9\Gamma\,\delta(z) + 0.1\Gamma\,\mathcal{N}(z|0,1), \tag{5.7}$$

In practical implementations the value for π is restricted to minimum value that will still produce useful information which is application depended. If there will be to many samples allocated for the saliency learning stage, there will be not sufficient number samples for the second stage. Therefore, for this realization, the minimum value for π is selected to be 0.1 (representing 90% compression rate) as the maximum compression rate where CS can still produce reasonably reconstructed data. The samples from the first stage are reused in the second stage. In extreme cases where the requiered sampling rate is lower than 10%, half of the allocated samples will be used for the first stage and the rest will be used for teh second stage.

5.3 Realization of Saliency-guided Modeling Approach in Spatial Domain

In this practical realization of the proposed model and based on the aforementioned saliency function and probability density distributions, the proposed saliency-guided sampling process can be implemented in two stages. In the first stage, which we will refer to as the saliency learning stage (5.2), a rough saliency map of the scene based on the reconstructed image using data from the first stage is computed. Hence, a subset Ω_T (3.1) is determined using uniform distribution under-sampling for this saliency learning stage.

In the second stage, $f(r, c)$ is sampled by $\varphi_k^t(r, c)$ with pdf p_T(3.9) to measure regions of high saliency with higher accuracy where Γ is determined through learning. In this practical realization, $G(\theta_i^\mu)$ can be selected as a parametric function or as an optimized nonparametric function (3.14). The mapping function G is a function of cluster θ_i index sorted by the cluster mean, which is the average saliency of region i. The parametric function was selected to be logarithmic to emphasizes high saliency more than lower salient levels while the nonparametric function is optimized by MCMC process (3.4.2).

The samples from subset Ω_T are used to create the sampling basis $\varphi_{T,k}$, with the acquired samples used to reconstruct the image at a higher reconstruction accuracy than the one obtained in the first stage.

To reconstruct the image, an $\ell 1$-based total variation minimization approach was employed [120]. The noise-free image f can be approximated via the following $\ell 1$-based total variation minimization formulation:

$$\arg\min_f \left\{ \lambda \|f\|_{TV_{l_1}} + \frac{1}{2} \|\Phi^t f - y\|_2^2 \right\} \tag{5.8}$$

where $\|\cdot\|_{TV_{l_1}}$ denotes the $\ell 1$-based anisotropic total variation norm defined by [9]:

$$\|x\|_{TV_{l_1}} = \sum_{r=0}^{R-1}\sum_{c=0}^{C-1}\left\{|x_{r,c} - x_{r+1,c}| + |x_{r,c} - x_{r,c+1}|\right\},\ x \in \mathbb{R}^{R\times C} \tag{5.9}$$

where $\lambda > 0$ is a regularization constant, and where $\|\cdot\|_2$ stands for ℓ_2 norm defined by

$$\|x\|_2 = \sqrt{\sum_{i=0}^{M-1} x_i^2} \tag{5.10}$$

The minimization problem can be solved using the fast iterative shrinkage-thresholding algorithm (FISTA) [9, 10].

5.3.1 3D Laser Measurements

The measurement of height/depth information is used in many applications such as parts inspection [76], reverse engineering, production validation [13], object recognition, object registration [12], and others [26]. Industrial systems use range measurement machines for manufacturing and quality assurance. Traditionally, mechanical systems were implemented for range measurement purposes. However, the use of optical measurement systems such as laser measurement systems for range measurements have become increasingly popular and widespread, and has been a strong focus for research in the last several decades [13, 26].

Range sensing and measuring is in general a time consuming process, especially if high resolution is required. Laser range measurement has made its progress from single light-spot scanning to a complex coded pattern that speeds up the scanning operation [29]. Despite the reduced scanning time achieved using complex coded patterns, single light-spot scanners have several important advantages over complex coded pattern-based systems. First, single light-spot scanners avoid geometry discontinuity due to occlusion that commonly affects the quality of complex coded pattern-based systems. Second, single light-spot scanners have a less complicated implementation. Therefore, the ability to acquire laser range data using single light-spot scanners in a highly efficient manner is desired for achieving high speed, high quality data acquisitions while maintaining low system complexity.

For the 3D laser measurements realization, a simplified version of saliency-guided approach in spatial domain was implemented, where saliency guided function is binary in

nature and not multi-scale (5.3). This simplified version is used where images are less complicated scenes. In other words the saliency function $\Gamma(r, c)$ has a value of 1 at location (r, c) for situations characterized by large spatial range variations, based on a frequency-tuned saliency map strategy [2] and 0 otherwise. The saliency function used in this simplified binary version is:

$$\Gamma(r, c) = 1, \text{ if } S(r, c) > \tau, \ \forall(r, c) \tag{5.11}$$

where

$$S(r, c) = \|I_\mu - I(r, c)\| \tag{5.12}$$

and where I_μ is the mean, $I(r, c)$ is the corresponding vector of the Laplacian of the Gaussian filtered image, and τ is the threshold value (set at two times the mean saliency $S(r, c)$ of a given image [2]).

In the first experiment, reconstruction performance was evaluated in situations where the measurements were free of noise. In the second experiment, reconstruction performance was evaluated in situations where the measurements were contaminated by different levels of noise. In the third experiment, a reconstruction repeatability test was performed to evaluate the consistency of reconstruction performance using the proposed approach.

Experimental Setup

The first three experiments were performed with 35 laser range datasets from the National Research Council (NRC) Three-dimensional Image Data Files [116] database. The measurement system used by the NRC institute is based on triangulation through a synchronized laser scanning apparatus that was developed at the Division of Electrical Engineering at the NRC [116]. The NRC database contained two data structures: i) raw data, and ii) interpolated data (z). Due to the scanner geometry (curved), the raw data exhibits distortion. The interpolated data provided by the NRC compensates for this geometry distortion, providing a symmetric lattice with corresponding height values. For experimental purposes, interpolated data was used with horizontal and vertical resolutions Δx and Δy in the range of 0.5mm - 2.0mm. Examples of fully sampled images (where each sampling location is measured) from the NRC range image database, along with corresponding noise contaminated versions used for the second experiment, are shown in Fig. 5.1 - Fig. 5.4.

Experiment 1 - Noiseless Range Measurements

In the first experiment, the reconstruction performance was evaluated in the situation where the measurements were free of noise. This was achieved via a parametric analysis

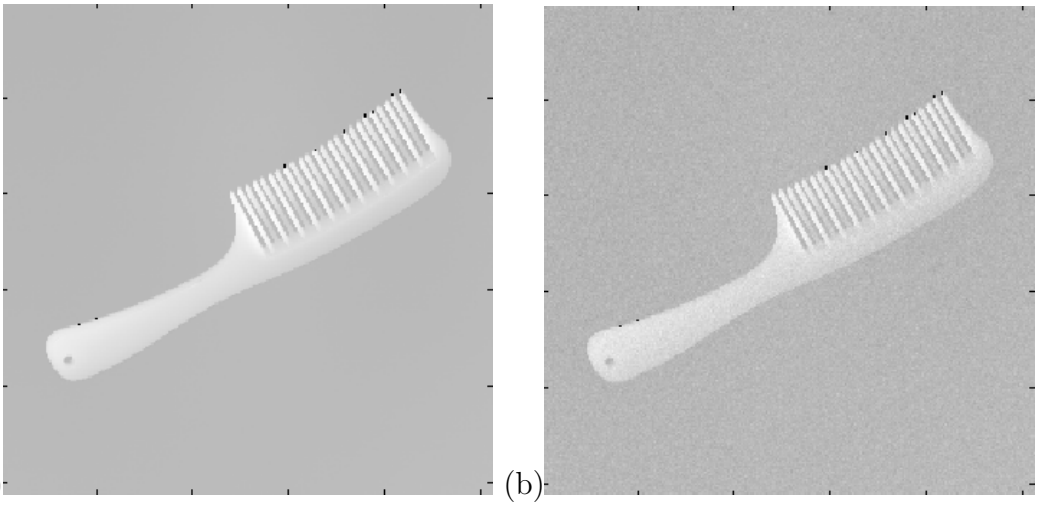


Figure 5.1: Example laser range measurement image used for testing ('comb'):
(a) Original fully sampled range measurements, (b) Noisy range measurements (standard deviation = 2%)

by computing the signal-to-noise ratio (SNR) of the reconstructed image for a wide range of compression rates, where the compression rate ρ is defined as one minus the ratio between the number of sampling locations measured and the total number of sampling locations. Therefore, the higher the compression rate achieved, the fewer the number of range measurements made. For illustrative purposes, the SNR was measured for images reconstructed across the range of 0% - 90% compression rates.

From the SNR vs. compression rate plot shown in Fig. 5.5(a), the SNR achieved using the conventional CS approach decreases almost linearly between ρ=0% and ρ=90%. On the other hand, the SNR achieved using the proposed saliency-guided sparse measurements model remains largely consistent between ρ=0% and ρ=70%, only decreasing noticeably from ρ=70% and ρ=90% due to insufficient salient samples at such a high compression rate. Furthermore, the SNR achieved using the proposed model is significantly higher than the one achieved using the conventional CS approach for all compression rates. Alternatively, the proposed model can achieve the same SNR value at a compression rate of ρ=70% as that achieved by the conventional CS approach at a significantly lower compression rate of ρ=3%.

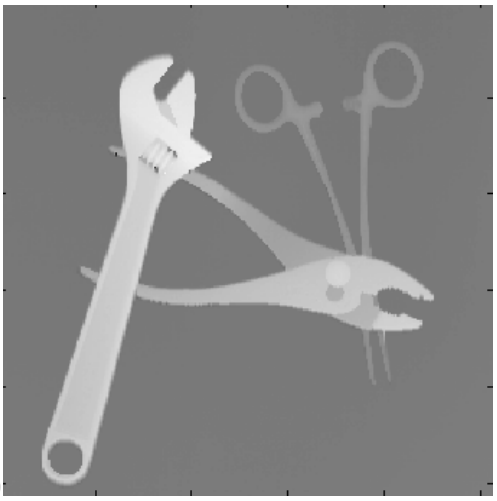
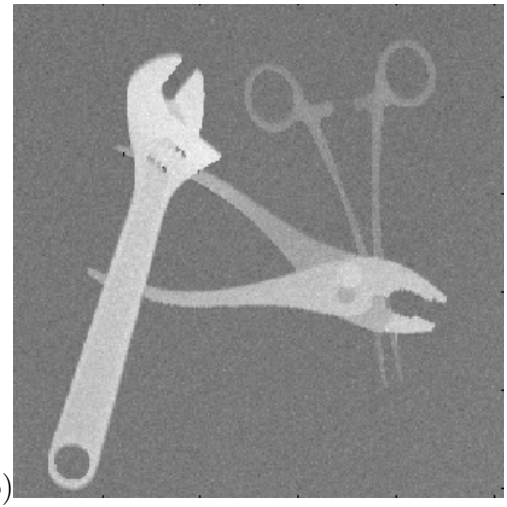
(a)  (b) 

Figure 5.2: Example laser range measurement image used for testing ('tools'): (a) Original fully sampled range measurements, (b) Noisy range measurements (standard deviation = 2%).

Experiment 2 - Noisy Images Reconstruction

In the second experiment, the reconstruction performance was evaluated in the situation where the measurements made were contaminated by noise. For testing purposes, the SNR was measured for images reconstructed under two different noise levels (additive white Gaussian noise with standard deviations of 2% and 3% of the dynamic range) across the range of 0% - 90% compression rates.

From the SNR vs. compression rate plots shown in Fig. 5.5(b) (Gaussian noise with standard deviation of 2%) and Fig. 5.5(c) (Gaussian noise with standard deviation of 3%), as with the first experiment, the conventional CS approach achieves SNR that is significantly lower than the one achieved using the proposed model. For example, at the same compression rate (72% for example), the saliency-guided model outperforms the conventional CS approach by 18 dB (32 dB vs. 14 dB) for Gaussian noise with a standard deviation of 2%. Similarly, in the case of Gaussian noise with a standard deviation of 3%, the saliency-guided model outperforms the conventional CS approach by 14 dB under the same compression rate scenario.

Finally, a comparison of the proposed model's reconstruction performance at different noise levels is shown in Fig. 5.5(d). The overall SNR is above 28 dB for all noise levels for compression rates from 0% to 70%, thus illustrating the high level of reconstruction perfor-

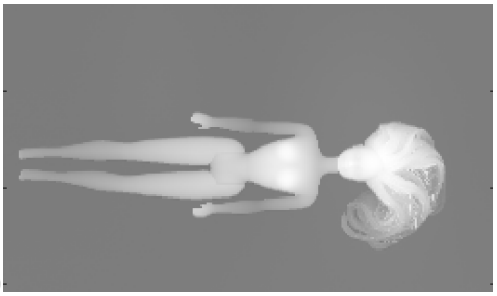
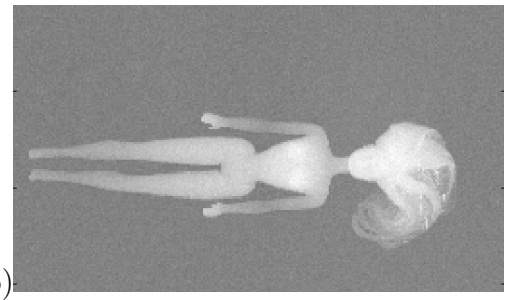
(a)  (b) 

Figure 5.3: Example laser range measurement image used for testing ('doll'): (a) Original fully sampled range measurements, (b) Noisy range measurements (standard deviation = 2%).

mance achieved using the proposed saliency-guided sparse measurements model. Fig. 5.6 provides complementary perspective for comparing the two models performance at different noise levels and at 50% compression rate. The saliency-guided model outperforms CS consistently through the entire selected noise range of 0-5%. While that at noise levels above 3%, saliency-guided performance is decreasing faster than in lower noise levels.

Experiment 3 - Reconstruction Repeatability Tests

Since there are stochastic processes embedded in the proposed model, reconstruction repeatability tests were performed to validate that consistent performance can be achieved. This is achieved by repeating the proposed approach 35 times for the noisy case with Gaussian noise of 2% standard deviation at 80% compression rate. As shown in Fig. 5.7, the SNR of the reconstructed data is very consistent with an average SNR of 27.8 dB and variance of 3.3 dB.

Experiment 4 - Reconstruction Examples

In order to visualize the reconstruction performance of the proposed model, several examples of reconstructed data are presented in Fig. 5.8 - Fig. 5.15.

Fig. 5.8 demonstrate reconstruction at 70% compression rate and noise-free measurements. The image produced using the proposed model contains significantly more important details than the conventional CS approach. This is most evident in the lack of threads

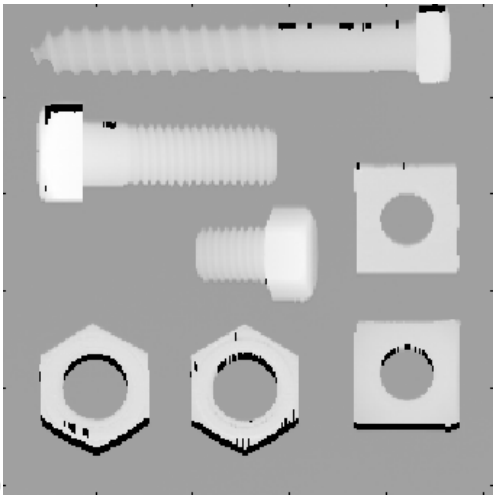
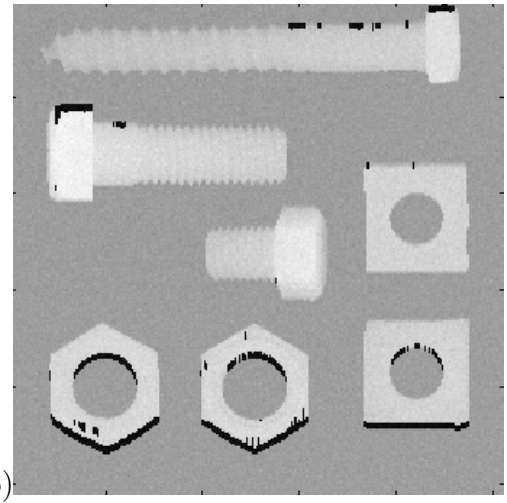
(a)  (b) 

Figure 5.4: Example laser range measurement image used for testing ('nuts and bolts'): (a) Original fully sampled range measurements, (b) Noisy range measurements (standard deviation = 2%).

on the bolts in the image constructed using the conventional CS approach, which is well captured in the image produced using the proposed model.

Fig. 5.9 and Fig. 5.10 are two images ('tools' and 'comb') at 70% compression rate and noisy measurements with noise standard deviation of 2%, while Fig. 5.11 shows the 'comb' image at the same compression rate but with noise standard deviation of 3%. Even under noise, the proposed model produces images that capture significantly more detail than the one produced by the conventional CS approach. This is most evident in the scissors in the 'tools' image, where the scissor appears disjointed in the image produced using the conventional CS approach while maintaining its structure in the image constructed using the proposed model. Similarly, in the 'comb' image, the teeth of the comb are well captured using the proposed model when compared to the conventional CS approach.

Fig. 5.13 and Fig. 5.14 shows the reconstructed images from noisy measurements with noise standard deviation of 2% at 75% and 80% compression rates, respectively. Even under the higher compression rates, the proposed model produces images that capture significantly more detail than the one produced by the conventional CS approach.

Fig. 5.15 shows the sampling distribution across the 'comb' image for the conventional CS approach and the proposed approach respectively. White pixels represent locations that are not sampled (80% of the pixels in this case). In contrast to the conventional CS approach where samples are distributed uniformly, the proposed saliency-guided approach

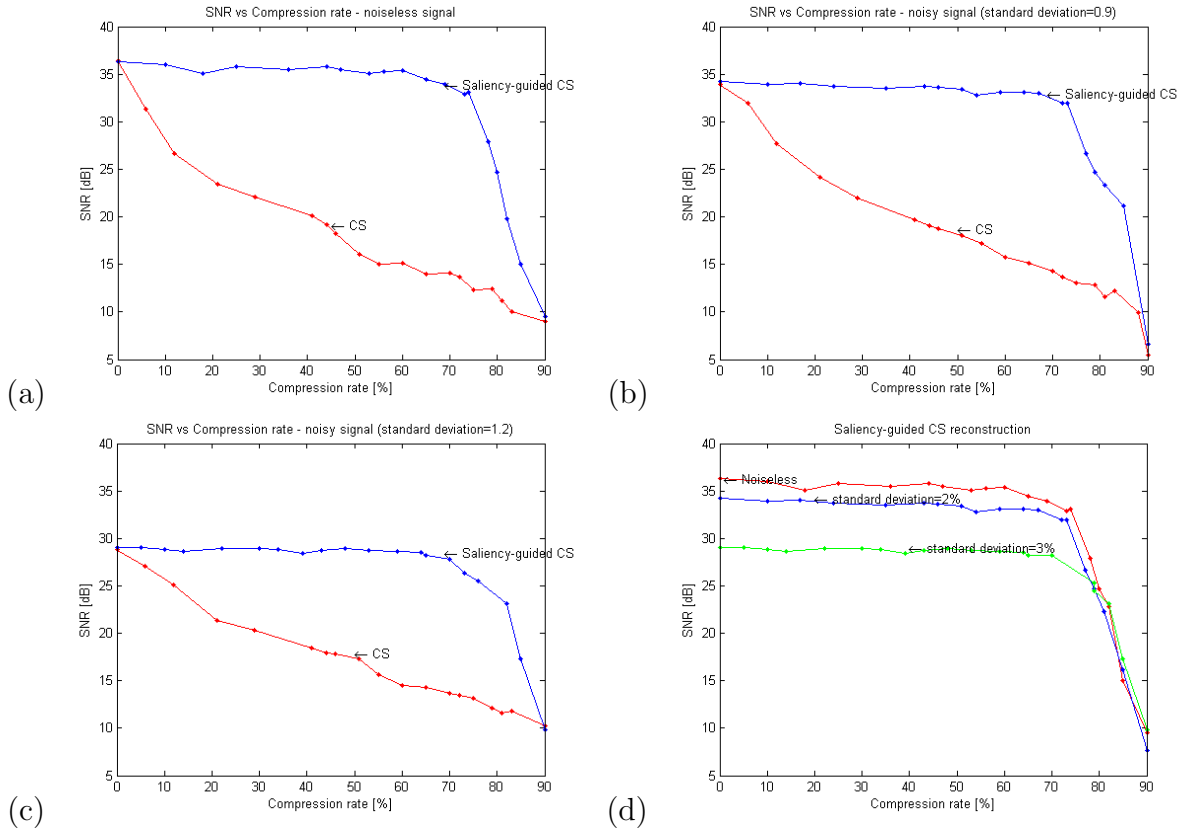


Figure 5.5: Noiseless image reconstruction at different compression rate ('nuts and bolts'):
(a) SNR vs. compression rate for noiseless range measurements,
(b) SNR vs. compression rate for range measurements contaminated by Gaussian noise with standard deviation of 2%
(c) SNR vs. compression rate for range measurements contaminated by Gaussian noise with standard deviation of 3%
(d) SNR vs. compression rate comparisons for different noise levels

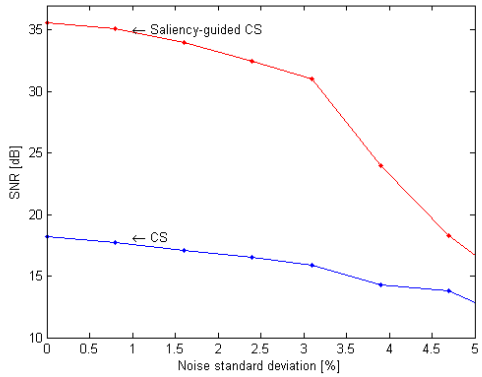


Figure 5.6: SNR vs. noise level at 50% compression rate

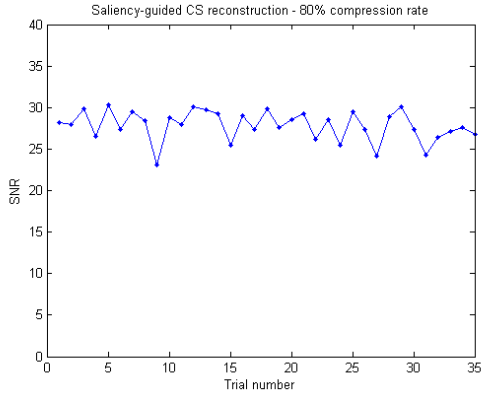


Figure 5.7: Repeatability test for saliency-guided reconstruction over 35 trials at 80% compression rate and Gaussian noise with standard deviation 2%. The SNR of the reconstructed data is very consistent, with an average SNR of 27.8 dB and variance of 3.3 dB.

distributes the samples according to the object of interest. Regions of high saliency are sampled densely while all other regions are sampled sparsely. As such, the proposed approach have low sampling frequency in the background (where few samples are needed for accurate reconstruction) and high sampling frequency around the object of interest (where more samples are needed for accurate reconstruction).

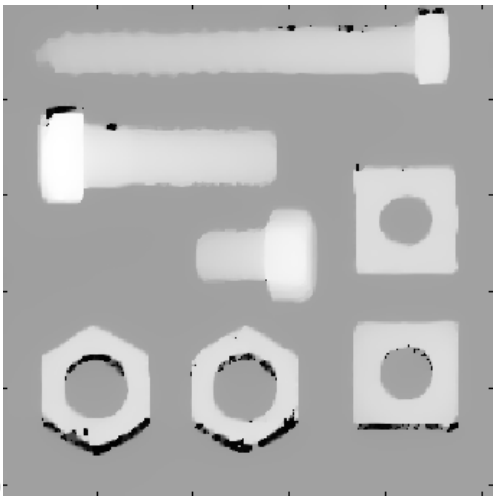
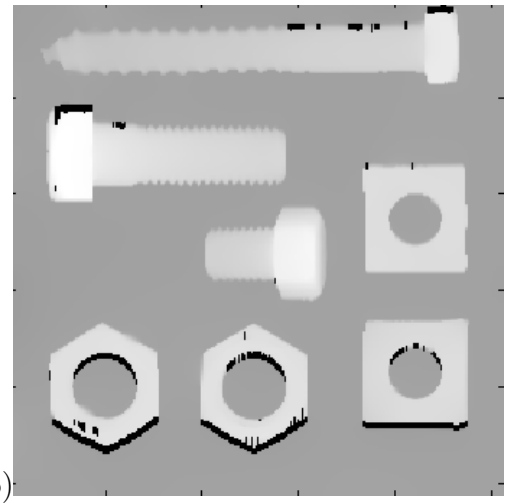
(a)  (b) 

Figure 5.8: Noiseless image reconstruction at 70% compression rate ('nuts and bolts'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction

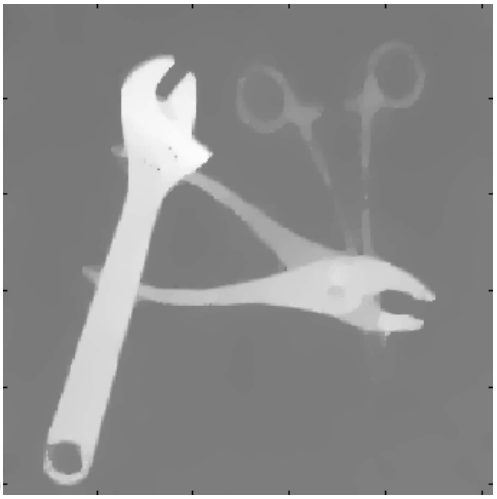
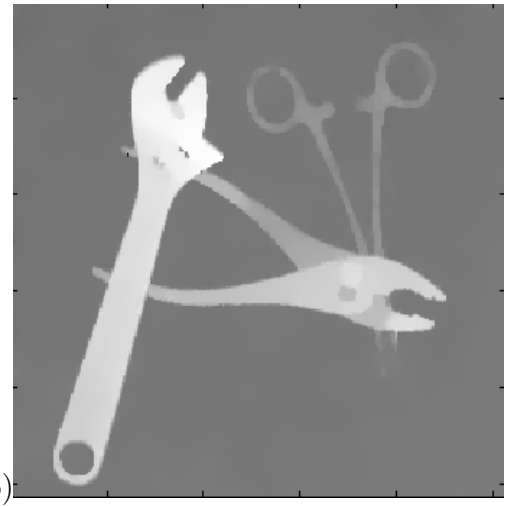
(a)  (b) 

Figure 5.9: Noisy image (standard deviation of 2%) reconstruction at 70% compression rate ('tools'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction

(a)  (b) 

Figure 5.10: Noisy image (standard deviation of 2%) reconstruction at 70% compression rate ('comb'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction

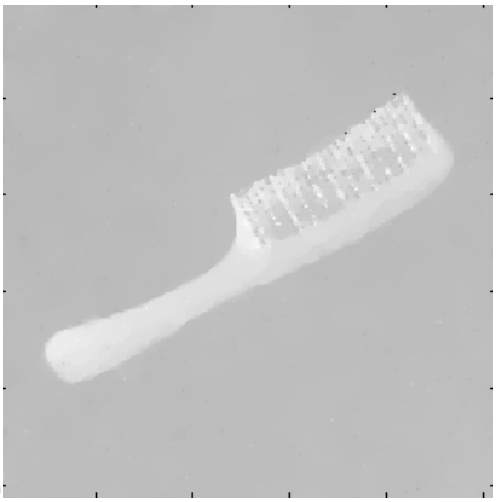
(a)  (b) 

Figure 5.11: Noisy image (standard deviation of 3%) reconstruction at 70% compression rate ('comb'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction

(a)  (b) 

Figure 5.12: Noisy image (standard deviation of 2%) reconstruction at 60% compression rate ('comb'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction

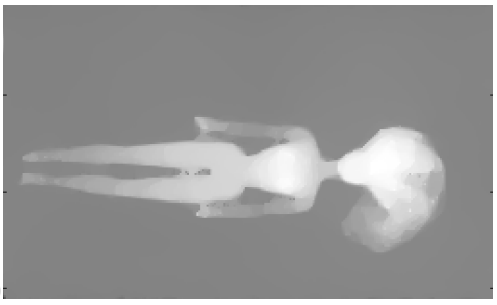
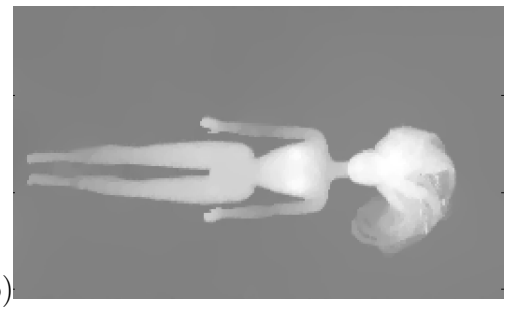
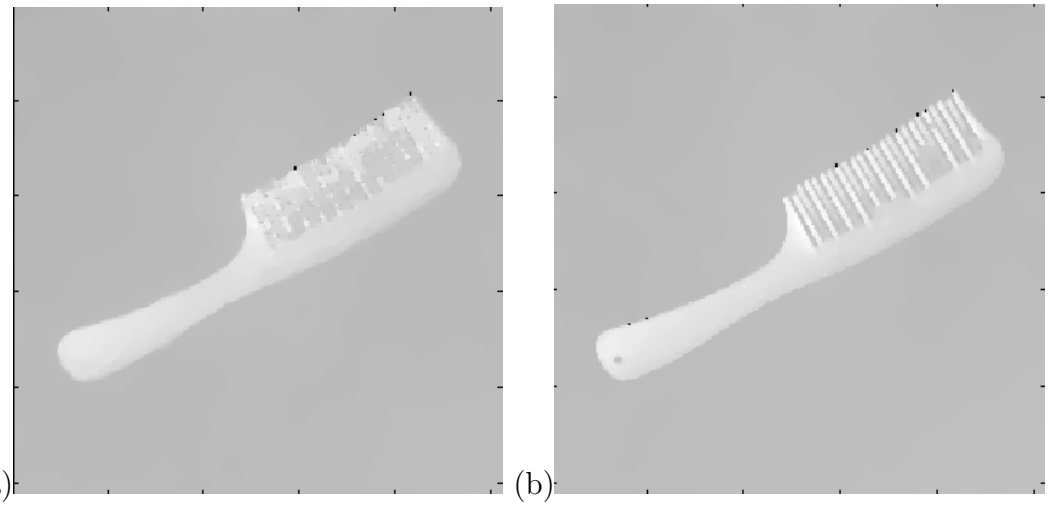
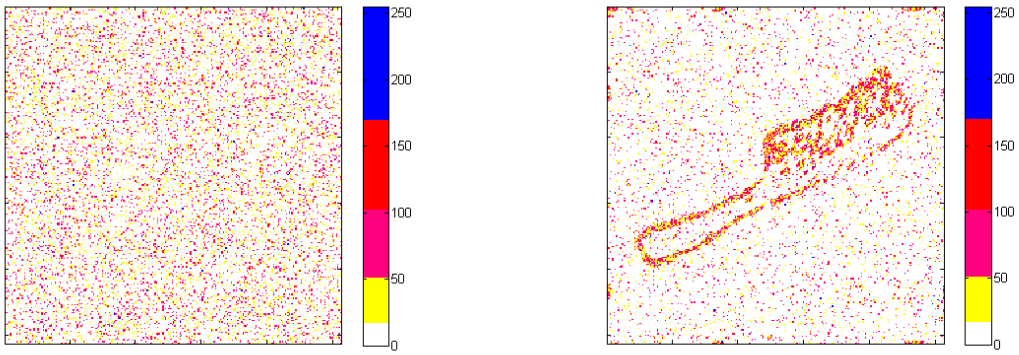
(a)  (b) 

Figure 5.13: Noisy image (standard deviation of 2%) reconstruction at 75% compression rate ('doll'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction



(a) (b)

Figure 5.14: Noisy image (standard deviation of 2%) reconstruction at 80% compression rate ('comb'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction



(a) (b)

Figure 5.15: Samples distribution at 80% compression rate:
(a) uniformly under-sampled, (b) Saliency-guided under-sampled

5.3.2 Robotic Vision

Robots are evolving from being remotely controlled to become fully autonomous. An example of a fully autonomous robot can be the mars exploration rovers [99] developed for navigating, sampling, and mapping regions of planet Mars. Robot localization is a key factor allowing autonomous navigation capabilities [36]. Autonomous robots can maintain an ongoing estimate of their location and orientation with respect to its environment through a localization process. Autonomous robots create maps of its environment, find current location, avoid obstacles, and navigate according to a planned path toward the target [133]. Measurement sensors are an important core part of robot localization [17, 50], of which laser range finders and cameras are commonly used as sensing devices [37]. Robotic localization is a multidimensional problem that can be complex to solve in real-time due to the amount of data that needs to be processed [137]. Even though many aspects of high dimensionality have been investigated [65], the large amount of data that needs to be acquired and processed constantly during robotic navigation makes this task challenging, especially in outdoor land-based [66], aerial [77], and underwater [101] environments. Natural objects in outdoor environments such as trees, residential construction, and even indoor objects such as furniture and staircases are complex and vary greatly in size, shape, orientation, and texture, and therefore require higher range measurement resolution. However, higher laser range measurement resolution can significantly effect acquisition and processing time.

The robotic vision tested images are more complected and contains much more fine details compared to the 3D laser measurements (5.3.1) tested scenes. Therefore, the multi-scale saliency-guided approach (5.3) in spatial domain was implemented. For the experiments at this stage, the mapping function G was selected to be logarithmic function (5.3).

In the first experiment, reconstruction performance was evaluated at different compression rates. In the second experiment, reconstruction performance was evaluated the situations where the measurements were contaminated by different levels of noise.

Experimental Setup

The experiments were performed with indoor and outdoor laser range data sets from Brown University range image database [105]. The laser range data in the Brown database were acquired using the Riegl LMS-Z210 laser range-finding apparatus with a rotating mirror. Each range image consists of 444×1440 measurements (each measurement based on time of flight of laser beam) with an angular separation of 0.18 degrees, resulting in a field of

view of 80 degrees vertically and 259 degrees horizontally. The operational range is around 2-200m.

Examples of fully sampled data (where each sampling location is measured), along with corresponding noise contaminated versions used for the second experiment, are shown in Fig. 5.16 - Fig. 5.18.

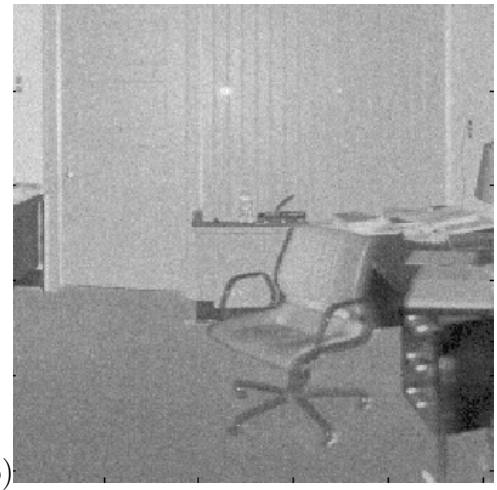
(a)  (b) 

Figure 5.16: Example laser range measurement image used for testing ('office'):
(a) Original fully sampled range measurements, (b) Noisy range measurements (standard deviation = 3%).

Experiment 1 - Reconstruction at Different Compression Rates

In the first set of experiments, a parametric analysis is performed to evaluate the reconstruction performance of the proposed method at different compression rates. This is achieved by computing the signal-to-noise ratio (SNR) of the reconstructed image for a wide range of compression rates. Within the context of CS, the compression rate ρ is defined as one minus the ratio between the number of measured sampling locations and the total number of sampling locations. Hence, a higher compression rate means that fewer range measurements are made. The SNR was measured for noiseless images reconstructed via the tested methods for the compression rate range of 0% - 90%.

Based on the SNR vs. compression rate plot shown in Fig. 5.19, the SNR achieved using the conventional CS approach is significantly lower than the one achieved using the proposed saliency-guided sparse measurements model across all compression rates.

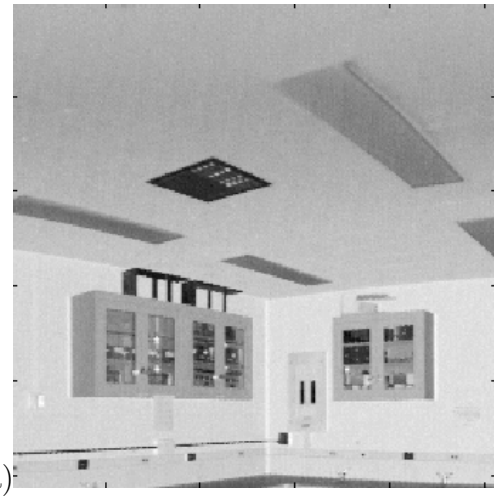
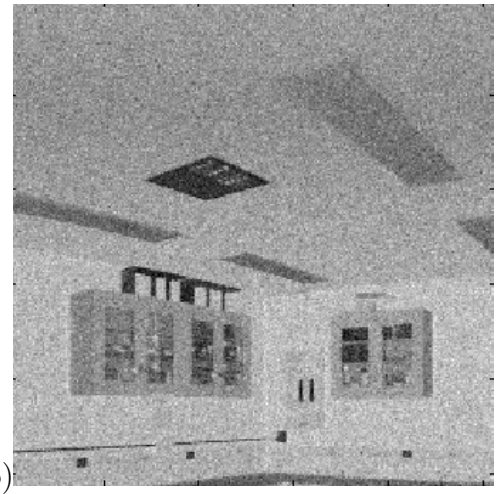
(a)  (b) 

Figure 5.17: Example laser range measurement image used for testing ('cabinet'):
(a) Original fully sampled range measurements, (b) Noisy range measurements (standard deviation = 4.7%).

Another way of looking at it is that the proposed model can achieve the same SNR value at higher compression rates than the value achieved by the conventional CS approach at a significantly lower compression rates.

Experiment 2 - Noisy Images Reconstruction

In the second set of experiments, the reconstruction performance of the proposed model is evaluated at different noise levels (additive Gaussian noise of 1.5% and 15.5% of the dynamic range). Fig. 5.20 shows that the proposed multi-scale saliency guided model with continuous saliency function $\Gamma(r, c)$ (5.6) outperforms both the conventional CS approach and the recent binary saliency-guided model (5.11) proposed in [130] at all noise levels. This illustrates the noise robustness of the proposed model, which is important in real-world scenarios where the measurements made are noisy.

Reconstruction Examples

For visualization purposes, several examples of indoor and outdoor reconstructed laser range data are shown in Fig. 5.21 - Fig. 5.23.

(a)  (b) 

Figure 5.18: Example laser range measurement image used for testing ('house'):
(a) Original fully sampled range measurements, (b) Noisy range measurements (standard deviation = 4.7%).

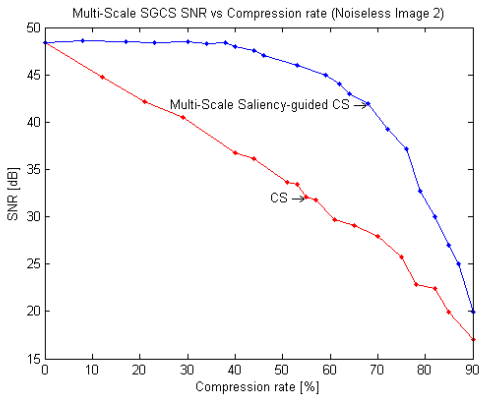


Figure 5.19: SNR vs. compression rate for noiseless range measurement. SNR achieved using the conventional CS approach is significantly lower than the one achieved using the proposed saliency-guided sparse measurements model across the entire compression rates range. Alternately, the saliency-guided model achieves the same SNR value at much higher compression rates compared to the conventional CS approach.

Fig. 5.21 demonstrates noiseless measurements reconstruction at 70% compression rate. The image produced using the proposed model contains significantly more important details than the conventional CS approach. This is most evident in the lack of fine details in the

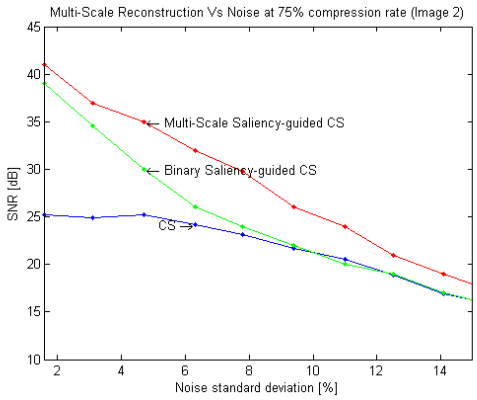


Figure 5.20: SNR vs. noise level at 75% compression rate for the multi-scale saliency guided model with continuous saliency function $\Gamma(r, c)$ (5.2), outperforms both the conventional CS approach and the binary saliency-guided model (5.11) at all noise levels.

tree as well as windows and roof frames in the image constructed using the conventional CS approach, which is well captured in the image produced using the proposed model.

Fig. 5.22 is a reconstructed images ('cabinet') at 70% compression rate and with noise standard deviation of 5%, while Fig. 5.23 shows the 'office' reconstructed image at 75% compression rate and with noise standard deviation of 3% . Even using noisy measurements, the proposed model produces images that capture significantly more detail than the one produced by the conventional CS approach. This is most evident in the 'office' image, where objects on the desk and drawer handles are largely missing in the image produced using the conventional CS approach compared to the proposed model.

5.3.3 Fluorescence Microscopy

Fluorescence is broadly used in many biological microscopy applications such as single molecular studies of individual proteins and living cells, in-vivo tracking of targets using fluorescent labeling, and molecular associations in live cells [108]. One key advantage of fluorescence microscopy is its inherently greater optical sensitivity and dynamic range in comparison to other methods based on optical density changes or chemiluminescent emission [108]. Fluorescence microscopy approaches are usually categorized by the way in which excitation light is delivered to the sample: i) wide-field and structured illumination, ii) surface illumination, and iii) scanning illumination. One of common most approaches in fluorescence microscopy is scanning confocal microscopy, where the sample is scanned by

(a)  (b) 

Figure 5.21: Noiseless image reconstruction at 70% compression rate ('house'): (a) CS reconstruction, (b) Saliency-guided CS reconstruction

a laser in two dimensions to reconstruct the image [106]. One drawback to confocal scan microscopy is its inherent slow acquisition speed since the image is been acquired pixel-by-pixel. Furthermore, the quantum efficiencies of common confocal microscopy CCD detectors is lower [108], resulting in lower signal to noise ratio (SNR). Achieving high SNR is of particular interest in biological fluorescence microscopy since the reflective index of cells is similar to water which forms the background. Therefore, the intensity difference between the regions of interest and the background is low and sensitivity to noise is high.

To deal with the issue of SNR, many physical noise reduction methods have been developed for fluorescence microscopy. Such methods include: reducing unwanted room background light, reducing unwanted fluorescence emission through a diaphragm that filters reflections or scattering light path [74, 157], time-gated imaging [33, 34, 35] using short laser pulses for sample illumination, as well as minimizing illumination time. Another way to improve SNR is maximizing signal cleanliness through specialized lenses [110].

Another issue that can affect SNR is the chemical destruction phenomenon known as photobleaching, where higher illumination levels beyond the fluorescence saturation point do not contribute to signal amplitude but continue to increase background scattering leading to increased noise levels and SNR degradation [94, 108, 138]. Photobleaching can occur at timescales of milliseconds to minutes. Common strategies to decrease photobleaching effects can be by reducing illumination levels, applying specialized filters and lenses, and using less sensitive fluorophore [103, 108, 114]. The combination of very weak fluorescence

(a)  (b) 

Figure 5.22: Noisy image (standard deviation of 5%) reconstruction at 70% compression rate ('cabinet'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction

signal (10^{-6} of the level of excitation light that produced it [106]) and photobleaching effects that limits the excitation light levels provides low SNR. Those limitations can lead to laser confocal microscopy that needs to measure 10-20 photons from brightest pixels in the image and as low as zero or one photon from the background [106]. Beyond physical methods, computational methods have been attempted reduce noise levels through image filters such as anisotropic diffusion and wavelet thresholding [98, 107, 123].

One promising approach for improving acquisition speed while maintaining low system complexity and high reconstruction quality for fluorescence microscopy is the concept of compressive fluorescence microscopy (CFM) [91, 92, 141, 153]. CFM makes use of compressive sensing (CS) theory, which allows for greatly reduced fluorescence microscopy acquisition times through the use of sparse measurements (samples). The theory of compressive sensing (CS) provides an unambiguous proof that discrete signals, which represent a sparse representation in the domain of a suitable linear transform, can be accurately recovered from their sub-Nyquist measurements [8, 18, 19, 42, 51]. Such reconstructions can be performed through solution of a convex optimization problem, which maximizes the sparseness of the signal representation coefficients subject to a set of measurement constraints. In such cases, the measurement constraints are derived from a signal/image acquisition model. This model normally describes the measurements as a (noise-contaminated) sequence of inner products between the quantity of interest and the elements of a sampling

(a)  (b) 

Figure 5.23: Noisy image (standard deviation of 3%) reconstruction at 75% compression rate ('office'):
(a) CS reconstruction, (b) Saliency-guided CS reconstruction

basis. Wu et al. [153] demonstrated the applicability of CFM in fast optical-sectioning imaging. Studer et al. [141] demonstrated the use of CFM for fast hyperspectral imaging. In addition to tackling the issue of acquisition speed, Marim et al.[91, 92] recently proposed a CFM system for improving SNR by fusing multiple CFM reconstructions.

To implement the ensemble of saliency-guided reconstructions for data fusion, an additional stage is added to the saliency-guided approach in spatial domain (5.3) implementation. This additional stage is the ensemble expectation merging stage for fusing multiple reconstructions and improving fluorescence microscopy SNR. At this stage, the final reconstruction $\widetilde{f}$ is computed via ensemble expectation merging [92]:

$$\widetilde{f} = E\left\{\bar{f}_t\right\}, t = t_1, t_2, \cdots t_n \tag{5.13}$$

where $E\{.\}$ denotes average, f_t is the image reconstruction at $t, t = t_1, t_2, \cdots t_n$ and n is the ensemble size. Similarly to the 3D laser measurements realization, a simplified version of saliency-guided approach in spatial domain was implemented (5.11).

In the first experiment, reconstruction performance was evaluated in the situation where the measurements made were contaminated by different levels of synthetic noise. In the second experiment, reconstruction performance was evaluated at different compression rates. In the third experiment, reconstruction performance of real noisy fluorescence microscopy data was evaluated.

Experimental Setup

The experiments were performed with existing fluorescence microscopy data sets provided by [115]. Five data sets where used for the synthetic noise experiments: YRC PIR ID: 5, 12, 27, 62 and 64. The data sets are largely noiseless, and have the following imaging settings: pixel size: 0.12758 μm $\times$ 0.12758 μm, objective: 100$\times$ and image size 512$\times$512. In addition, three noisy data sets were used for the real noise experiments: YRC PIR ID: 3499, 5352 and 8565. The data sets contain time series of at least 10 images, with the following imaging settings: pixel size: 0.12758 μm $\times$ 0.12758 μm, objective: 100$\times$ and image size 512$\times$512.

Examples of fully sampled data (where each sampling location is measured), along with corresponding noise contaminated versions used for the first experiment, are shown in Fig. 5.24 and Fig. 5.25. Fully sampled real noisy images are shown in Fig. 5.26.

Experiment 1 - Noise Sensitivity Tests

In the first experiment, a comparison of the proposed Saliency-guided sparse reconstruction ensemble fusion (SSREF) model's reconstruction performance vs. noise levels (additive white Gaussian noise) with different ensembles is shown in Fig. 5.27. Up to 4% noise standard deviation there is insignificant gains from increasing ensemble size. From the other end, at higher noise level, larger ensemble sizes lead to improved SNR reconstruction, with a 4dB difference between ensemble sizes of 4 and 7, and 3dB between ensemble sizes of 7 to 10. Since the ensemble expectation merging oparation has some level of bluring effect, lower ensemble size produces slidly higher SNR at very low noise levels. It might be interesting to investigate in the future an improved ensemble expectation merging method.

At this point, an ensemble size of 10 is used for further experiments. A comparison between the fusion of compressive fluorescence microscopy (FCFM) [91, 92] approach and the proposed SSREF model's reconstruction performance at different noise levels is shown in Fig. 5.28. The compression rate is set to 75%. The proposed model outperforms FCFM significantly and consistently through the entire selected noise range of 1.5-10%. For example, a 11.8 dB difference is observed at 5% noise level and a 4 dB difference is observed at 10% noise. Fluorescence microscopy images contains structures and definite regions of interest, especially at the edges of of the objects in the images. Therefore, adjusting the sampling probability according to underlying data provides improved SNR performance compared to the FCFM, where the entire scene is considered equally important. At very high noise level, the saliency map is not performing as well as at lower noise levels to

identify regions of interest. Therefore, at higher noise levels the SSREF performance is getting closer to the FCFM reconstruction performance.

Experiment 2 - Compression Rate Sensitivity Tests

In the second experiment, the reconstruction performance was evaluated via a parametric analysis by computing the signal-to-noise ratio (SNR) of the reconstructed fluorescence microscopy image for a wide range of compression rates, where the compression rate ρ is defined as one minus the ratio between the number of sampling locations measured and the total number of sampling locations. Therefore, the higher the compression rate achieved, the fewer the number of measurements made. For illustrative purposes, the SNR was measured for fluorescence microscopy image, contaminated with 3% standard deviation noise, reconstructed via ensembles of 10 reconstructions across the range of 0% - 80% compression rates.

From the SNR vs. compression rate plots shown in Fig. 5.29 (Gaussian noise with standard deviation of 3%), the FCFM approach achieves SNR that is significantly lower than the one achieved using the proposed SSREF model. For example, at the same compression rate (60% for example), the SSREF model outperforms the FCFM approach by 10.7 dB (32.5 dB vs. 21.9 dB) for Gaussian noise with a standard deviation of 3%. At 0% compression rate, the FCFM and SSREF achives similar performance, since 100% of samples are beed used. At very high compression rate, the SSREF is not performing as well as at lower rates since the initial learning phase do not have sufficient samples to produce useful information to identify regions of interest. At the mid-compression rate, between 30% and 70%, SSREF achives much better performance compared to FCFM.

To visualize the reconstruction performance of the proposed SSREF model, Fig. 5.30 and Fig. 5.31 demonstrate reconstruction of FCFM as well as the proposed SSREF model at different levels of synthetic noise. The fluorescence microscopy images produced using the proposed SSREF model contain significantly more important details than the FCFM approach. This is most evident in the weak structure definition in the fluorescence microscopy images constructed using FCFM, which is well captured in the fluorescence microscopy images produced using the SSREF model. In addition, the proposed model provides improved noise suppression in the background regions more efficiently than FCFM. SSREF preserves edges much better in both noise level cases (3% and 5.5%). SSREF reconstruct much better objects of interest since the sampling distribution is saliency guided while the conventional CS sample the entire scene by uniform distribution.

Experiment 3 - Noisy Fluorescence Microscopy Reconstruction Tests

Reconstruction Examples

In the third set of tests, real noisy fluorescence microscopy images were used to compare reconstruction performance of FCFM vs. the proposed SSREF model Fig. 5.32 - Fig. 5.34. The proposed SSREF model outperforms the FCFM approach for real noisy images in terms of image quality namely better structure definition and reduced noise. SSREF preserve edges of noisy fluorescence microscopy much better than FCFM. For example, elements in cluster Fig. 5.32 can be identified as separate object at the image produced by SSREF, while conventional CS reconstruction produced clusters with blurred and noisy edges. It is most evident that the saliency guided sampling distribution reconstruct regions of interest such as edges, much better compared to conventional CS approach which sample the entire scene with uniform distribution.

5.3.4 Learning Nonparametric Sampling Function for Compressive Fluorescence Microscopy

The indirect learning model in the spatial domain (5.2) defines the continuous saliency function $\Gamma(r, c)$ (5.6) where $G(\theta_i^{\mu})$ returns a sparsity level value between 0 and 1 and is a function of the cluster θ_i index sorted by the cluster mean, which is the average saliency of region i. $G(\theta_i^{\mu})$ can be selected to be a function that emphasizes high saliency more than lower salient levels. In the realization of saliency-guided modeling approach in spatial domain (5.3), $G(\theta_i^{\mu})$ can be selected to be the logarithmic function of cluster θ_i. This logarithmic function emphasizes high saliency more than lower salient levels. The logarithmic function selection for $G(\theta_i^{\mu})$ was shown to improve reconstruction performance, but is this arbitrary selection the most efficient choice? In addition this choice is not necessarily an optimized choice for other type of images.
Construct a sampling pattern based on the nonparametric modeling approach can be extended for optimizing the model by mapping the extracted features to an optimized pdf through learning. The function $G(\theta_i^{\mu})$ can be optimized through learning for maximizing reconstruction performance. In the indirect learning model in the spatial domain (5.2) the saliency map is clustered into R clusters θ_i, $i = \{1, \ldots, R\}$ with common saliency properties. In order for optimizing the mapping function G for learning the pdf Γ, Markov chain Monte Carlo (MCMC) optimization method (3.14) is selected and implemented through Metropolis-Hastings sampling algorithm (3.1). By solving this optimization (3.14) one can maximize the CS reconstruction performance.

Let define the mapping function $G(\overline{z_i})$ at iteration i^{th} where the argument is the Markov chain sequence $\overline{z_i}$. Let $\overline{z_1}$ the initial proposed sequence with initial arbitrary logarithmic function that showed good results in initial experiments. At step k, a sequence $\overline{z}^*$ is proposed to be assigned to sequence $\overline{z}_k$ for the mapping function $G(\overline{z}_k)$ through random walk chain where $\overline{z}^* = \overline{z}_{k-1} + \overline{w}$ where $\overline{w}$ come from uniform distribution. An acceptance parameter α (3.15) with appropriate function Δ (3.15) for this application, is calculated (5.14) for each iteration k :

$$\alpha(\overline{z}^*, \overline{z}_{k-1}) = min\{1, \lambda \cdot \exp(SNR_{\widetilde{f}(\overline{z}^*)} - SNR_{\widetilde{f}(\overline{z}_{k-1})})\} \tag{5.14}$$

where $SNR_{\widetilde{f}(\overline{z}^*)}$ is the signal to noise ratio of a reconstructed image $\widetilde{f}$, while the sequence $\overline{z}^*$ is been used for the mapping function G and λ is the regularization parameter. The implementation uses random number u, uniform distributed in the range $(0, 1)$, to implement the accept/reject method. This random number is compared with the probability criterion $\alpha(\overline{z}^*, \overline{z}_{k-1})$. In other words, if $u < \alpha(\overline{z}^*, \overline{z}_{k-1})$ it is accepted and $\overline{z}_k = \overline{z}^*$ is assigned, otherwise it is rejected and the previous variable in the sequence $\overline{z}_k = \overline{z}_{k-1}$ is assigned. This iterative process is used to find an optimized mapping function G. Since the number of clusters R might be different for every image or noise level, a continues function G is estimated that is used an outcome of the learning process.

Experimental Results and Discussions

The sampling function learning process was implemented for fluorescence microscopy images as well as indoors and outdoors laser range images used for learning an optimized function G which is been used as the sampling probability function. Examples of learned an optimized sampling probability functions are presented at (Fig-5.35 and Fig-5.36) as well as static log function (Fig-5.37) used as reference.

Reconstruction performance is evaluated for saliency-guided sparse reconstruction through an optimized function G that was learned in an initial process. The learned process and experiments used different images from the same type. In other words, one fluorescence microscopy image was used for learning and the learned function G is used for other fluorescence microscopy images. Similarly for laser range images, one image was used for learning an optimized G function which is then used for other laser range images. As reference, the reconstruction based on static log function is presented. This is achieved by computing the signal-to-noise ratio (SNR) of the reconstructed image for a wide range of compression rates. Within the context of CS, the compression rate ρ is defined as one minus the ratio between the number of measured sampling locations and the total number

of sampling locations. Hence, a higher compression rate means that fewer measurements are made. The SNR was measured for noiseless images and compression rate range of 40% - 80%. Based on the SNR vs. compression rate plot shown in Fig-5.38, the SNR achieved based on the learned function approach is higher than the one achieved using static log function across all compression rates, average of 6 dB for fluorescence microscopy image ID1-1 5.38 (a), average of 4.4 dB for fluorescence microscopy image ID1-4 5.38 (b), average of 3.5 dB for range measurement indoors image ID2-7 5.38 (c) and average of 5.7 dB for range measurement outdoors image ID2-11 5.38(d).

A comparison between reconstruction with static log sampling probability function and the proposed reconstruction performance which is based on dynamically learned sampling probability function model at different noise levels is shown in Fig-5.39. The compression rate is set to 50%. It can be observed that the proposed learning model outperforms the static logarithmic function through the entire selected noise range. average of 3.1 dB for fluorescence microscopy image ID1-1 5.39 (a), average of 1.4 dB for fluorescence microscopy image ID1-4 5.39 (b), average of 1.9 dB for range measurement outdoors image ID2-7 5.39 (c) and average of 3.7 dB for range measurement outdoors image ID2-11 5.39 (d). It can be seen that at lower noise level the difference between the two approaches is larger compared to higher noise levels. It can be interesting to study in the future implementing filters for the saliency map in order to improve reconstruction performance at higher noise levels.

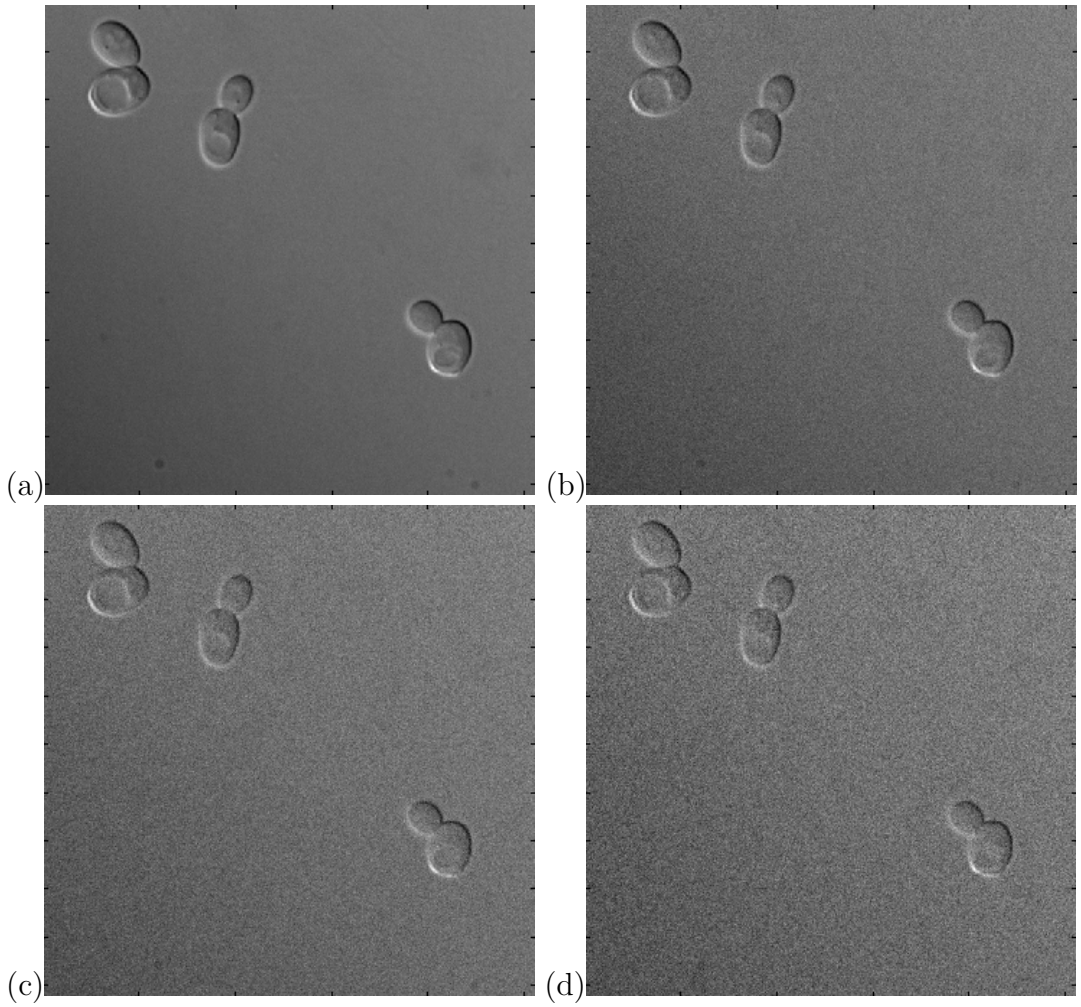


Figure 5.24: Example fluorescence microscopy image used for testing ('ID 12'):
(a) Fully sampled data, (b) Noisy measurements (standard deviation = 3%) (c) Noisy measurements (standard deviation = 5.5%), (d) Noisy measurements (standard deviation = 7%)

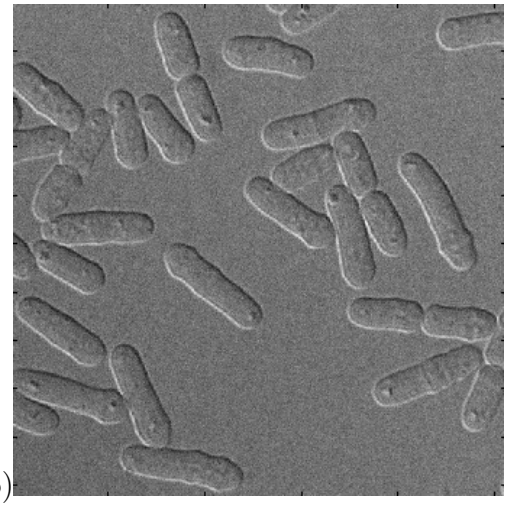
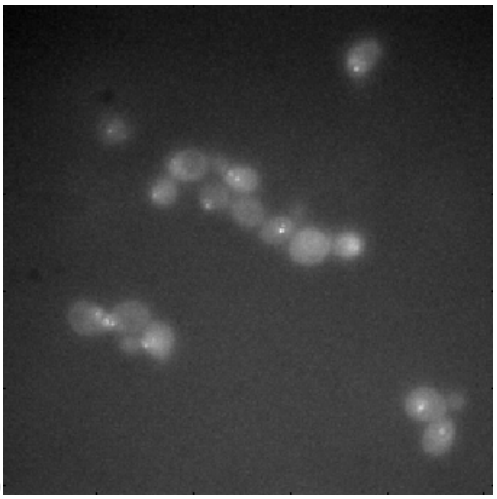
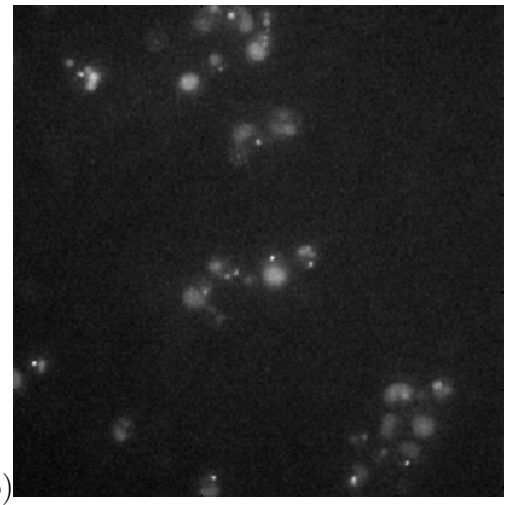
(a)  (b) 

Figure 5.25: Example fluorescence microscopy data used for testing ('ID 5'):
(a) Fully sampled data, (b) Noisy measurements (standard deviation = 5.5%)

(a)  (b) 

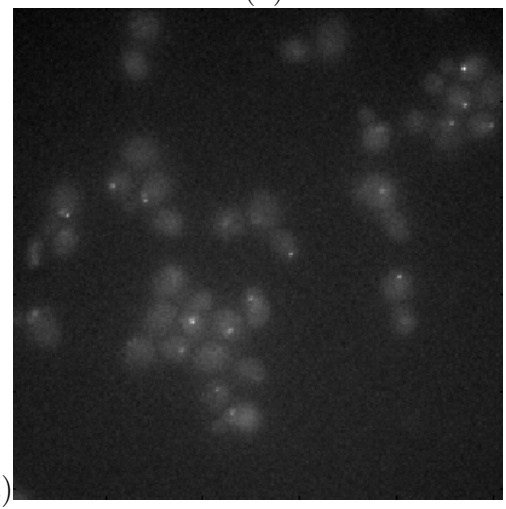
(c) 

Figure 5.26: Example fluorescence microscopy measurement image used for testing:
(a) Original fully sampled real noisy fluorescence microscopy measurements ('ID 8565'),
(b) Original fully sampled real noisy fluorescence microscopy measurements ('ID 3499'),
(c) Original fully sampled real noisy fluorescence microscopy measurements ('ID 5352')

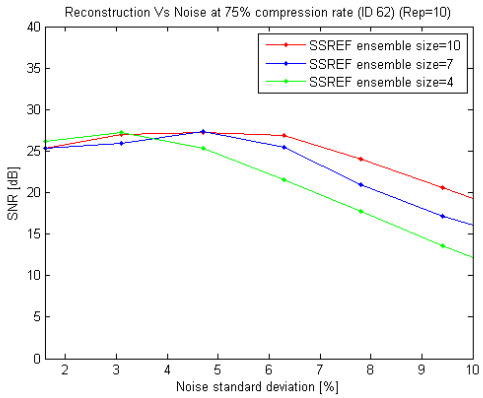


Figure 5.27: SNR vs. noise levels with different ensembles. At higher than 4% standard deviation noise, larger ensemble sizes lead to improved SNR reconstruction, with a 4dB difference between ensemble sizes of 4 and 7, and 3dB between ensemble sizes of 7 to 10.

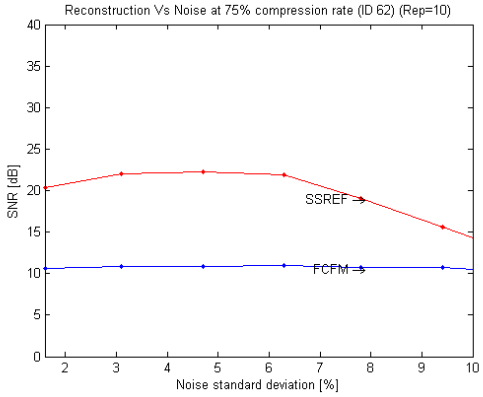


Figure 5.28: SNR vs. noise level at 75% compression rate and ensemble size of 10. The proposed SSREF model outperforms FCFM significantly and consistently through the entire selected noise range of 1.5-10%. For example, a 11.8 dB difference is observed at 5% noise level and a 4 dB difference is observed at 10% noise

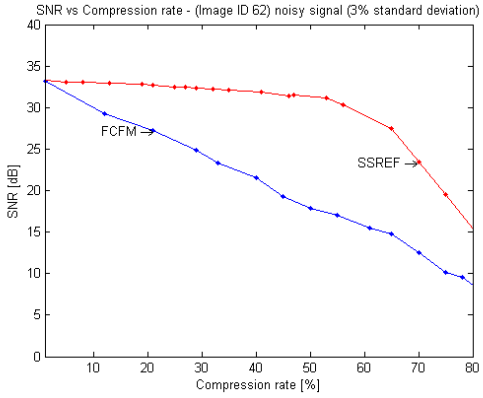


Figure 5.29: SNR vs. compression rate for fluorescence microscopy measurements contaminated by Gaussian noise with standard deviation of 3%. FCFM approach achieves SNR that is significantly lower than the one achieved using the proposed SSREF model. For example, at the same compression rate (60% for example), the SSREF model outperforms the FCFM approach by 10.7 dB (32.5 dB vs. 21.9 dB) for Gaussian noise with a standard deviation of 3%

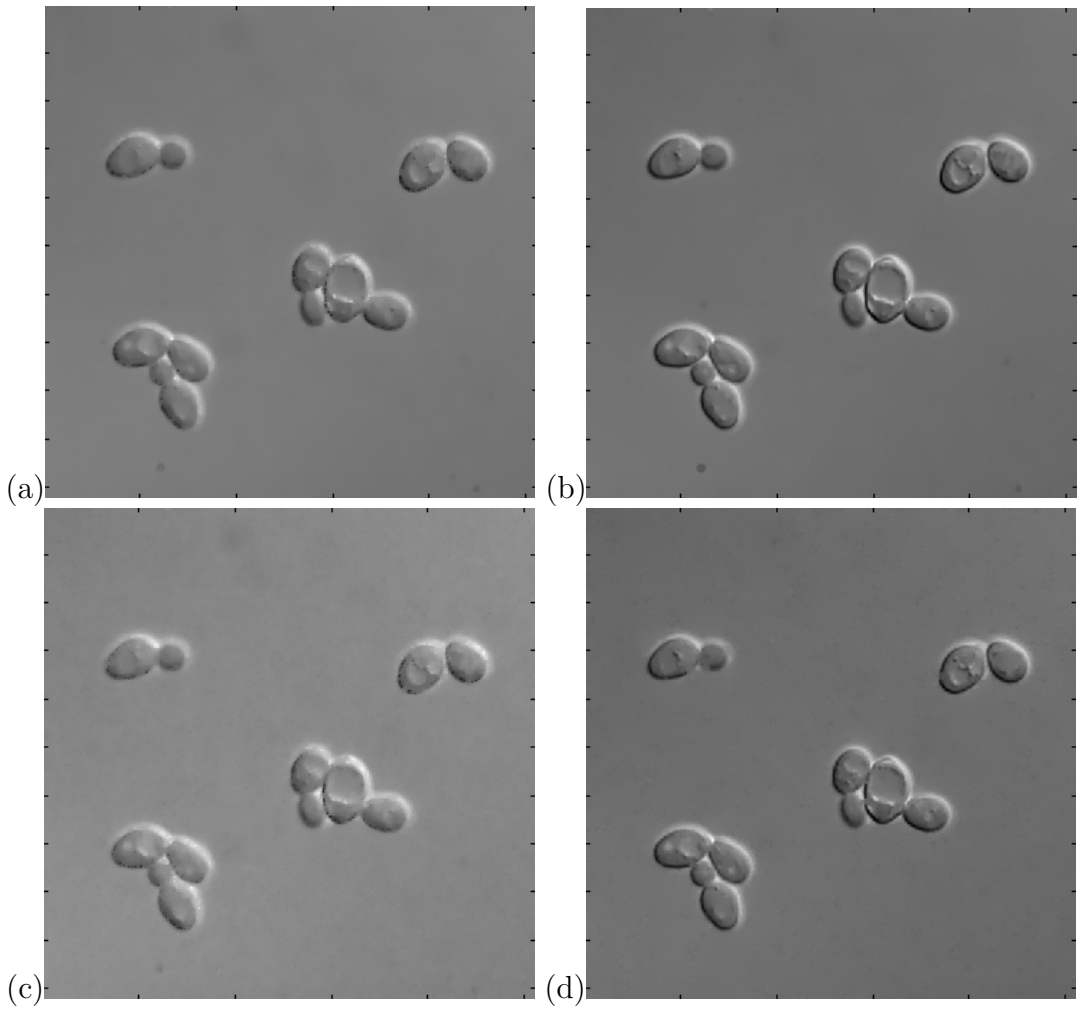


Figure 5.30: Image ID: 64 with synthetic noise. The fluorescence microscopy images produced using the proposed SSREF model contain significantly more important details compared to the FCFM approach. SSREF preserves edges much better in both noise level cases (3% and 5.5%). SSREF reconstruct much better objects of interest since the sampling distribution is saliency guided while the conventional CS sample the entire scene with uniform distribution:
(a) FCFM reconstruction at 75% compression rate and 3% noise level, (b) SSREF reconstruction at 75% compression rate and 3% noise level, (c) FCFM reconstruction at 75% compression rate and 5.5% noise level, (d) SSREF reconstruction at 75% compression rate and 5.5% noise level

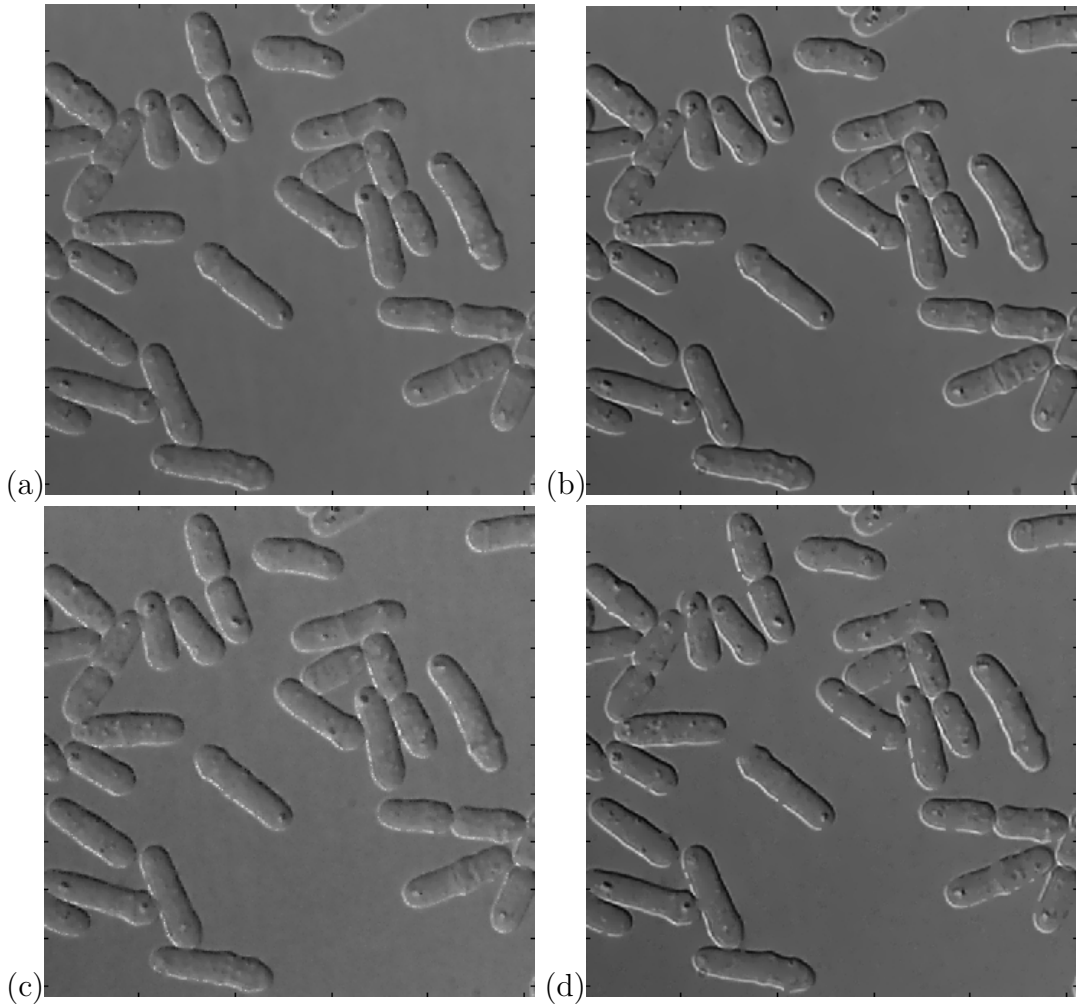


Figure 5.31: Image ID: 5 with synthetic noise. Similarly to Fig. 5.30, SSREF image reconstruction performance outperforms FCFM. This is most evident in the weak structure definition in the fluorescence microscopy images constructed using FCFM, which is well captured in the fluorescence microscopy images produced using the SSREF model. In addition to better edge preserves, SSREF provides improved noise suppression in the background regions more efficiently compared to FCFM:
(a) FCFM reconstruction at 60% compression rate and 3% noise level, (b) SSREF reconstruction at 60% compression rate and 3% noise level, (c) FCFM reconstruction at 65% compression rate and 5.5% noise level, (d) SSREF reconstruction at 65% compression rate and 5.5% noise level

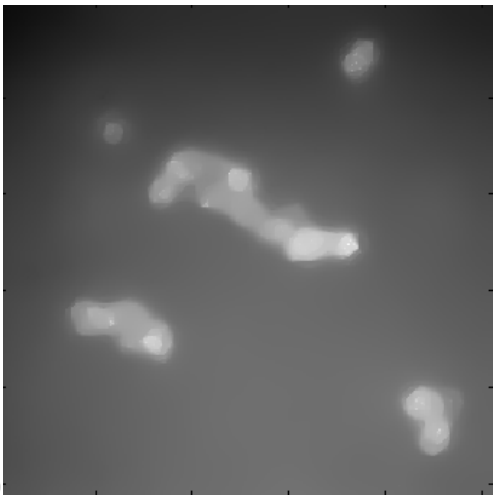
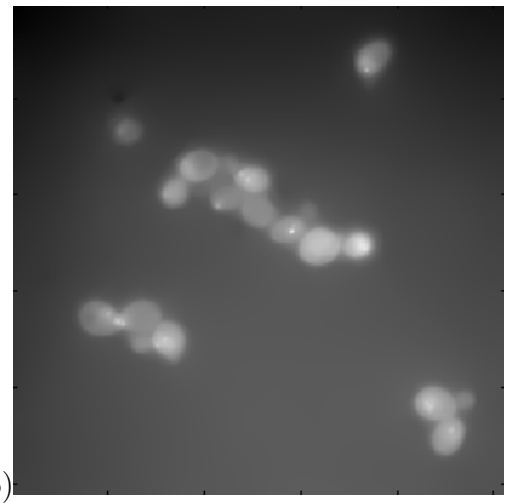
(a)  (b) 

Figure 5.32: Image ID: 8565. SSREF preserve edges of noisy fluorescence microscopy much better than FCFM. For example, elements in the cluster can be identified as separate object at the image produced by SSREF while conventional CS reconstruction produced clusters with blurred and noisy edges. It is most evident that the saliency guided sampling distribution reconstruct regions of interest such as edges, much better compared to conventional CS approach which samples the entire scene with uniform distribution.
(a) FCFM at 84% compression rate, (b) SSREF at 84% compression rate

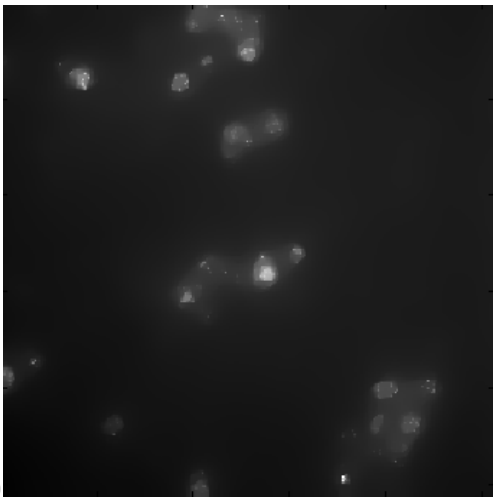
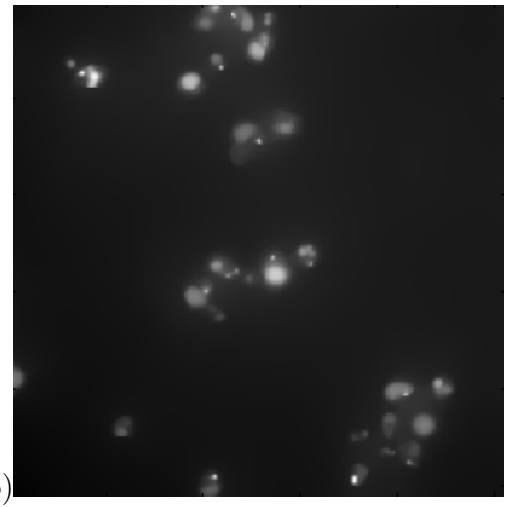
(a)  (b) 

Figure 5.33: Image ID: 3499:
(a) FCFM at 85% compression rate, (b) SSREF at 85% compression rate

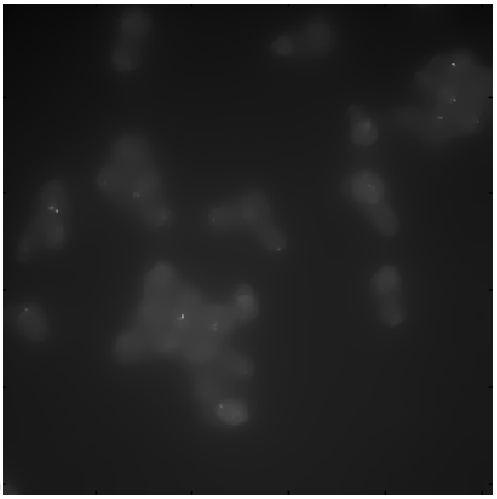
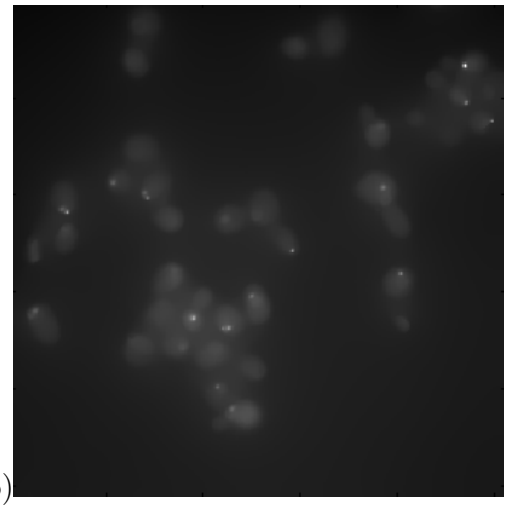
(a)  (b) 

Figure 5.34: Image ID: 5352:
(a) FCFM at 78% compression rate, (b) SSREF at 78% compression rate

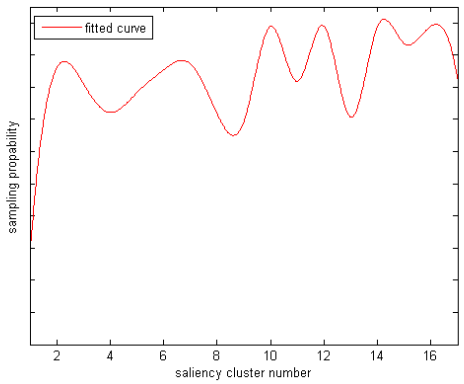


Figure 5.35: Learning an optimized sampling probability function for fluorescence microscopy image for testing ('ID 1-1') a continues function is estimated by a curve fitting to $G(\theta_j^{\mu})$

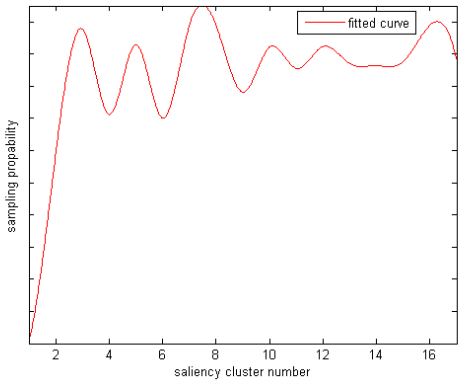


Figure 5.36: Learning an optimized sampling probability function for laser range image for testing ('ID 2-11') a continues function is estimated by a curve fitting to $G(\theta_j^{\mu})$

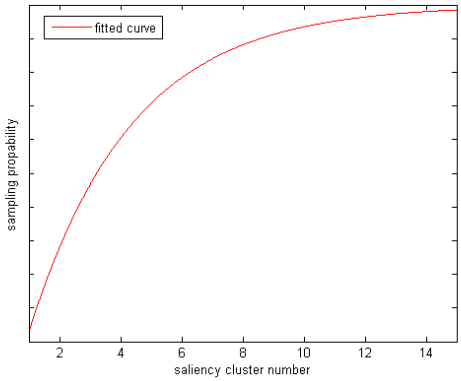


Figure 5.37: Static logarithmic sampling probability function used as reference a continues function is estimated by a curve fitting to $G(\theta_j^{\mu})$

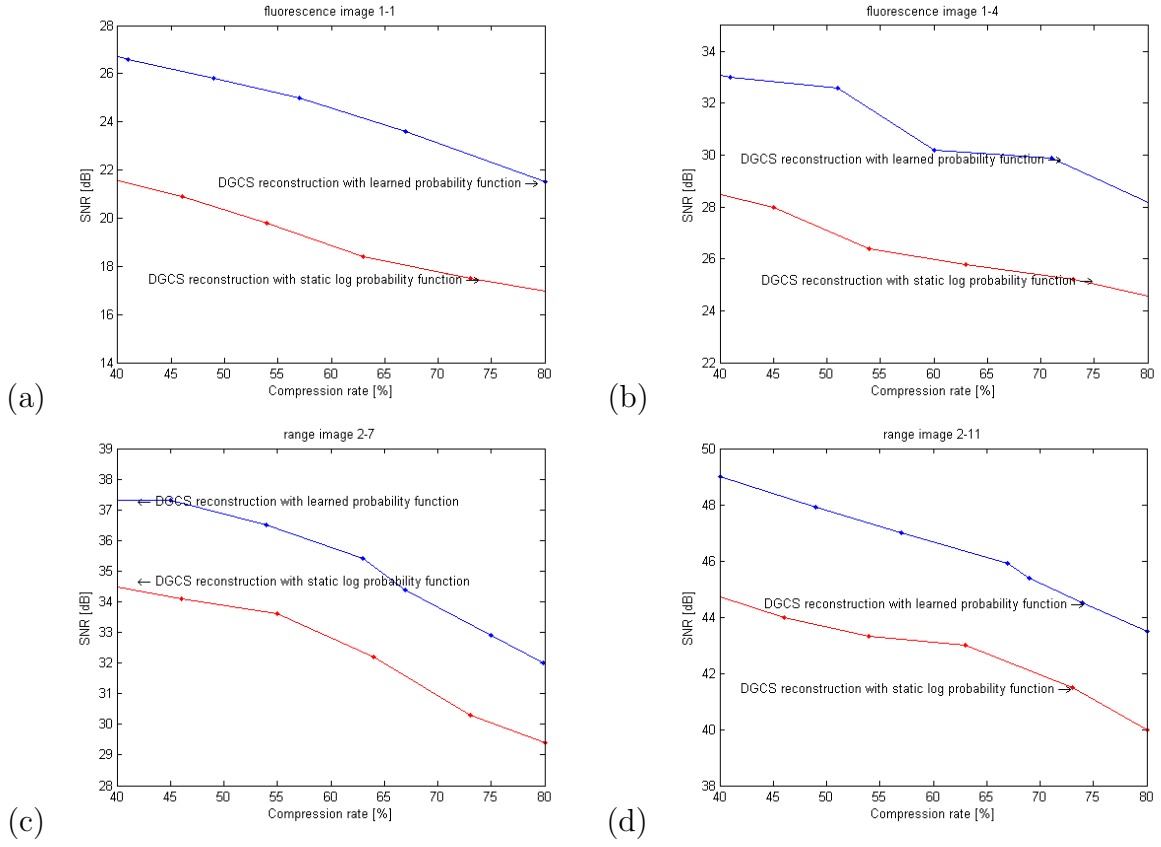


Figure 5.38: Reconstruction performance at different compression rates of dynamically learned sampling probability function vs. static log function for different images
(a) Fluorescence microscopy image ID1-1,
(b) Fluorescence microscopy image ID1-4,
(c) Range measurement indoors image ID2-7,
(d) Range measurement outdoors image ID2-11

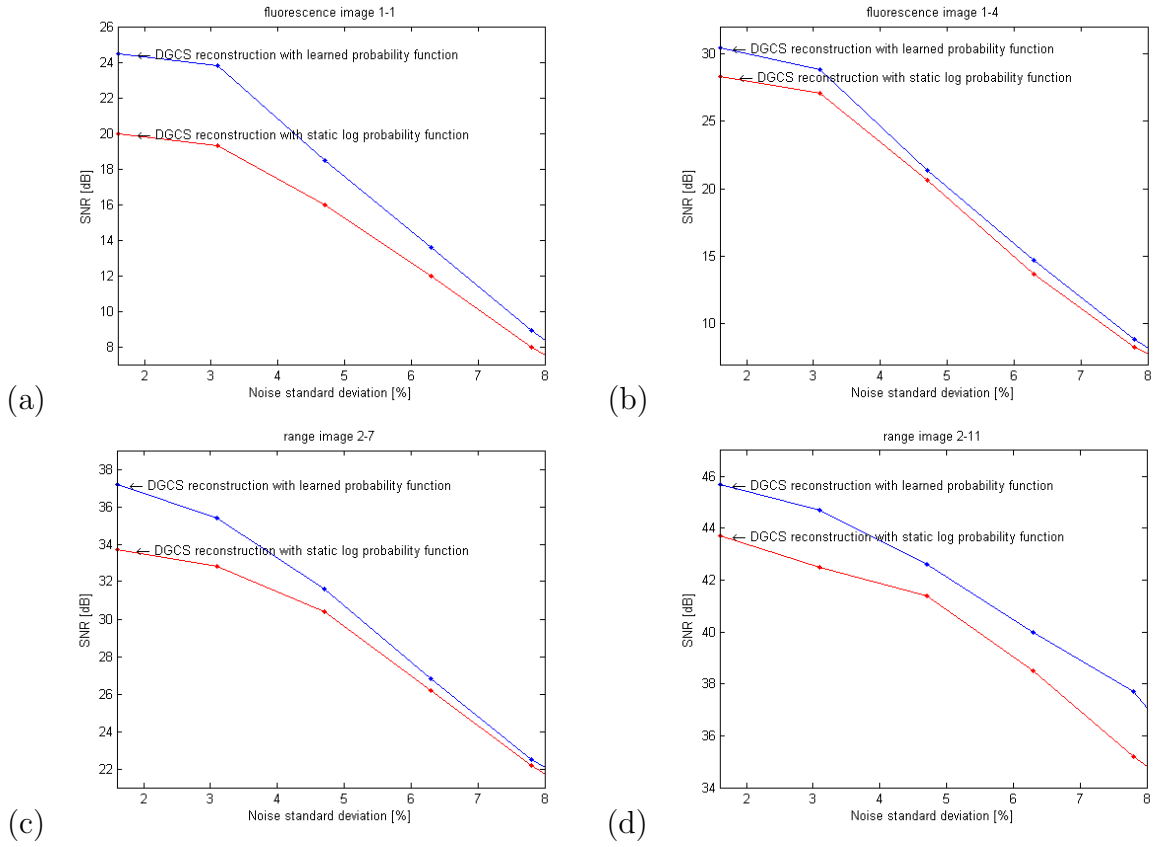


Figure 5.39: Noise analysis of dynamically learned sampling probability function vs. static log function for fluorescence microscopy and range measurements at 50% compression rate at different noise levels
(a) Fluorescence microscopy image ID1-1,
(b) Fluorescence microscopy image ID1-4,
(c) Range measurement indoors image ID2-7,
(d) Range measurement outdoors image ID2-11

Chapter 6

Summary of Contributions and Future Research

In this work, compressive sensing based data-guided statistical sparse measurements method is proposed, implemented and evaluated. This method provides an efficient acquisition approach at multiple sampling domains and can be implemented for wide range of applications. A significant image reconstruction performance improvement was shown, compared to the conventional CS. The core idea of data-guided statistical sparse measurements approach, is image sampling by an optimized sampling patterns based on learning data statistical properties. The novelty of this method is the constructing sampling patterns distribution based on underlying data rather than that commonly used in the CS community, namely, the uniform distribution.

The learning process of nonparametric pdf is demonstrated through two methods. The direct approach as the name imples, learns directly from the dataset. The indirect approach learns indirectly after performing mapping between extracted features and the pdf. A unified model is developed and demonstrated for different representation domains, such as frequency domain and spatial domain. Implementation of a unified model is implemented and evaluated for several applications, such as optical coherence tomography, bridge structure vibration, robotic vision, 3D laser measurements and fluorescence microscopy.

Experimental results are consistent through different applications and sampling domains, showing that data-guided statistical sparse measurements method provides significant improvement compared to the conventional CS. For example, in the OCT application where sampling is in the frequency domain, it was shown that 4.1dB PSNR improvement can be achieved for retinal data at 65% sampling rate compared with conventional CS

sampling method. Furthermore the proposed design allows for substantial reduction in the number of samples needed to achieve similar performance compared to conventional uniform CS sampling approach. For example, it was shown that to achieve retina reconstruction PSNR of 16 dB, the data-guided statistical sparse measurements approach achieves similar performance as the conventional uniform CS sampling approach using 45% fewer samples.

Another example where the performance of the proposed model is demonstrated for laser range data acquisition in the spatial domain. In this case, the proposed design allows for substantial reduction in the number of samples needed to achieve similar performance as the conventional CS approach. For example, to achieve reconstruction SNR of 40 dB, the data-guided statistical sparse measurements approach achieves similar performance as the standard CS approach using 60% fewer samples. In further experiments for compressive fluorescence microscopy application the data-guided statistical sparse measurements method demonstrated its robustness to noise for experiments involves synthetic noise as well as real noisy data.

The data pdf can be selected as a static function based on some previous knowledge or optimized learning based function. Experiments were performed with fluorescence microscopy and range measurement images, for comparing a static parametric function with an optimized nonparametric based on learning pdf. Experimental results show significant improvement in reconstruction SNR were the optimized learned based approach is used. For example, in the case of fluorescence microscopy images, the nonparametric learned based approach outperforms the static approach by an average of 4.4 dB in reconstruction SNR for compression rates range between 40% and 80%.

This work offers a core technology for improved and robust CS based acquisition design. In future, it will be interesting to investigate additional features that can be mapped to sampling distribution to improve further CS reconstruction. Multi-dimensional feature base can be researched, considering features vector which includes color saliency map, texture based features and more. In this case the learned mapping function will map multi dimensional feature vector to optimal sampling distribution. The learning methods can be investigated further in future research for cases where the training dataset is not rich enough, therefore more relevant data is needed. In this case, an adaptive learning method can be researched, where tested data is added adaptively to the training dataset to improve the quality of the training dataset for next tests. In addition, it can be interesting to research in the future the correlation and trade-offs between sparse representation domains and sparse sampling for optimizing further CS reconstruction. In future, it will be interesting to investigate hardware implementation of the data-guided CS approach as well as integration within a system such as fluorescence microscopy or OCT. Finally, this data-

guided statistical sparse measurements system can be easily extended to other imaging applications as well.

References

[1] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk. Salient region detection and segmentation. *Lecture Notes in Computer Science*, 5008:66–75, 2008.

[2] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. *IEEE International Conference on Computer Vision and Pattern Recognitio*, pages 1597–1604, June 2009.

[3] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statistics and Computing*, 18(1):343–373, 2008.

[4] J. Antoni. Blind separation of vibration components: Principles and demonstrations. In *Mechanical Systems and Signal Processing*, volume 19, pages 1166–1180, 2005.

[5] E. Arias-Castro, E. Candes, and M. Davenport. On the fundamental limits of adaptive sensing. *Preprint: arXiv:1111.4646*, 2012.

[6] G.Meynants W. Ogiers B. Dierickx, D. Scheffer and J. Vlummens. Random addressable active pixel image sensors. In *Proc. SPIE*, volume 2950, pages 2–7, 1996.

[7] Y. Bao, J.L Beck, and H. Li. Compressive sampling for accelerometer signals in structural health monitoring. *Structural Health Monitoring*, 10:235–246, 2010.

[8] R. Baraniuk, M. Davenport, R. DeVore, , and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[9] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419 – 2434, 2009.

[10] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.*, 1:183–202, 2009.

[11] M.C. Jones Bernard and W. Silverman. Fix, e. and hodges, j. l. (1951): An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951). *International Statistical Review / Revue Internationale de Statistique*, 57(3):233247, 1989.

[12] P.J. Besl and N.D. McKays. A method for registration of 3d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14:239–254, 1992.

[13] F. Blais. Review of 20 years of range sensor development. *J. Electron. Imag.*, 13(1):231–240, 2004.

[14] M. Bocca, L.M. Eriksson, A. Mahmood, R. Jantti, and J. Kullaa. A synchronized wireless sensor network for experimental modal analysis in structural health monitoring. *Computer-Aieded Civil and Infrastucture Engineering*, 26:483–499, 2011.

[15] L. Breiman, W.S. Meisel, and E. Purcell. Variable kernel estimates of multivariate densities. *Technometrics*, 19:1351144, 1977.

[16] A.M. Bruckstein, M. Elad, and M. Zibulevsky. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. *IEEE Trans. on Information Theory*, 54(11):4813 – 4820, 2008.

[17] W. Burgard and et al. Experiences with an interactive museum tourguide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.

[18] E. Candes and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Computational Mathematics*, 6(2):227–254, 2006.

[19] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1221, 2006.

[20] E. J. Candes. Restricted isometry property and its implications for compressed sensing. *Comptes rendus - Mathematique*, 346(9-10):589–592, May 2008.

[21] E. J. Candes and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Probl.*, 23(3):969–985, 2007.

[22] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inform. Theory*, 52(3):5406–5425, 2004.

[23] E.J. Candes and M.B. Wakin. An introduction to compressive sampling. *IEEE Sig. Proc. Mag.*, 25(2):21–30, 2008.

[24] R.M. Castro. Adaptive sensing performance lower bounds for sparse signal estimation and testing. *Preprint: arXiv:1206.0648*, 2012.

[25] N.N. Cencov. Evaluation of an unknown distribution density from observations. *Doklady*, 3:1559–1562, 1962.

[26] F. Chen, G. M. Brown, and M. Song. Overview of three-dimensional shape measurement using optical methods. *Opt. Eng.*, 39:10–22, 2000.

[27] G. H. Chen, J. Tang, and S. H. Lengr. Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Med. Phys.*, 35(2):660–663, 2008.

[28] H. Chen. Tutorial on monte carlo sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, 2008.

[29] S. Chen, Y. Li, and J. Zhang. Vision processing for realtime 3-d data acquisition based on coded structured light. *IEEE Trans. on Image Processing*, 12(2), 2008.

[30] S. S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sc. Comp.*, 20(1):33–61, 1998.

[31] S. Chib and I. Jeliazkov. Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation. *Statistica Neerlandica*, 59(1):30–44, 2005.

[32] M.A. Choma, M.V. Sarunic, C. Yang, and J. A. Izatt. Sensitivity advantage of swept source and Fourier domain optical coherence tomography. *Optics Express*, 11:2183, 2003.

[33] R. Connally, D. Veal, and J. Piper. High resolution detection of fluorescently labeled microorganisms in environmental samples using time-resolved fluorescence microscopy. *FEMS Microbiol Ecol*, 41:239245, 2002.

[34] R. Connally, D. Veal, and J. Piper. Novel flashlamp based timeresolved fluorescence microscope reduces autofluorescence for 30-fold contrast enhancement in environmental samples. *Proc. SPIE*, 4964:1423, 2003.

[35] R. Connally, D. Veal, and J. Piper. Flash lamp-excited time-resolved fluorescence microscope suppresses autofluorescence in water concentrates to deliver an 11-fold increase in signal-to-noise ratio. *J Biomed Opt*, 9:725734, 2004.

[36] I. Cox. Blanche an experiment in guidance and navigation of an autonomous robot vehicle. *IEEE Transactions on Robotics and Automation*, 7(2):193–204, 1991.

[37] J.S. Cui, H.B. Zha, H.J. Zhao, and R. Shibasaki. Multi-modal tracking of people using laser scanners and video camera. *Image Vision Computing*, (2):240–252, 2008.

[38] C. Daniel and F.S. Wood. In *Fitting Equations to Data: Computer Analysis of Multifactor Data, 2nd Edition, John Wiley and Sons*, 1980.

[39] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–3457, 2004.

[40] M. Davenport and et al. Joint manifolds for data fusion. *IEEE Trans. Image Processing*, 19(10):2580–2594, 2010.

[41] M. A. Davenport and et al. Introduction to compressed sensing. in compressed sensing: Theory and applications, chapter 1. In *Cambridge University Press*, 2012.

[42] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[43] D. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J. AMS*, 22(1):1–15, January 2009.

[44] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202, 2003.

[45] D.L. Donoho and X. Huo. Uncertainty principles and ideal atom decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, November 2001.

[46] D.L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 53(12):4655–4666, 1989.

[47] W. Drexler and J. G. Fujimoto. Optical coherence tomography. In *Springer Berlin Heidelberg*, 2008.

[48] M.F. Duarte. Single-pixel imaging via compressive sampling (building simpler, smaller, and less-expensive digital cameras). *IEEE Sig. Proc. Mag.*, 25(2):83–91, 2008.

[49] M.F. Duarte and Y.C. Eldar. Structured compressed sensing from theory to applications. *IEEE Tran. Sig. Proc.*, 59(9):4053–4085, 2011.

[50] H. Durrant-Whyte. An autonomous guided vehicle for cargo handling applications. *International Journal of Robotics Research*, 15(2):407–441, 1996.

[51] J. Romberg E. Candes and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[52] M. Elad. Sparse and redundant representations: From theory to applications in signal and image processing. In *ISBN 978-1-4419-7010-7*, pages 3–16, 2010.

[53] M. Elad. Sparse and redundant representations: From theory to applications in signal and image processing. In *ISBN 978-1-4419-7010-7*, pages 17–34, 2010.

[54] M. Elad. Sparse and redundant representations: From theory to applications in signal and image processing. In *ISBN 978-1-4419-7010-7*, pages 35–55, 2010.

[55] M. Elad. Sparse and redundant representations: From theory to applications in signal and image processing. In *ISBN 978-1-4419-7010-7*, pages 79–110, 2010.

[56] M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 48(9):2558–2567, 2002.

[57] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky. A wide-angle view at iterated shrinkage algorithms. In *Proceedings of SPIE (Wavelet XII)*, San-Diego CA, USA, 2007.

[58] M.L Gabriele, G. Wollstein, H. Ishikawa, L. Kagemann, J. Xu, L.S Folio, and J.S Schuman. Optical coherence tomography: history, current status, and laboratory work. *Invest Ophthalmol Vis Sci.*, 52:2425–2436, 2011.

[59] D. Gamerman. Markov chain monte carlo. In *Stochastic simulation for Bayesian inference, Chapman and Hall*, 1997.

[60] Y. Gao, J.B.F Spencer, and M. Ruiz-Sandoval. Distributed computing strategy for structural health monitoring. *Structural control and health monitoring*, 13:488–507, 2006.

[61] K. Gedalyahu and et al. Time-delay estimation from low-rate samples: A union of subspaces approach. *IEEE Trans. Image Processing*, 58(6):3017–3031, 2010.

[62] W. Geitzenauer, C.K Hitzenberger, and U.M. Schmidt-Erfurth. Retinal optical coherence tomography: past, present and future perspectives. *Br J Ophthalmol*, 95:171–177, 2011.

[63] I.J. Good and R.A. Gaskins. Nonparametric roughness penalties for probability densities,. *Technometrics*, 58:255–277, 1971.

[64] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.

[65] J. Guivant, J. Masson, and E. Nebot. Navigation and mapping in large unstructured environments. *International Journal of Robotic Research*, 2003.

[66] J. Guivant and E. Nebotl. Optimization of the simultaneous localization and map-building algorithm for real-time implementation. *IEEE Journal of Robotics and Automation*, 17(3):242–257, 2001.

[67] W. K. Hastings. Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[68] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak. Compressive sensing for networked data. *IEEE Signal Processing Magazine*, 25(2):92–101, 2008.

[69] J. Haupt, R. Baraniuk, R. Castro, and R.Nowak. Sequentially designed compressed sensing. *Preprint: (available at http://www.win.tue.nl/ rmcastro/publications/SCS.pdf)*, 2012.

[70] B. Hazra, A. Sadhu, R. Lourenco, and S. Narasimhan. Retuning tuned mass dampers using ambient vibration response. *Smart Materials and Structures,*, 19(11):115002, 2010.

[71] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[72] Y. Hu. Matlab central. In *http://www.mathworks.com/matlabcentral/fileexchange/22274-hill-climbing-color-image-segmentation*.

[73] D. Huang, E.A. Swanson, C.P. Lin, J.S. Schuman, W.G. Stinson, W. Chang, M.R. Hee, T. Flotte, K. Gregory, C. A. Puliato, and J. G. Fujimoto. Optical coherence tomography. *Science*, 254:1178–1181, 1991.

[74] S. Inoue and K. R. Spring. Video microscopy. *New York: Plenum Press*, 13, 1997.

[75] A. J. Izenmana. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, 1991.

[76] J. A. Jalkio, R. C. Kim, and S. K. Casei. Three dimensional inspection using multi-stripe structured light. *Opt. Eng.*, 24(6):966–974, 1985.

[77] J.-H. Kim and S. Sukkarieh. Airborne simultaneous localization and map building. In *IEEE Proceedings on the Conference on Robotics and Automation, Taipei, Taiwan*, 2003.

[78] T. Klein, W. Wieser, C. M. Eigenwillig, B. R. Biedermann, and R. Huber. Megahertz OCT for ultrawide-field retinal imaging with a 1050 nm Fourier domain mode-locked laser. *Optics Express*, 19(4):3044–3062, 2011.

[79] M.D Kohler, M.P. Davis, and E. Safak. Earthquake and ambient vibration monitoring of the steel-frame ucla factor building. *Earthquake Spectra, Earthquake Engineering Research Institute,*, 21(3):1–22, 2005.

[80] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1281–2011, 2011.

[81] M. Ledoux. The concentration of measure phenomenon. *American Mathematical Society*, 89, 2001.

[82] R. Leitgeb, W. Drexler, A. Unterhuber, B. Hermann, T. Bajraszewski, T. Le, A. Stingl, and A. Fercher. Ultrahigh resolution Fourier domain optical coherence tomography. *Opt. Express*, 12:2156–2165, 2004.

[83] C. Liu, A. Wong, K. Bizheva, P. Fieguth, and H. Bie. Homotopic, non-local sparse reconstruction of optical coherence tomography imagery. *Opt. Express*, 20:10200–10211, 2012.

[84] X. Liu and J. U. Kang. Compressive SD-OCT: the application of compressed sensing in spectral domain optical coherence tomography. *Opt. Express*, 18:22010–22019, 2010.

[85] D.O. Loftsgaarden and C.P. Quesenberry. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, 36:10491051, 1965.

[86] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[87] J.P. Lynch. An overview of wireless structural health monitoring for civil structures. *Philosophical Transcations of the Royal Society,*, 365:345–372, 2007.

[88] J.P. Lynch, A. Sundararajan, K.H Law, A.S. Kiremidjian, E. Carryer, H. Sohn, and C.R. Farrar. Field validation of a wireless structural monitoring system on the alamosa canyon bridge. In *SPIE's 10th Annual International Symposium on Smart Structures and Materials, San Diego, CA, USA*, pages 267–278, 2003.

[89] F. Magalhaes, E. Caetano, and A. Cunha. Challenges in the application of stochastic modal identification methods to a cable-stayed bridge. *Journal of Bridge Engineering,*, 12(6):746–764, 2007.

[90] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Tran. Sig. Proc.*, 41(12):3397–3415, 1993.

[91] M Marim and E Angelini. Denoising in fluorescence microscopy using compressed sensing with multiple reconstructions and non-local merging. *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 3394(7):3394–3397, August 2010.

[92] M. Marim, E. Angelini, and J. C. Olivo-Marin. Compressed sensing in biological microscopy. *in Proc. SPIE Wavelets XIII*, 7446:3394–3397, 2009.

[93] S. Martinez-Conde, S.L. Macknik, and D.H. Hubel. The role of fixational eye movements in visual perception. *Br J Ophthalmol*, 5:229–240, 2004.

[94] RA. Mathies, K. Peck, and L. Stryer. Optimization of high-sensitivity fluorescence detection. *Analytical chemistry*, 62:17861791, 1990.

[95] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for bernoulli and subgaussian ensembles. *Constructive Approximation*, 28(3):277–289, 2008.

[96] M. Mishali and et al. Xampling: Analog to digital at sub-nyquist rates. *IET Circuits, Devices, and Systems*, 5(1):8–20, 2011.

[97] N. Mohan, I. Stojanovic, W. C. Karl, B. E. A. Saleh, and M. C. Teich. Compressed sensing in optical coherence tomography. *SPIE*, 7570:75700L–75700L–5, 2010.

[98] WC. Moss and et al. A novel 3d wavelet-based filter for visualizing features in noisy biological data. *J Microscopy*, 219:4349, 2005.

[99] NASA. Mars exploration rovers. In *http://marsrovers.jpl.nasa.gov/home/*.

[100] B.K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comp.*, 24(2):227–234, 1995.

[101] P. Newman and J. L. Rikovski. Towards constant-time SLAM on an autonomous underwater vehicle using synthetic aperture sonar. In *International Symposium of Robotics Research, Siena, Italy*, 2003.

[102] H. Nyquist. Certain topics in telegraph transmission theory. *EE Trans*, 47:617644, January 1928.

[103] N. Panchuk-Voloshina and et al. Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *J Histochem Cytochem*, 47:11791188, 1999.

[104] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. 27th Annual Asilomar Conf. Sig. Sys. Comp.*, 1993.

[105] The Brown University pattern theory group. Brown range image database. In *http://www.dam.brown.edu/ptg/brid/range/index.html*.

[106] J. B. Pawley and B. R. Masters. Handbook of biological confocal microscopy, third edition. *Journal of Biomedical Optics*, 13(029902), 2008.

[107] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Pattern Anal Mach Intell*, 12:629639, 1990.

[108] H R. Petty. Fluorescence microscopy: Established and emerging methods, experimental strategies, and applications in immunology. *Microscopy Research and Technique*, 70(8):687–709, 2007.

[109] S. Pfetsch and et al. On the feasibility of hardware implementation of sub-nyquist random-sampling based analog-to-information conversion. In *IEEE Int. Symp. Circuits and Systems (ISCAS), Seattle, WA*, page 14801483, 2008.

[110] D. Piston. Choosing objective lenses: The importance of numerical aperture and magnification in digital microscopy. *The Biological Bulletin*, 195:14, 1998.

[111] P. Puvanathasan, P. Forbes, Z. Ren, D. Malchow, S. Boyd, and K. Bizheva. High-speed, high-resolution Fourier-domain optical coherence tomography system for retinal imaging in the 1060 nm wavelength region. *Opt. Lett*, 33:2479–2481, 2008.

[112] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *IEEE Signal Processing Letters*, 18(10):595–598, 2011.

[113] F. Geshwind R. Coifman and Y. Meyer. Noiselets. *App. Comp. Harm. Anal.*, 10(1):27–44, 2001.

[114] BR Renikuntla and et al. Improved photostability and fluorescence properties through polyfluorination of a cyanine dye. *Organic Lett.*, 6(6):909912, 2004.

[115] Davis TN (2010) The Yeast Resource Center Public Image Repository: A large database of fluorescence microscopy images. . 2010 May 19;11(1):263 Riffle M. The yeast resource center public image repository: A large database of fluorescence microscopy images. volume 11, page 263, 2010.

[116] M. Rioux and L. Cournoyer. The nrcc three-dimensional image data files. *Technical Report CNRC No. 29077, National Research Council of Canada, Ottawa*, 1988.

[117] C. P. Robert and G. Casella. Monte carlo statistical methods. In *New York: Springer-Verlag*, 1999.

[118] R. Robucci, L.K. Chiu, J. Gray, J. Romberg, P. Hasler, and D. Anderson. Compressive sensing on a cmos separable transform image sensor. In *IEEE Int. Conf. Ac. Speech Sig. Proc.*, pages 5125–5128, 2008.

[119] B. Hazraand A.J. Roffel, S. Narasimhan, and M.D. Pandey. Modified crosscorrelation method for the blind identification of structures. *Journal of Engineering Mechanics*, 136(2):880–897, 2010.

[120] J. Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, 2008.

[121] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832837, 1956.

[122] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. *40th An. Conf. Inf. Sc. Sys.*, pages 207–212, 2006.

[123] S. Sabri and et al. Interest of image processing in cell biology and immunology. *J Immunol Methods*, 208:127, 1997.

[124] A. Sadhu, B. Hazra, S. Narasimhan, and M.D. Pandey. Decentralized modal identification using sparse blind source separation. In *Smart Materials and Structures, IOP Publishing*, volume 20, page 125009, 2011.

[125] S.Chiba and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[126] S. Schwartz, C. Liu, A. Wong, D. A. Clausi, P. Fieguth, and K. Bizheva. Energy-guided learning approach to compressive FD-OCT. *journal Optics Express*, page submitted, 2012.

[127] S. Schwartz, A. Wong, and D. A. Clausi. Compressive fluorescence microscopy using saliency-guided sparse reconstruction ensemble fusion. *journal Optics Express*, 20(16):17281–17296, 2012.

[128] S. Schwartz, A. Wong, and D. A. Clausi. Multi-scale saliency-guided compressive sensing approach to efficient robotic laser range measurements. In *Ninth Conference on Computer and Robot Vision (CRV 2012), San Diego, CA USA*, 2012.

[129] S. Schwartz, A. Wong, and D. A. Clausi. Saliency-guided compressive fluorescence microscopy. In *34th Annual International Conference of the IEEE Engineering in Medicine and Biology (EMBC12), Toronto, Ontario Canada*, number http://doi.ieeecomputersociety.org/10.1109/CRV.2012.8, 2012.

[130] S. Schwartz, A. Wong, and D. A. Clausi. Saliency-guided compressive sensing approach to efficient laser range measurement. *Journal of Visual Communication and Image Representation*, 24(2):160–170, 2013.

[131] C. E. Shannon. Communications in the presence of noise. *in Proc. IRE*, 37:10–21, January 1949.

[132] C. E. Shannon. Classic paper: Communication in the presence of noise. *in Proc. IEEE*, 86(2):447457, January 1998.

[133] R. Siegwart and I. R. Nourbakhsh. Introduction to autonomous mobile robots. *MIT press*, pages 181–256, 2004.

[134] B.W. Silverman. Density estimation for statistics and data analysis. In *Monographs on Statistics and Applied Probability, London: Chapman and Hall*, 1986.

[135] S.H. Sim, B.F. Spencer, M. Zhang, and H. Xie. Automated decentralized modal analysis using smart sensors. *Structural Control and Health Monitoring, Wiley Interscience,*, 17(8):872–894, 2009.

[136] A. Mart S.M. Potter and J. Pine. High-speed CCD movie camera with random pixel selection for neurobiology research. In *Proc. SPIE*, volume 2869, page 243253, 1997.

[137] R. Smith and P. Cheeseman. On the representation and esitmation of spatial uncertainty. *International Journal of Robotics Research*, 5(4):56–68, 1986.

[138] L. Song, E. J. Hennink, T. Young, and H. J. Tanke. Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophysical Journal*, 68:25882600, 1995.

[139] A.A. Portillo S.P. Monacos, R.K. Lam and G.G. Ortiz. Design of an event-driven random-access-windowing CCD-based camera. In *Proc. SPIE*, volume 4975, page 115, 2003.

[140] J. Starck, F. Murtagh, and J. Fadili. Sparse image and signal processing - wavelets, curvelets, morphological diversity. In *ISBN 9780521119139*, pages 16–75, 2010.

[141] V. Studer, J. Bobin, M. Chahid, H. Moussavi, E. J. Candes, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences of the United States of America*, page 10 pages, 2011.

[142] J. Treichler and et al. Application of compressive sensing to the design of wideband signal acquisition receivers. In *Proc. U.S./Australia Joint Work. Defense Apps. of Signal Processing (DASP), Lihue, Hawaii*, 2009.

[143] J. Tropp and A.C. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53(12):4655–4666, 2007.

[144] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk. Nyquist: Efficient sampling of sparse, bandlimited signals. *IEEE Trans. Inform. Theory*, 56(1):520–544, 2010.

[145] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[146] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory*, 51(3):1030–1051, 2006.

[147] J. Trzasko, A. Manduca, and E. Borisch. Highly undersampled magnetic resonance image reconstruction via homotopic L0-minimization. *IEEE Transactions on Medical Imaging*, 28(1):106–121, 2009.

[148] M. Unser. Sampling-50 years after Shannon. *Proceedings of the IEEE*, 88(4):569–587, April 2000.

[149] J. Wang, M. Abou Shousha, V.L Perez, C.L Karp, S.H Yoo, M. Shen, L. Cui, V. Hurmeric, C.Du, D. Zhu, Q. Chen Q, and M. Li. Ultra-high resolution optical coherence tomography for imaging the anterior segment of the eye. *Ophthalmic Surg Lasers Imaging*, doi: 10.3928/15428877-20110627-02, 2011.

[150] Z. Wang and G. Arce. Variable density compressed image sampling. *IEEE Transactions on Image Processing*, 19(1):264–270, 2010.

[151] P. Whittle. On the smoothing of probability density functions. *Soc. Ser. B*, 55:549557, 1958.

[152] R. Willett, A. Martin, and R. Nowak. Backcasting: adaptive sampling for sensor networks. In *Third International Symposium on Information Processing in Sensor Networks (IPSN 2004)*, page 124133, 2004.

[153] Yuehao Wu, Peng Ye, Iftekhar O. Mirza, Gonzalo R. Arce, and Dennis W. Prather. Experimental demonstration of an optical-sectioning compressive sensing microscope (csm). *Opt. Express*, 18:24565–24578, 2010.

[154] M. Young, E. Lebed, Y. Jian, P. J. Mackenzie, M. F. Beg, and M. V. Sarunic. Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography. *Optics Express*, 2:2690–2697, 2011.

[155] M. Young, E. Lebed, Y. Jian, P.J. Mackenzie, M.F. Beg, and M.V. Sarunic. Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography. *Biomedical Optics Express*, 2(9):2690, 2011.

[156] Y. Yu, B. Wang, and L. Zhang. Saliency-based compressive sampling for image signals. *IEEE Signal Processing Letters*, 17(11):973 – 976, 2010.

[157] J. Zakrzewski. Integrating a spectrometer with an optical microscope presents challenges. *SPIE OE Magazine*, page 29, November 2003.

[158] Y. Zhang and J. Li. Wavelet-based vibration sensor data compression technique for civil infrastructure condition monitoring. *Journal of Computing in Civil Engineering,*, 20(6):1461–1466, 2006.

[159] A.T. Zimmerman, M. Shiraishi, R.A. Swartz, and J.P Lynch. Automated modal parameter estimation by parallel processing within wireless monitoring systems. *Journal of Infrastructure System*, 22:102–113, 2008.