PRACTICAL CONSIDERATIONS IN EXPERIMENTAL COMPUTATIONAL SENSING

by

Phillip K. Poon

A Dissertation Submitted to the Faculty of the

COLLEGE OF OPTICAL SCIENCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2016

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Phillip K. Poon titled Practical Considerations in Experimental Computational Sensing and recommend that it be accepted as fulfilling the dissertation requirement for the degree of Doctor of Philosophy.

Date: 12 December 2016
Amit Ashok

Date: 12 December 2016
Rongguang Liang

Date: 12 December 2016
Michael E. Gehm

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: 12 December 2016
Dissertation Director: Amit Ashok

STATEMENT BY AUTHOR

SIGNED: Phillip K. Poon

ACKNOWLEDGEMENTS

Graduate school is an arduous and enlightening experience. It is not difficult by design but by nature it forces one into a state of mind which embraces the edge of knowledge and trek into the unknown. I was fortunate to have many guides who showed me the path, even when there were times when I wandered off to get my bearings. Along the way I encountered many people who not only helped me with the journey but bestowed kindness and friendship, asking for nothing in return.

My main guide along the journey was Professor Michael Gehm. I first met him when I took a graduate level Linear Algebra course which I found particularly challenging. I often went to his office hours asking for help and his ability to be patient and explain concepts from different perspectives is a gift few teachers have. As an advisor, I would like to thank him for all of the help and guidance he has given me over the years. His generosity for funding my graduate studies as well trips to conferences is appreciated. He believed in me more than I believed in myself. I consider him not only as a mentor but as a father figure. If I can become half the scientist that he is, I would consider that a successful career.

I especially want to thank Professor Esteban Vera, who I first met as a postdoctoral researcher in the Laboratory for Engineering Non-Traditional Sensors (LENS) and supervised me for the majority of my graduate studies. Much of the work and results in this disseration is due to his guidance. Even after he started his professorship in Chile, he was willing to review my data and suggest different methods of analysis. Professor Vera is directly responsible for much of my training as an experimentalist. I consider Professor Vera as an older brother who was always there to protect me from the pitfalls of the graduate school journey.

I also thank Doctor Dathon Golish. His approach to work and life was a calming effect in often stressful times. He made major contributions to the Adaptive Feature Specific Spectral Imaging-Classifier (AFSSI-C) and provided valuable feedback on various research projects and conference presentations.

Thank you Professor Mark Neifeld and Professor Amit Ashok for being my advisor and supervisor during my first year as a PhD student. They were the first to introduce me to many of the techniques and subjects related to computational sensing. They taught me fundamental concepts in optics, statistical signal processing and programming. Many of the results in this dissertation would not have been possible without their teachings.

I've also had many other supervisors along the way whose effort must also be acknowledged: My undergraduate advisor at San Diego State University, Professor Matthew Anderson. Doctor John Crane, who was my supervisor during my internship at the Lawrence Livermore National Laboratory. Professor Joseph Eberly and Professor Gary Wicks who were my advisors at the Institute of Optics.

I would like to formally express gratitude to a number of exceptional teachers throughout my life. Professor Tom Milster who taught Diffraction and Interference and allowed me to be a teaching assistant for that course. Professor Masud Mansuripur, whose course in Electromagnetic Waves was the most elegant and well taught version of the classical nature of light that I have ever had the pleasure to experience. Professor Jeff Davis, who first ignited my passion for optics while I was an undergraduate physics student at San Diego State University.

I also want to thank several faculty members who committed time from their busy schedules to help with several milestones of my graduate school experience. Special thanks to Professor Julie Bentley, Doctor James Oliver, and Professor Richard Morris who wrote letters of recommendation for me. Appreciation goes to Professor Tom Milster, Professor Harrison Barrett, Professor Russell Chipman, and Professor John Greivenkamp who formed my oral comprehensive exam committee. Thank you to Professor Rongguang Liang who served on my doctoral dissertation committee.

I would like to thank several members of the Duke Imaging and Spectroscopy Program (DISP) laboratory for their friendship: Patrick Llull, Mehadi Hassan, and Evan Chen. Tsung Han Tsai was not only a colleague but his work on computational polarimetry and spectroscopy using an Liquid Crystal on Silicon (LCOS) Spatial Light Modulator (SLM) was the inspiration which directly lead my idea of using the same architecture for computational spectral unmixing.

Other graduate students, colleagues, and faculty must also be thanked, for at one time or another they all helped me: Basel Salahieh, Vicha Treeaporn, John Hughes, Steve Feller, Myungjun Lee, Sarmad H. Albanna, Professor Lars Furenlid, Doctor Joseph Dagher, Professor Daniel Marks, Professor Janick Roland-Thompson, Mary Pope, Mark Rodriguez, and Amanda Ferris.

I've had the good fortune to form friendships with an amazing set of groupmates as part of the LENS. David Coccarelli invited our family to spend our first Thanksgiving in North Carolina with him and we had many discussions about college basketball and life. I would express my sincere gratitude to Matthew Dunlop-Gray, who designed and constructed the AFSSI-C which is the foundation for much the work in this disseration. I learned so much from Matthew especially much of my practical skills. Tariq Osman constructed the Static Computational Optical Undersampled Tracker (SCOUT) which is also a major part of this dissertation. Alyssa Jenkins whose combination of sense of humor and intelligence is unmatched. Thank you to Qian Gong for your kindness, positivity, and generousity. Thank you Xiaohan Li for keeping me company that final year of graduate school, conversations about basketball and helping me with my math. Thank you David Landry for helping me with all software and computer programming related issues. Thank you Kevin Kelly, Adriana DeRoos, Andrew Stevens and Dineshbabu Dinakarababu for your friendship. Finally, I consider Wei-Ren Ng as one of my best friends and as a brother. Our time in the LENS group was marked by many late nights spent working in the lab and office. He was generous in sharing his knowledge and gave me the advice that I often did not want to hear but was true.

Appreciation goes to the all the staff at the College of Optical Sciences at the University of Arizona. It is one of the most friendly and well run academic departments I have ever had the fortune to be a part of. I hope my career will reflect well upon the college.

Finally, I would like to thank my closest friends that I've met throughout the years. They often provided much needed respites during my journey—Christopher MacGahan, Ricky Gibson, Krista MacGahan, Kristi Behnke, Michael Gehl, Carlos Montances, Matthew Reaves, Vijay Parachuru, Eric Vasquez. Thank you for letting me into your lives and being part of mine.

Last but not least, to my family. You make me happy.

DEDICATION

For my wife. We moved from city to city. You stuck with me through the highs and lows. You cooked dinner for me when I came home from a long day. You did the chores so I could concentrate on research. You acted as both mother and father to our son while I wrote. You believed in me even when I did not. You sacrificed your dreams and goals so I could accomplish mine.

You're the real Ph.D.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Implementing computational optical sensors often comes with various issues that many traditional sensors may not encounter.

CHAPTER 1

Introduction

This chapter introduces the reader to the concepts of isomorphic sensing, multiplexing, indirect imaging, task-specific sensing, compressive sensing and computational sensing. It also provides the motivation for the need to address the practical issues in experimental computational sensing.

A measurement is a process that converts a physical phenomena to a collection of data. The signal-of-interest is the physical phenomena that we are interested in quantifying. We will call the collection of data the measurement data. Often the measurement is *isomorphic.* Isomorphic sensing is the concept that a sensor's measurement data resembles the signal-of-interest. Isomorhic sensing is called traditional sensing. In isomorphic sensing the analog hardware, analog-to-digital converter (ADC), and processing algorithms are all seperate components, see Figure 1.1. *Computational sensing* is the concept that a joint design of the sensor, often though *coding* of the analog signal, with inversion algorithms can exceed the performance of an isomorphic sensor. Computational sensors exist at the intersection of this processes, see Figure 1.2 [1]. While isomorphic sensors can provide flexible sensing in multiple applications, a computational sensor's joint design can lead to performance increases. Throughout this chapter and the rest of this dissertation, we will provide many examples that highlight the differences between computational and isomorphic sensing.

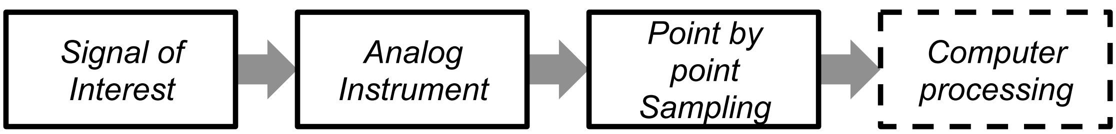

Figure 1.1: A systems view of a traditional sensing scheme. The signal-of-interest is incident upon the analog instrument. The analog instrument forms an isomorphism of the signal which is then periodically sampled point-by-point through an ADC device. Once the signal is in digital form, post-processing algorithms are often used to perform various tasks such as noise reduction, detection, and classification. Notice that the analog instrument, sampling scheme, and processing are all seperated.

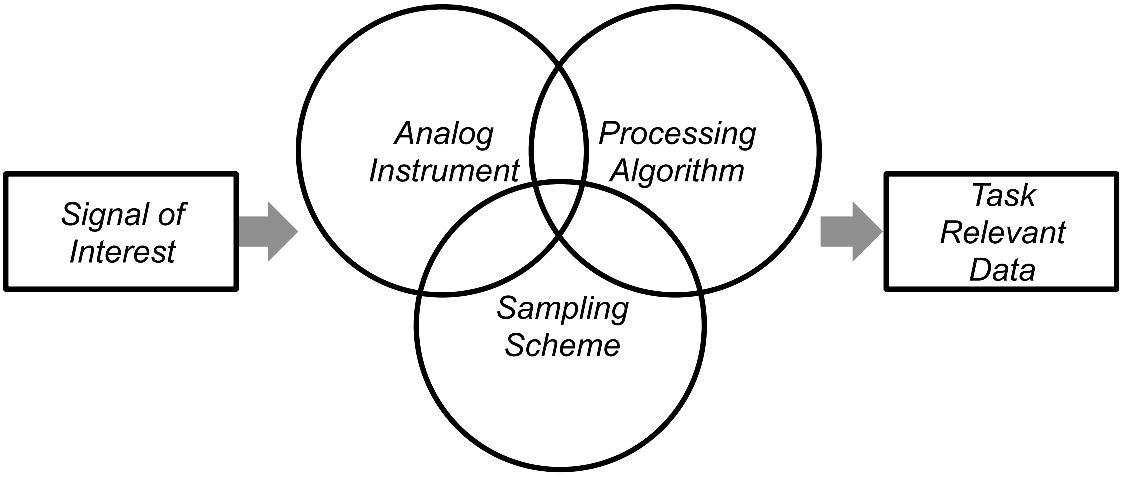

Figure 1.2: A systems view of a computational sensing scheme. The signal-of-interest is incident upon the analog instrument.

Rather than a rigorous discussion, this chapter will discuss some of the major developments and concepts in the field of computational sensing on an intuitive level. This will familiarize the reader with important terminology and techniques common in the field of computational sensing. A rigorous discussion with mathematical formalism of the concepts is presented in chapter 2. This chapter will also discuss some of the challenges I and many other experimentalists and engineers have faced when developing computational sensing prototypes. I will close this chapter with a brief look ahead to the rest of the dissertation.

1.1 Isomorphic Sensing

In Greek, the word isomorphic loosely translates to equal in form. Traditional sensors perform isomorphic sensing. In the context of this dissertation, an isomorphic sensor is any sensor which attempts to produces measurement data that resembles the signal-of-interest. In this paradigm, the analog instrument, sampling scheme, and post-processing algorithms are seperate components and processes.

We will discuss three important examples of isomorphic sensors: the pinhole camera, the photographic camera and the optical spectrometer (which I will just call a spectrometer from now on, even though there are many instruments called spectromers that not concerned with optical spectra). These sensors have had major roles throughout the history of optics and in the physical sciences, so it is natural to use them as examples of computational optical sensing. Therefore it is important we first understand the isomorphic version of these sensors.

In the photographic camera, the signal-of-interest is the object's intensity distribution. This can be the scatter or emitted light from a person, a tree or a distant group of stars. The analog instrument consists of the lenses which are designed and fabricated to produce an intensity disibution (optical image) that looks like the object at the focal-plane array (FPA). The more that the image resembles the object the better the optics. The FPA then samples and quantizes the image and produces a digital representation of the object's intensity distribution, the measurement data. If one is interested in performing a task such as detection or classification, the measurement data is often is post-processed to perform those tasks.

There are two major sub-systems in the photographic camera which determine how well it performs: the optics and the FPA. Ideally, the optics (the analog instrument in this case) will produce a point spread function (PSF) which is infinitely small in diameter. For example, in a task such as the detection of a star from several neighboring stars in the night sky, if the PSF is much larger than the center to center seperation of the two stars in the optical image, it will be quite difficult to detect. A careful reader will note that this is the same argument used by Lord Raleigh in proposing his resolution criterion [2]. Even if the PSF is small enough, the FPA must sample at a fine enough pixel-to-pixel spacing, called the *pixel pitch*, to accurately reproduce the intensity variations at the scale which is pertinant to the task. Intuitively, this makes sense because if the image of both stars and the decreased intensity which signifies a certain amount of seperation between the two stars are imaged onto a single pixel, the one cannot ever hope to be able to accurately the detect the star without some other prior or side information.

The pinhole camera predates the photographic camera over a thousand years. It consists of a small hole and a box which prevents any light except from the pinhole to enter, see Figure 1.3. The pinhole camera is useful for imaging in parts of the electromagnetic spectrum and particles for which there is no direct analog to refractive lens or reflective mirror.

In the spectrometer, the signal of interest is the spectrum of the object. The optics are designed to take the incoming light and seperate various wavelength components, see Figure 1.4. The part of the spectrometer which is used to physically isolate the wavelengths is called a *spectrograph*. The result is a spectral intensity as a function of position at the FPA. The FPA and post-processing algorithms are

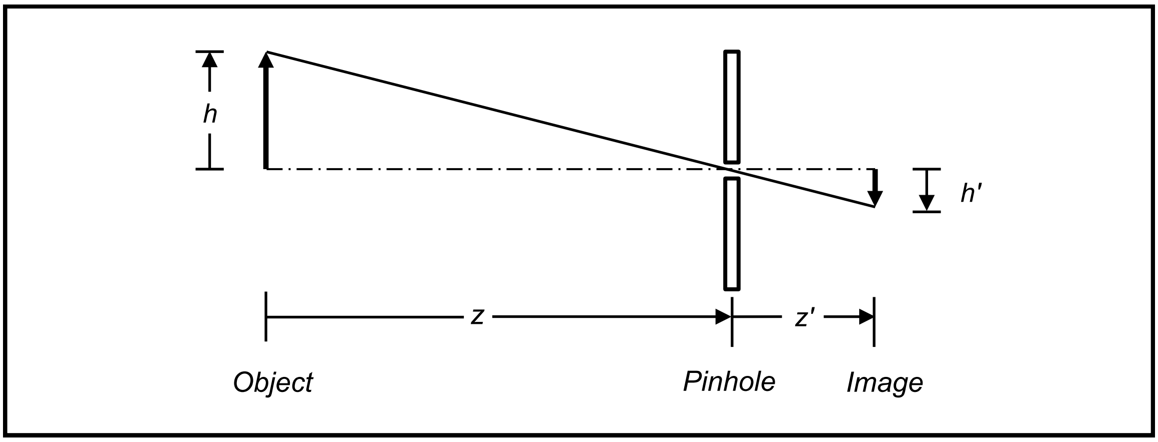

Figure 1.3: A pinhole camera.

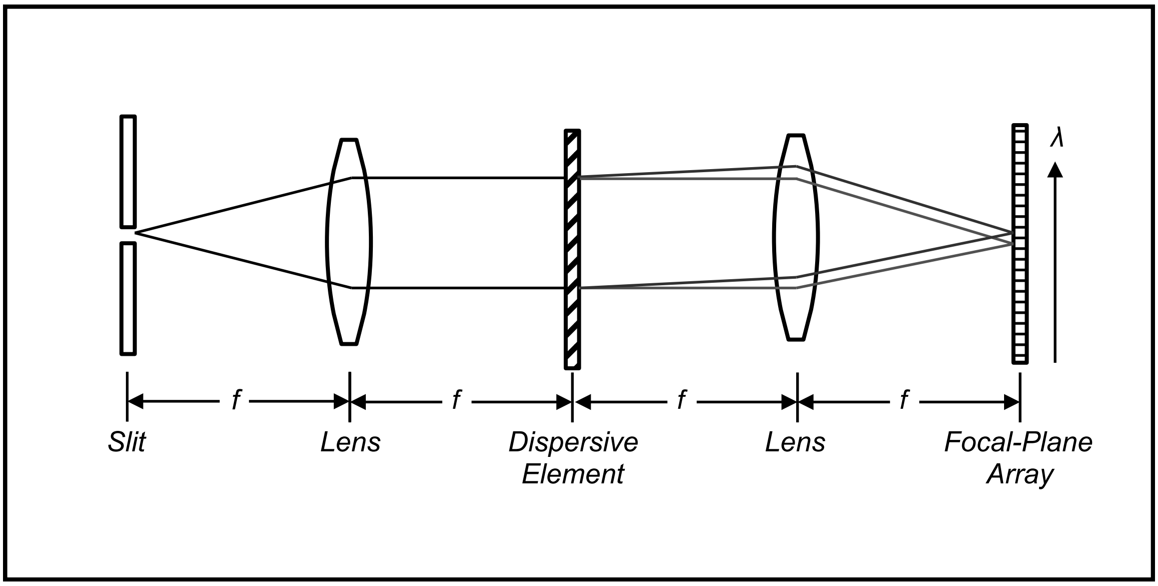

Figure 1.4: An isomorphic slit spectrometer with a 4F configuration.

used in the same manner as the photographic camera, which is to sample the optical spectrum creating a digital version of it and to perform various tasks on the measurement data. For now, we will concentrate on the slit spectrometer, which measures the spectrum at a single point on the object.

In the spectrometer, one of the important performance metrics is *spectral resolution*, which we denote $\delta\lambda$. The spectral resolution is the smallest difference in wavelength the instrument can discern. Large spectral resolutions can degrade the spectrometers ability to discern important parts of the spectrum. Similarly with the camera, the FPA must have a pixel pitch which is small enough in order to correctly sample the variations in the spectrum.

The point-by-point nature of isomorphic sensing is both a strength and a source of weakness.

The strength comes from the straightforward and intuitive architecture of the isomorphic sensor. Each subsystem: the optics, the focal-plane array (FPA), and the post-processing can be designed and constructed seperately as long as they meet their individual specifications. As long as the signal-to-noise ratio (SNR) is sufficient and the sampling rate is high enough, we are guaranteed to recover the signal.

One of the weaknessess of the isomorphic approach is the the ability to measure low SNR signals. Because the signal-of-interest is sampled in a completely parallel fashion at each exposure, each pixel contributes a certain amount of noise. If the noise dominates, the measurement fidelity decreases often forcing the operator to increase the exposure time. For weak signals, the exposure time can become prohibitive and for temporally dynamic signals this may lead to a loss of resolution. Indeed, one of the major engineering trade-offs faced by traditional spectrometer designers is that when one attempts to increase the light collection (increased slit-width) the spectral resolution $\delta\lambda$ degrades. Similarly, in the pinhole camera, there is a throughput versus spatial resolution trade-off, increasing the size of the pinhole degrades the PSF.

It would be easy to assume that with the recent revolution in machine learning and statistical signal processing combined with the dramatic increase in computing power that we could simply post-process poor measurments and obtain useful data. However, this isn't possible due to the an important theorem in information theory called the *data processing inequality* [3]. In layman's terms, it means "garbage in, garbage out".

Another weakness of isomorhic sensing is that the seperation of the analog instrument, the sampling scheme, and the data processing algorithms lead to increased size, weight and power-cost (SWAP-C). As we mentioned in the photographic camera, the optics must be designed to produce a small PSF. For demanding applications, the optical design and fabrication can be the most expensive component of the sensor. While FPA prices in the visible have fallen, FPAs in certain parts of the electromagnetic spectrum can be quite expensive or non-existant [4, 5].

In many cases, the signal is redundant and high resolution sampling becomes a waste of resources, such as data storage and communications bandwidth. A good example is in photography where often the post-processing takes the digital image and applies a compression algorithm which looks for patterns in the signal and reduces the file size, discarding much of the sample data [6].

The isomorphic sensor has served humanity well, however with all the weakness that I have discussed, I will now begin to discuss some of major techniques in computational sensing that can be used to address some or all of the issues that I just stated.

1.2 Development of Multiplexing in Sensing

Multiplexing in sensing allows each measurement sample to be a combination of multiple points of the signal-of-interest. Multiplexing is a powerful tool that can be exploited by the sensor designer to eliminate or relax design trade-offs.

A simple example which illustrates the usefulness of multiplex sensing is weighing objects. In this example, we are given a 100 sheets of paper. Let's assume that the scale's measurement error is insignificant. Isomorphic measurement sensing means one would need to measure each sheet of paper individually. Requiring 100 measurements.

Now, let's say the scale's measurement error is on the order of the weight of the sheet. Measuring each sheet invidiually produces a large measurement error. In order to reduce the error to an acceptable SNR we need to make several measurements per sheet to reduce the error to an acceptable level.

However, we can be a little smarter. We can measure all 100 sheets at the same time. Since the weight of all 100 sheets is much larger than the measurement error of the scale, we can dramatically increase the precision of the measurement. If we can assume that each sheet is the same weight, then we are done.

The weighing problem is analogous to the spectroscopy example. As discussed earlier in section 1.1, there is trade-off between light collection and spectral resolution. Increasing the slit-width to increase the amount of light has the effect of degrading the spectral resolution $\delta\lambda$. Around the late 1940's and early 1950's, several important papers and inventions demonstrated the effectiveness of multiplexing in spectroscopy. At the time the FPA was non-existant, so in the slit spectrometer shown in Figure 1.4, where the FPA is pictured, there was actually another slit. To record the intensity at each spectral channel, either the dispersive element or the exit slit had to be mechanically translated, making the measurements even slower by a factor of N_λ, the number of spectral channels of interest.

Golay was the first to propose multiplexing the slit spectrometer by creating a pattern of binary (1's and 0's) entrance and exit slits [7]. The idea borrowed heavily from communications theory which is concerned with the reliable transmission of information over a noisey channel. In the Golay multi-slit spectrometer, the patterns of entrance and exit slits are matched based on mathematically useful properties, this is similar to coding and decoding signals in commucations. In communications theory the process of structuring the data from the source to the receiver is referred to as *coding*. Similarly, in computational sensing, the transmission of information between a object signal-of-interest and the sensor is considered a coding problem [8]. In the multi-slit spectrometer, the entrance slits act to code the spectrum while the exit slits decoded the coded spectrum. Intuitively, the ability to use multiple entrance and exit slits increases the optical throughput of the spectrometer. Golay's idea dramatically increased the optical throughput without degrading the spectral resolution.

Fellgett devised an alternative approach to multiplex spectroscopy, the Fourier Transform spectrometer, and was the first to note that multiplex measurements compared to isomorphic measurements improve the signal-to-noise ratio on the order of $\sqrt{N_\lambda}$ [9]. He is generally credited with discovering the multiplex advantage, so it is often called the Fellgett advantage instead.

Another example that is pertinant to this dissertation is coded aperture imaging. Coded aperture imaging can be thought of as the multiplexed version of a pinhole camera. As mentioned earlier in section 1.1, there is a trade-off between the throughput and spatial resolution. However in many fields, such as high-energy particle imaging, refractive lenses and reflective mirrors are non-existant or underdeveloped. By using multiple pinholes the throughput is increased without sacraficing spatial resolution. However, the pattern of the pinholes, which is the code, in this case must be carefully designed in order to the reconstruction to be feasible. Fenimore, Canon, and Gottesman were the first to create an elegant solution to coded aperture design called uniformly redundant arrays [10, 11].

In summary, multiplexing has the ability to eliminate classic trade-offs in isomoprhic sensors: signal strength or resolution. Modern researchers are still actively developing novel ways to implement multiplexing to increase resolution and senstivity in the spatial domain [12, 13], spectral domain [14, 15], and temporal domain [16, 17]. However, multiplexing is not without its own set of challenges. As we mentioned, the coding must often be designed to obtain feasible signal reconstruction. We now discuss inverse problems in computational sensing.

1.3 Forward Models and Inverse Problems

In the computational sensing community, a model explains the mapping of the signal-of-interest to the measurement data is called the *forward model.* The problem of taking the observed data and calculating a reconstruction of the signal-of-interst or task-specific parameters is called the *inverse problem.*

As you can imagine, solving inverse problems of isomorphic measurements, when one is concerned with reconstruction of the signal-of-interest, tend to be straight forward. In the weighing problem, the measurement is also the reconstruction. In the slit spectromer, where the forward model can be simply the continuous to discrete mapping of the spectrum. The spectrum is the interpolated measurement.

Of course, we can begin to add various levels of complexity to the forward model to account for various physical aspects of the sensor, such as the fact the FPA can't measure certain wavelength regions or the noise in our measurements. But again, assuming proper sampling and enough SNR, the reconstruction of the isomorphic signal is the measurement. This simplicity is one reason why isomorphic sensing still dominates at the consumer level despite all of the drawbacks I discussed earlier in section 1.1.

However, the multiplexing of signal information forces us to develop computational steps to solving the inverse problem. In the multiplexed weighing problem, a significant complication occurs when each sheet of paper has a different weight. Now solving the inverse problem is not as straight forward. A single measurement of all 100 sheets in this case is an *underdetermined problem* since we have 1 equation and 100 unknowns. What we can do is try measuring different combinations of the 100 sheets, each new combination provides us with a new equation to work with reducing the error. Niavely, we might assume that we can randomly choose 100 unique combinations and solve 100 equations using the algebra we were taught in high school. This works fine when there is no measurement error. However, in the presense of noise, in many applications including the weighing problem, random combinations are not the best way to conduct the coding. They are sub-optimal in terms of reconstruction error. This lead many to begin working on optimal coding strategies of signals for sensing and is major topic in this disseration.

In fact, the idea of multiplexing was not invented by Golay or Fenimore, what they contibuted was the creation of effective ways to code and decode their signal-of-interest, making their sensors more appealling to a broader community of scientists and engineers. A good example is the Hadamard matrix based code [18]. The Hadamard code is appealling because reconstruction is simply the transpose of the original matrix.

In summary, the forward model of a sensor is essentially accounting for the physics which govern the measurement. While the solving the inverse problem is a mathematical problem which attempts to either reconstruct the object or to calculate task-specific data from the measurement data. Unfortunately, not all multiplexing forward models codes have mathematically elegent inversion steps. Often the physics of the situation force non-isomorphic measurements which require a computational step to solve the inverse problem.

1.4 Indirect Imaging

While Golay, Fennimore, and others were levering multiplexing to eliminate tradeoffs in traditional sensors, an entirely disparate group of researchers were working on imaging techniques for which there was no isomorphic analog. In these cases the physics of the sensing modality prevents a point-by-point sampling of the signal-of-interest. Indirect imaging refer to sensing schemes which include X-ray Computed Tomography (CT), Single-Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI) and certain forms of sonic and radio wave imaging all require a data-processing or reconstruction step to solve an inverse problem [19].

Perhaps one of the most successful early examples of indirect imaging and the rise of inverse problems in sensing is the development of radar. While early radar was concerned with the detection and distance of an object, development of imaging radar began after World War II. Imaging radar and specifically Synthetic Aperture Radar (SAR) can use time delay information combined with the doppler effects and interference patterns of coherent radio waves to create high resolutions of terrain and buildings.

In medicine, a common imaging modality is X-Ray CT. In X-Ray CT, computational inversion is required to reconstruct a 2 or 3-dimensional function from 1 or 2-dimensional measurement data. The forward model can be simple: In a collimated beam architecture with a 1-dimensional detector array, we say that each sample from each pixel on the array is proportional to the total number of x-ray photons that have not been absorbed by the object [20] plus noise. The inversion of course is not straight forward. The culmination of the work related to the inversion techniques and the actual prototype resulted in the Nobel Prize for Physiology or Medicine in 1979 [21].

Indirect imaging is a subfield of computational sensing. Due to the medical or military applications of these computational sensors, there has been an intense push to reduce measurement time and improve task-specific and reconstruction results. Many of the techniques from other subfields of computational sensing have been brought to bear for indirect imaging [22, 23].

We have discussed the development of multiplex sensing and indirect imaging and how the ideas from both subfields are analogous in terms of producing a non-isomorphic measurement. However a major step in practical implementation of computational sensing is being able to obtain the measurements in a quick, reliable and efficient manner. Computational sensing as a field would not exist without the most important invention in optics and photonics of the 20th century.

1.5 The Digital Imaging Revolution

In 1969 Boyle and Smith invented the Charge-Coupled Device (CCD) [24]. The CCD is the first integrated circuit device which using a 2-dimensional arrangement of pixels which could reliably convert an intensity distribution to a digital signal. The CCD is a type of FPA.

The CCD was a major breakthrough for entire fields and industries who depended on the reliable sampling, storage and transmission of optical signals. Until then one either had to use film or bulky tubes that required an electron beam to be scanned across an image scene, such as the Image orthicon [25]. For their invention, Boyle and Smith both received the Nobel Prize in Physics in 2009 [26]

The invention of the digital camera by Sasson followed shortly after [27]. Several years later the first digital spectrometer was invented. The exit slit was replacing by the CCD, whiched allowed for instant and simultaenous measurement of the entire spectrum in a compact architecture [28].

The development of the Complementary Metal–Oxide–Semiconductor (CMOS) FPA was also important. While in scientific settings, it could not rival the quality of the CCD, its cheaper cost brought digital imaging to the consumer level. Other technology like the digital computers and computer networking also provided major contributions to the democratization of imaging and optical sensing. While scientific

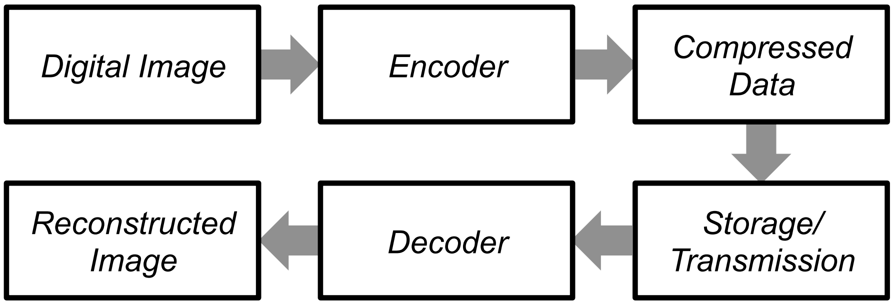

Figure 1.5: An general flowchart for image and data compression techniques.

grade optical instrumentation was and is still expensive, the researcher could at least capture, process, and share measurement data with significantly less effort. Without it, the field of computational sensing would not exist.

Algorithms for efficient and reliable storage and transmission of digital images became more important. Over time the pixel count continued to increase and the sheer volume of digtal image and video data being generated and transmitted over networks began to outpaced improvements in storage and transmission capacity. While many engineers developed new technology to combat the hardware limitations of storage and transmission. This also led to a renewed effort by researchers to develop more efficient image and video compression algorithms [29, 30].

While a multitude of compression techniques exist, they all follow the same basic process, see Figure 1.5. Once the CCD samples the optical signal and produes the measurement data, the encoder usings the compression algorithm to look for redundancies in the data and produce a lower dimensional representation of the image. The compressed data can then either be stored or transmitted or both. The decoder solves the inverse problem of reconstructing the image.

As mentioned in at the beginning of this chapter, computational sensing lies at the intersection of the design of the analog hardware, sampling schemes, and processing algorithms. Many of the algorithms used in computational sensing are the same as or inspired by the techniques used by the image processing community. This is because a major effort of computational sensing is the desire to make more resource efficient measurements.

1.6 Compressive Sensing

Traditionally, in order to increase the resolution of a sensor, one had to increase the number of measurements. This means that the SWAP-C must also increase. A camera with just a few megapixels FPA costs less than one with hundreds of megapixels. The cost of designing the optics will also need to scale to provide enough optical resolution. In a perfect world, we could capture all the information we need from just a few measurements.

Conventional signal processing dictates that accurate reconstruction of the signal-of-interest is highly improbable. If we have a discrete signal we need at least as many measurements as there are signal elements to solve the inverse problem. If the number of measurements is less, then the inverse problem is underdeterimined. Fortunately, a signal acquisition technique called *compressive sensing* allows us to design sensors that solve these types of highly underdetermined inverse problems.

As discussed earlier, much of the data being generated by sensors are redundant. Images, spectra, video, and audio data of real-world signals tend to exhibit patterns or redundancies that can be exploited. This allows a compression algorithm to significantly reduce the amount of data needed to represent the signal.

There is a class of compression algorithms called *lossy* [31]. In lossy compression, not only are redundancies exploited but data that is deemed insignificant to the signal quality is discarded. Only the most important part of the signal is kept as part of the compressed representation of the original signal. When the signal is uncompressed, the amount of data is less that the original measured data. The difference in quality is often unnoticeable to a human observer. In both lossless and lossy compression, the goal is to obtain a *sparse* representation of the signal. A sparse representation means that the signal can be well approximated with only a few non-zero elements in a representation basis. A representation basis is a basis in which the signal-of-interest is sparse. For example, most natural images are sparse in the Fourier basis. The representation basis is typically not the native basis of the signal-of-interest, i.e. pixel number or spectral channel.

In 2006, David Donoho argued that traditional sensors tend to produce vast amounts of measurement data, but often the majority of data is redundant and discarded in the compression step. He proposed that sensors can be designed to directly measure the most relavent data in a signal, suggesting a measurement scheme that can measure a compressed form of the signal [32]. This is the idea behind *com-*

pressive sensing sometimes known as *compressive sampling*. If the measurements are compressive then it should be possible to significantly reduce the number of measurements to accurately reconstruct the signal.

Note that there is a subtle but powerful distinction between compressive sensing and the traditional sensing and then compressing. The difference between compressive sensing and the traditional approach is that traditional compression algorithms operate as a post-processing step. Therefore, a traditional compression algorithm will have access to the entire signal to look for redundancies and convert it into a sparse representation. In compressive sensing, we do the compression directly and therefore we do not have access to the entire uncompressed signal. The algorithms must assume that the signal has a sparse representation.

The question of how to actually measure or code the analog signal to directly obtain compressed data is also important. Fortunately, random coding tends to work well in many instances when the signal has a sparse representation. Much of the work in this dissertation will discuss other types of coding schemes that can be used to outperform random codes.

The idea of compressive sensing seems to be similar to the concept of multiplex sensing. However, there is an important distinction to be made. In compressive sensing, the aim is to obtain the relavent information in as few measurements as possible. In multiplexing, the goal is to overcome limitations mainly due to lack of SNR. Many compressive sensing schemes also employ multiplexing.

One useful example of compressive sensing versus traditional sensing is the single pixel camera [12]. The single pixel camera is a multiplexing camera architecture that uses time sequential random measurements and recovers the image in significantly less measurements (=number of exposures $\times$ pixels) than the conventional camera. Another example is the Coded Aperture Snapshot Spectral Imaging (CASSI) architecture [33], which can reconstruct a spectral data cube in significantly less FPA exposures than a traditional spectral imaging architecture.

Another important distinction is between reconstruction and task-specific sensing. Task-specific sensing tends to refer to measurement techniques that attempt to directly perform tasks such as detection, classification, and estimation without the intermiediate step of reconstructing the high-dimensional signal. Compressive sensing is useful not just of overcoming resolution limitations in reconstruction but for reducing measurement resources for task-specific sensing. For example, in facial recognition the goal is detection of an individual person. Reconstruction of the face image is simply an intermediate step, therefore, one can develop a compressive sensing scheme that is optimal for direct facial detection, skipping the step of image reconstruction [34].

Multiplex sensing, compressive sensing, and task-specific sensing are considered subfields of computational sensing. Computational sensors offer significant advantages that allow us to overcome classic engineering trade-offs in sensor design. However, computational sensing has its own unique set of engineering problems that we will now discuss.

1.7 Practical Considerations in Computational Sensing

Decoding the signal, which we can also think of as solving the inverse problem is not necessarily straight forward. Unfortunately the coded of the analog signal and inversion steps are often seperately designed in computational sensing. This becomes especially frequency because the practicing sensor engineer must piece together various techniques from theorestist. This

1.8 Dissertation Overview

CHAPTER 2

Formalism

This chapter introduces the reader to the more rigorous concepts and mathematical background that will required to fully understand the material presented in the later chapters of this dissertation.

A rigorous discussion of multiplexing and signal-to-noise ratio will be discussed, as well are various coding schemes used in various notable computational sensors as well as the ones in this disseration.

Since the AFSSI-C relies on a variation of Principal Component Analysis (PCA) and a Bayesian algorithm for coding design we will discuss some of the fundamental of Bayesian probability and the Log-Likelihood Ratios.

The measurement process implements a mapping from the

2.1 Isomorphic Sensing

$$\mathbf{g} = \mathbf{H}\mathbf{r} \tag{2.1}$$

where $\mathbf{g}$ is the measurement vector.

2.2 Multiplexing

2.3 Principal Component Analysis

2.4 Bayesian Rules and Log-Likelihood Ratios

2.5 Compressive Sensing

2.5.1 The Nyquist-Shannon Sampling Theorem

The Nyquist-Shannon Sampling Theorem states one must sample a signal with a sampling rate that is at least twice the maximum frequency of the signal to prevent aliasing [35].

2.5.2 Sparsity, Incoherence, and the Restricted Isometry Property

At first glance, compressive sensing is unintuitive. Going back to the weighing example, if we have hundreds of itmes we can group the items together to increase the precision of our measurements. However, we still have to take as many measurements as their are items, to solve for their weight. If the number of measurements is less than the number of objects, then a unique solution is impossible. If we have some sort of helpful information, we can significantly reduce the number of solutions, discarding away solutions that are inconsistent with this prior information.

In compressive sensing, this helpful knowledge is called *sparsity*. A signal is sparse if only a few elements of the signal is non-zero. For compressive sensing to work, the signal-of-interst itself doesn't need to be sparse. if one only needs to describe it with a few representation vectors relative to the native dimensionality of the signal. Note that the sign Sparsity is quite similar to how compressiable a signal is. Most real world signals aren't sparse, they are compressible. However, we know from lossy compression techinques that we can throw away most of the data, that much of the information is not sparse. The reconstruction algorithms rely on the assumption of sparsity to correctly estimate the original signal from the small number of measurements [].

The actual question of how to code the analog signal-of-interest to produce the most compressive measurements as possible is still an active area of research []. Many in the compressive sensing community refer to the codes as the measurement basis. This interpretation of the coding scheme allows us to think of the measurement data as projections of the signal-of-interst onto a set of vectors. Incoherence is the concept that the measurement basis and the represetation basis should have low correlation with each other. The more incoherent—lower correlation—the two bases are, the higher the probability of succesful reconstruction of compressive measurements. Random matrices have a high probability of being incoherent with any basis [36], which is why they are so popular in experimental prototypes [] since it simplified the task of having to design the measurement basis. However, as we will see random measurements do not always lead to the best reconstruction results since they are agnostic to the signal-of-interest. The choice of the measurement basis can be a difficult task depending on the application, however it has

oherence, However, mathematicians have shown that the measurement basis should have a low maximum correlation with the representation basis, the basis in which the object signal has a sparse representation. This allows the signal to get

scheme relies of the coherence of the representation basis and the measurement basis. Up till now the representation basis has been the canoncial Dirac basis.

Quite surpriseling, Other than sparsity, when other knowledge of the signal can be made often random measurements are the optimal measurements.

2.5.3 Inversion

2.5.3.1 Least Squares

Suppose

$$\mathbf{g} = \mathbf{H}\mathbf{r} \tag{2.2}$$

Given $\mathbf{g}$ and $\mathbf{H}$ we want to solve for $\mathbf{r}$. If the matrix is full rank then we can simply multiply both sides of equation 2.2 by $\mathbf{H}^{-1}$

$$\mathbf{H}^{-1}\mathbf{g} = \mathbf{H}^{-1}\mathbf{H}\mathbf{r} = \mathbf{I}\mathbf{r} = \mathbf{r} \tag{2.3}$$

If $\mathbf{H}$ is not full rank then its inverse does not exist. However we can try to find a solution $\hat{\mathbf{r}}$ that minimizes the least squared error. This is called the *Least Squares Solution* also known as the *Least Squares Estimator*, *Ordinary Least Squares* and by many other names. We define the squared error as

$$\|\mathbf{e}\|^2 = \|\mathbf{H}\mathbf{r} - \mathbf{g}\|^2 \tag{2.4}$$

To minimize the error, we take the derivative of equation 2.4 with respect to $\mathbf{r}$ and set it equal to zero and solve for $\mathbf{r}$. The full derivation which shows each step is given in Appendix A. The least squares estimate:

$$\hat{\mathbf{r}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{g} \tag{2.5}$$

2.5.3.2 L0 and L1 Norm Minimization

2.5.3.3 LASSO and sparsity regularization

$\mathfrak{F}\{\iint\limits_a^b g(\xi,\eta)h(x-\xi,y-\eta)d\xi d\eta\}$

APPENDIX A

Derivation of the Least Squares Estimator

Suppose

$$\mathbf{g} = \mathbf{H}\mathbf{r} \tag{A.1}$$

Given $\mathbf{g}$ and $\mathbf{H}$ we want to solve for $\mathbf{r}$. If the matrix is full rank then we can simply multiply both sides of equation A.1 by $\mathbf{H}^{-1}$

$$\mathbf{H}^{-1}\mathbf{g} = \mathbf{H}^{-1}\mathbf{H}\mathbf{r} = \mathbf{I}\mathbf{r} = \mathbf{r} \tag{A.2}$$

If $\mathbf{H}$ is not full rank then its inverse does not exist. However we can try to find a solution $\hat{\mathbf{r}}$ that minimizes the least squared error. This is called the *Least Squares Solution* also known as the *Least Squares Estimator*, *Ordinary Least Squares* and by many other names. We define the squared error as

$$\|\mathbf{e}\|^2 = \|\mathbf{H}\mathbf{r} - \mathbf{g}\|^2 \tag{A.3}$$

To minimize the error, we take the derivative of equation A.3 with respect to $\mathbf{r}$ and set it equal to zero and solve for $\mathbf{r}$. Note that the equation A.3 can be expanded in terms to a inner product

$$\|\mathbf{e}\|^2 = \|\mathbf{H}\mathbf{r} - \mathbf{g}\|^2 = \sum_{i=1}^{N} e_i^2 = \mathbf{e}^T\mathbf{e} = (\mathbf{H}\mathbf{r} - \mathbf{g})^T(\mathbf{H}\mathbf{r} - \mathbf{g}) \tag{A.4}$$

The transpose is distributive

$$(\mathbf{H}\mathbf{r} - \mathbf{g})^T = (\mathbf{H}\mathbf{r})^T - \mathbf{g}^T \tag{A.5}$$

The transpose of a product of matrices equals the product of their transposes in reverse order

$$(\mathbf{H}\mathbf{r})^T = \mathbf{r}^T\mathbf{H}^T \tag{A.6}$$

So equation A.4 becomes

$$\begin{aligned} \|\mathbf{e}\|^2 &= (\mathbf{r}^T\mathbf{H}^T - \mathbf{g}^T)(\mathbf{H}\mathbf{r} - \mathbf{g}) \\ &= \mathbf{r}^T\mathbf{H}^T\mathbf{H}\mathbf{r} - \mathbf{r}^T\mathbf{H}^T\mathbf{g} - \mathbf{g}^T\mathbf{H}\mathbf{r} + \mathbf{g}^T\mathbf{g} \end{aligned} \tag{A.7}$$

We can see that the two middle terms $\mathbf{r}^T\mathbf{H}^T\mathbf{r} = \mathbf{g}^T\mathbf{H}\mathbf{r}$ because they are just scalars.

$$\|\mathbf{e}\|^2 = \mathbf{r}^T\mathbf{H}^T\mathbf{H}\mathbf{r} - 2\mathbf{g}^T\mathbf{H}\mathbf{r} + \mathbf{g}^T\mathbf{g} \tag{A.8}$$

To find the least squares solution, take the gradient with respect to $\mathbf{r}$ and set it equal to zero.

It should be noted that there are two different notations for writing the derivative of a vector with respect to a vector $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$. If the numerator $\mathbf{y}$ is of size m and the denominator $\mathbf{x}$ of size n, then the result can be laid out as either an $m \times n$ matrix or $n \times m$ matrix, i.e. the elements of $\mathbf{y}$ laid out in columns and the elements of $\mathbf{x}$ laid out in rows, or vice versa. They are both correct and equal, which leads to confusion when switching back in forth. I will write both to reduce confusion.

Clearly the gradient of the third term in equation A.8 w.r.t $\mathbf{r}$ is 0, so it goes away. We first tackle the first term on the right hand side in equation A.8

$$\frac{\partial}{\partial \mathbf{r}}\mathbf{r}^T\mathbf{H}^T\mathbf{H}\mathbf{r} \tag{A.9}$$

Let $\mathbf{K} = \mathbf{H}^T\mathbf{H}$. Since $\mathbf{K}$ is symmetric, we can use the identity

$$\frac{\partial}{\partial \mathbf{r}}\mathbf{r}^T\mathbf{K}\mathbf{r} = 2\mathbf{r}^T\mathbf{K} = 2\mathbf{K}^T\mathbf{r} \tag{A.10}$$

since $\mathbf{K} = \mathbf{K}^T$ then

$$\frac{\partial}{\partial \mathbf{r}}\mathbf{r}^T\mathbf{H}^T\mathbf{H}\mathbf{r} = 2\mathbf{r}^T\mathbf{H}^T\mathbf{H} = 2\mathbf{H}^T\mathbf{H}\mathbf{r} \tag{A.11}$$

and the gradient of the middle term in equation A.8 is simply $-2\mathbf{H}^T\mathbf{g}$ so

$$\frac{\partial}{\partial \mathbf{r}}\|\mathbf{e}\|^2 = 2\mathbf{H}^T\mathbf{H}\mathbf{r} - 2\mathbf{g}^T\mathbf{H} \tag{A.12}$$

setting it equal to zero and solving for $\mathbf{r}$ gives the least squares estimate

$$\hat{\mathbf{r}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{g} \tag{A.13}$$

Acronyms

ADC analog-to-digital converter. 8, 11

AFSSI-C Adaptive Feature Specific Spectral Imaging-Classifier. 4, 5, 25

CASSI Coded Aperture Snapshot Spectral Imaging. 23

CCD Charge-Coupled Device. 20, 21

CMOS Complementary Metal–Oxide–Semiconductor. 20

CT Computed Tomography. 19

DISP Duke Imaging and Spectroscopy Program. 5

FPA focal-plane array. 13–16, 18, 20, 22, 23

LCOS Liquid Crystal on Silicon. 5

LENS Laboratory for Engineering Non-Traditional Sensors. 4, 5

MRI Magnetic Resonance Imaging. 19

PET Positron Emission Tomography. 19

PSF point spread function. 13, 15

SAR Synthetic Aperture Radar. 19

SCOUT Static Computational Optical Undersampled Tracker. 5

SLM Spatial Light Modulator. 5

SNR signal-to-noise ratio. 15, 16, 18, 23

SPECT Single-Photon Emission Computed Tomography. 19

SWAP-C size, weight and power-cost. 15, 22

Symbols

$\delta\lambda$ Spectral Resolution. 14–16

$\mathbf{g}$ The Measurement Vector. 25

N_λ The Number of Spectral Channels. 16

REFERENCES

[1] M. A. Neifeld, A. Mahalanobis, and D. J. Brady, "Task-specific sensing-introduction," *Appl. Opt.*, vol. 45, no. 13, pp. 2857–2858, May 2006. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-45-13-2857

[2] L. Rayleigh, "Investigations in optics, with special reference to the spectroscope," *The London, Edinburgh, and Dublin Philosophical Magasize and Journal of Science*, 1879.

[3] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

[4] C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, and W. J. Padilla, "Terahertz compressive imaging with metamaterial spatial light modulators," *Nature Photonics*, vol. 8, no. 8, pp. 605–609, 2014.

[5] I. Noor, O. Furxhi, and E. L. Jacobs, "Compressive sensing for a sub-millimeter-wave single pixel imager," in *SPIE Defense, Security, and Sensing.* International Society for Optics and Photonics, 2011, pp. 80 220K–80 220K.

[6] D. Taubman and M. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards and Practice.* Springer Science & Business Media, 2012, vol. 642.

[7] M. J. Golay, "Multi-slit spectrometry," *JOSA*, vol. 39, no. 6, pp. 437–444, 1949.

[8] D. J. Brady, *Optical imaging and spectroscopy.* John Wiley & Sons, 2009.

[9] P. Fellgett, "I.—les principes généraux des méthodes nouvelles en spectroscopie interférentielle-a propos de la théorie du spectromètre interférentiel multiplex," *J. phys. radium*, vol. 19, no. 3, pp. 187–191, 1958.

[10] E. E. Fenimore and T. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Applied optics*, vol. 17, no. 3, pp. 337–347, 1978.

[11] S. R. Gottesman and E. Fenimore, "New family of binary arrays for coded aperture imaging," *Applied optics*, vol. 28, no. 20, pp. 4344–4352, 1989.

[12] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. E. Kelly, R. G. Baraniuk *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, p. 83, 2008.

[13] D. Townsend, P. Poon, S. Wehrwein, T. Osman, A. Mariano, E. Vera, M. Stenner, and M. Gehm, "Static compressive tracking," *Optics express*, vol. 20, no. 19, pp. 21 160–21 172, 2012.

[14] M. E. Gehm, S. T. McCain, N. P. Pitsianis, D. J. Brady, P. Potuluri, and M. E. Sullivan, "Static two-dimensional aperture coding for multimodal, multiplex spectroscopy," *Applied optics*, vol. 45, no. 13, pp. 2965–2974, 2006.

[15] T.-H. Tsai and D. J. Brady, "Coded aperture snapshot spectral polarization imaging," *Applied optics*, vol. 52, no. 10, pp. 2153–2161, 2013.

[16] J. Holloway, A. C. Sankaranarayanan, A. Veeraraghavan, and S. Tambe, "Flutter shutter video camera for compressive sensing of videos," in *Computational Photography (ICCP), 2012 IEEE International Conference on.* IEEE, 2012, pp. 1–9.

[17] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," *Optics express*, vol. 21, no. 9, pp. 10 526–10 545, 2013.

[18] N. Sloane, T. Fine, P. Phillips, and M. Harwit, "Codes for multiplex spectrometry," *Applied optics*, vol. 8, no. 10, pp. 2103–2106, 1969.

[19] H. H. Barrett and K. J. Myers, *Foundations of Image Science.* John Wiley & Sons, 2013.

[20] J. Radon, "1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten," *Classic papers in modern diagnostic radiology*, vol. 5, 2005.

[21] "The Nobel Prize in Physiology or Medicine, 1979," https://www.nobelprize.org/nobel_prizes/medicine/laureates/1979/perspectives.html, accessed: 2016-08-22.

[22] X. X. Zhu and R. Bamler, "Tomographic sar inversion by-norm regularization—the compressive sensing approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, 2010.

[23] C. Chen and J. Huang, "Compressive sensing mri with wavelet tree sparsity," in *Advances in neural information processing systems*, 2012, pp. 1115–1123.

[24] W. S. Boyle and G. E. Smith, "Charge coupled semiconductor devices," *Bell System Technical Journal*, vol. 49, no. 4, pp. 587–593, 1970.

[25] H. W, N. L, and S. Joe, "Image orthicon," Feb. 11 1975, uS Patent 3,866,078. [Online]. Available: https://www.google.com/patents/US3866078

[26] "The Nobel Prize in Physics, 2009," http://www.nobelprize.org/nobel_prizes/physics/laureates/2009/press.html, accessed: 2016-08-24.

[27] J. Estrom, "Kodak's First Digital Moment," http://lens.blogs.nytimes.com/2015/08/12/kodaks-first-digital-moment/, August 12 2015, accessed: 2016-08-24.

[28] K. L. Moore, "Spectrometer with electronic readout," Mar. 27 1979, uS Patent 4,146,332.

[29] H. Kobayashi and L. R. Bahl, "Image data compression by predictive coding i: Prediction algorithms," *IBM Journal of Research and Development*, vol. 18, no. 2, pp. 164–171, 1974.

[30] J. Ziv and A. Lempel, “Compression of individual sequences via variable-rate coding,” *IEEE transactions on Information Theory*, vol. 24, no. 5, pp. 530–536, 1978.

[31] B. E. Usevitch, “A tutorial on modern lossy wavelet image compression: foundations of jpeg 2000,” *IEEE signal processing magazine*, vol. 18, no. 5, pp. 22–35, 2001.

[32] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[33] A. Wagadarikar, R. John, R. Willett, and D. Brady, “Single disperser design for coded aperture snapshot spectral imaging,” *Applied optics*, vol. 47, no. 10, pp. B44–B51, 2008.

[34] H. S. Pal, D. Ganotra, and M. A. Neifeld, “Face recognition by using feature-specific imaging,” *Applied optics*, vol. 44, no. 18, pp. 3784–3794, 2005.

[35] C. E. Shannon, “Communication in the presence of noise,” *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[36] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.