Title
Physical Optics Based Computational Imaging Systems

Permalink
https://escholarship.org/uc/item/1765j3mh

Author
Olivas, Stephen Joseph

Publication Date
2015

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Physical Optics Based Computational Imaging Systems

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in

Electrical Engineering (Photonics)

by

Stephen Joseph Olivas

Committee in charge:

Professor Joseph E. Ford, Chair
Serge Belongie
Jules Jaffe
Jan Kleissl
Truong Nguyen
George Papen

2015
The dissertation of Stephen Joseph Olivas is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015
DEDICATION

This dissertation is dedicated to my family.
# TABLE OF CONTENTS

Signature Page ....................................................... iii
Dedication ............................................................... iv
Table of Contents .................................................... v
List of Figures ........................................................... vii
List of Tables ........................................................... xii
Acknowledgements ...................................................... xiii
Vita .......................................................................... xvi
Abstract of the Dissertation .......................................... xviii

I Introduction ........................................................... 1

II Imaging Camera Fundamentals and Trade-offs ...................... 4
II Bibliography ........................................................... 16

III Inverse Problems .................................................... 17
III Bibliography ........................................................... 30

IV Image processing for cameras with fiber bundle image relay .......... 31
A Introduction: Traditional Imagers and Fiber-Coupled Imagers ... 31
B Component and System Point Spread Function ....................... 34
C Global Artifact Removal Calibration of the Fiber-Coupled Sensor and Prototype ................................................................. 41
D Spherical to Planar Mapping, Image Alignment, and Stitching Together Image Seams ................................................................. 47
E Experimental Setup .................................................... 49
F Discussion and Conclusion ............................................ 53
IV Bibliography ........................................................... 55

V Digital Image Processing for Wide-Angle Highly Spatially-Variant Imagers ................................................................. 58
A Introduction - Traditional and Fiber Coupled Imagers ............... 58
B Experimental Setup .................................................... 59
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Spatially Variant Extended Depth-of-Focus Algorithm</td>
<td>65</td>
</tr>
<tr>
<td>D</td>
<td>Discussion &amp; Conclusion</td>
<td>67</td>
</tr>
<tr>
<td>V</td>
<td>Bibliography</td>
<td>69</td>
</tr>
<tr>
<td>VI</td>
<td>Platform Motion Blur Image Restoration System</td>
<td>70</td>
</tr>
<tr>
<td>A</td>
<td>Introduction</td>
<td>70</td>
</tr>
<tr>
<td>B</td>
<td>Measuring Image Motion Using PSDs</td>
<td>72</td>
</tr>
<tr>
<td>C</td>
<td>PSD Image Motion Tracking Simulation</td>
<td>74</td>
</tr>
<tr>
<td>D</td>
<td>Generating a Spatially Variant PSF Using PSDs</td>
<td>75</td>
</tr>
<tr>
<td>E</td>
<td>System Prototype Design</td>
<td>78</td>
</tr>
<tr>
<td>F</td>
<td>Experimental Setup</td>
<td>80</td>
</tr>
<tr>
<td>G</td>
<td>Image Deblurring: Prototype Results</td>
<td>81</td>
</tr>
<tr>
<td>H</td>
<td>Future Directions</td>
<td>88</td>
</tr>
<tr>
<td>I</td>
<td>Conclusion</td>
<td>90</td>
</tr>
<tr>
<td>J</td>
<td>Appendix: Calibration - Mapping of Voltage to Pixels</td>
<td>91</td>
</tr>
<tr>
<td>VI</td>
<td>Bibliography</td>
<td>93</td>
</tr>
<tr>
<td>VII</td>
<td>Characterization of a Compressive Imaging System</td>
<td>97</td>
</tr>
<tr>
<td>A</td>
<td>Introduction</td>
<td>97</td>
</tr>
<tr>
<td>B</td>
<td>CI and Image Sampling Theory</td>
<td>98</td>
</tr>
<tr>
<td>C</td>
<td>System Design and Configuration</td>
<td>100</td>
</tr>
<tr>
<td>D</td>
<td>DMD™ Time Division Multiplexing</td>
<td>104</td>
</tr>
<tr>
<td>E</td>
<td>Experimental Setup</td>
<td>106</td>
</tr>
<tr>
<td>F</td>
<td>Experimental Results</td>
<td>110</td>
</tr>
<tr>
<td>G</td>
<td>Conclusion</td>
<td>117</td>
</tr>
<tr>
<td>VII</td>
<td>Bibliography</td>
<td>120</td>
</tr>
<tr>
<td>VIII</td>
<td>Conclusion</td>
<td>123</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure IV.1:</th>
<th>(a) A conventional lens forms an aberrated planar image (the prototype is inset to show relative size). The Zebase design F_005 has been re-optimized to achieve f = 12 mm, F/1.7 and 120° field-of-view. (b) Prototype: a monocentric.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure IV.2:</td>
<td>(a) Monocentric lens (MC) comprised of a symmetric 2-glass achromat (top) and meniscus lens (bottom). (b) A fiber bundle relays a business card logo from its flat to its curved surface. (c) A fiber bundle coupled to.</td>
</tr>
<tr>
<td>Figure IV.3:</td>
<td>(a) The fiber bundle structure and defects lead to image artifacts. Fiber absorbers between every other 2.5 µm fiber intersection prevent crosstalk. (b) An OVT5653 sensor is used to sample the image exiting the.</td>
</tr>
<tr>
<td>Figure IV.4:</td>
<td>(a) Side-by-side size comparison of the prototype with an F/4 commercial camera of similar field-of-view. (b) The F/1.35, 30 megapixel, 126° field-of-view fiber-coupled monocentric lens imager prototype.</td>
</tr>
<tr>
<td>Figure IV.5:</td>
<td>(a) Schematic and (b) cross-section showing the internal structure and scale of the fiber-coupled sensor. The fiber bundle relays light incident at its surface to the CMOS image sensor though an epoxy bonding layer,</td>
</tr>
<tr>
<td>Figure IV.6:</td>
<td>Schematic of the impulse response characterization setup where an LED, pinhole, and collimation lens (f = 360 mm, F/# = 5.14, Ø = 76mm) form a collimated point source. The setup in (a) uses a high resolution microscope.</td>
</tr>
<tr>
<td>Figure IV.7:</td>
<td>Data cube representation of the impulse response of various DUTs corresponding to the bottom row of Fig. IV.8 collected using the setup shown in Fig. IV.6(b). The x and y axes correspond to the image.</td>
</tr>
<tr>
<td>Figure IV.8:</td>
<td>Impulse characterization of system components where each subfigure contains (from top to bottom) the integration of the 200 images captured during a point source scan, a projection of the data onto translation-pixel space.</td>
</tr>
</tbody>
</table>
Figure IV.9: Two typical flat-field calibration images (left) and their histograms (right) captured using the fiber bundle relayed imager. The lowest intensities transmitted by the bundle are above 175 (65%) (a) on-axis and ........................................ 43

Figure IV.10: A calibration image of a flat white scene (a) shows artifacts due to broken fibers, scratches, and other imperfections as well as the fiber bundle’s internal structure, all of which lead to obscuration and color artifacts (b) ........................................ 44

Figure IV.11: (a) A flat-field calibration image of a uniform white scene shows artifacts due to broken fibers, scratches, and other imperfections as well as the fiber bundle’s internal structure, all of which lead to artifacts in the acquired ......................... 46

Figure IV.12: Image processing data flow. .................................................. 47

Figure IV.13: Schematic showing the fiber-coupled sensors in the prototype and their respective field-of-view. The six sensors gather image data across a 126° field-of-view. ................................. 48

Figure IV.14: Distortion correction and image stitching. ...................... 50

Figure IV.15: An image taken using a conventional wide field-of-view camera (Canon 5D mark II with EF 8-15mm F/4 fisheye lens) with similar characteristics as the fiber-coupled monocentric lens imager. Magnified regions show the image resolution. .............. 51

Figure IV.16: Prototype image showing significant image improvement due to distortion correction, image stitching, and interpolation to form panorama. Artifacts due to bubbles in the optical adhesive, broken fibers, and fiber bundle scratches are ............... 52

Figure V.1: Schematic of experimental setup and EDOF imaging system. .. 59
Figure V.2: EDOF imaging system and experimental lab setup. .............. 60
Figure V.3: Image taken with lens focused to 1 m. Five targets are placed in the scene at depths of 0.65 m, 0.83 m, 1 m, 1.33 m, and 2 m orthogonal to the optical axis. Magnified views show the limited depth-of-focus. ......................................................... 60

Figure V.4: Single exposure focal swept images of a point source placed at 1 m orthogonal along the optical axis for various field angles, and placed at 0° and 60° for a set of distances orthogonal to the optical axis. This shows the PSF dependence ............... 61

Figure V.5: The spatially variant PSF changes depending on object distance and angle (note PSF scale). Many exposures were taken as the EDOF system’s focus was incrementally stepped. Profiles of the captured images as a function of focal position .......... 62
Figure V.6: Single exposure focal swept images and PSFs. The focus is swept from 0.5m to infinite conjugate. Five targets are placed in the scene at depths of 0.65 m, 0.83 m, 1 m, 1.33 m, and 2 m orthogonal to the optical axis. Magnified views.

Figure V.7: EDOF image processed using the spatially variant PSF data and focal swept imagery in Fig. V.6. Magnified views from the various seen depths show the depth-of-focus of the processed image.

Figure VI.1: Platform motion deblur system schematic. A lens and beamsplitter form identical image planes for the image sensor and position sensing detector (PSD) array to provide data for image restoration.

Figure VI.2: Lateral photoeffect of a PSD. Differential currents from input (blue) and output (red) electrodes produce sum (VI.2) and position (VI.1) data.

Figure VI.3: Scenes input into PSD image motion tracking simulation to characterize the effects of background illumination. (a) Ground truth input image. (b) SV blurred image. (c) SV blurred image with Gaussian noise ($\sigma = 0.1$) with a positive mean.

Figure VI.4: Simulation results: Plot of PSD tracking error for various levels of background illumination showing the sensor is only capable of tracking image features much brighter than the background. Percentages represent the ratio between.

Figure VI.5: PSD centroid tracking of identical image movement is skewed toward the sensor’s center as background illumination increases. (a) Motion estimate from Fig. VI.3(b) with no background illumination and (b) from Fig. VI.3(c) with background.

Figure VI.6: PSD data at specific locations in the image field after calibration (Appendix J). These local PSF measurements are used to construct an SV PSF. Their relative location is given in Figs. VI.10 and VI.9.

Figure VI.7: PSF specific to pixel $(x, y)$ is recursively generated using affine model coefficients $a(t) - f(t)$ as the elementary time step $t$ is integrated over the sampling time.

Figure VI.8: Photo of the imaging, position sensing, motion actuated system.

Figure VI.9: Experimentally blurred color image. Magnified sections of the image confirm SV blur. The calculated SV PSF is shown superimposed (yellow lines) and is consistent with the image motion blur. The PSD’s relative locations are.
Figure VI.10: Experimentally blurred starfield image (contrast reversed). The calculated SV PSF is shown superimposed in green and is consistent with the SV image motion blur. The PSD’s relative locations are shown in cyan. The portion of...

Figure VI.11: Pixel-by-pixel deconvolution. A PSF is generated for the blurred pixel of interest (green). A section in the neighborhood of the pixel is deblurred using the PSF. Only the center pixel (yellow) of the deblurred section is used in the final reconstruction...

Figure VI.12: The region enclosed in red of the blurred image in Fig. VI.9 is deblurred using (c) bilinearly interpolated total variation regularization, (d) pixel-by-pixel Lucy-Richardson techniques, and (e) bilinearly interpolated Lucy-Richardson.

Figure VI.13: The region of Fig. VI.10 enclosed in red is deblurred using a pixel-by-pixel method with the iterative spatially invariant Lucy-Richardson algorithm. The (a) ground truth, (b) blurred, and (c) five-iteration reconstructed...

Figure VI.14: A PSD can track the moving centroid of features that remains on the sensor under low background illumination. They provide position (VI.1) and intensity (VI.2) outputs, which can reveal invalid data. (a) A PSD array provides...

Figure VI.15: (a) Depiction of a PSD and imager’s physical location in the image plane. PSD voltage and imager pixel information is paired at two separate image locations and used to form calibration vectors. (b) Depiction showing vectors...

Figure VII.1: Schematic of compressive imaging system. A scene is imaged onto a digital micromirror device (DMD™), which serves as a spatial light modulator (SLM). A sequence of specific basis patterns are emulated by the DMD™ mirrors...

Figure VII.2: Photo of experimental compressive imaging system. A scene is imaged onto the DMD™ which directs a portion of the imaged scene to a condenser lens, which then directs the optical energy to a single photo-detector. A focusing scope helps...

Figure VII.3: Grayscale representation through the use of time division multiplexing of the DMD™. An 8 bit PNG image loaded into the D4100 controller board is converted into eight 1 bit image files, each of which is displayed for a different time duration...

Figure VII.4: Two methods of sampling the CI system patterns are compared in order to improve 8 bit system performance. To perform the comparison, 256 PNG files are...

Figure VII.5: Sixteen of the basis patterns depicting the (b) Noiselet (c) Hadamard and (d) DCT basis transform sets. The colorbar is shown in (a).
Figure VII.6: Ground truth images taken with a Canon 5D Mark II DSLR camera. The scenes include (a) and (b) back-illuminated chrome on glass transparencies in a lab setting, and (c) a daylight illuminated building. ................................. 109

Figure VII.7: Spectral data of the outdoor image scene Fig. VII.6(c) taken using an Ocean Optics Spectrometer with and without color filters in place. ........................................... 110

Figure VII.8: Experimentally acquired (a), (d) Noiselet, (b), (e) Hadamard, and (c), (f) DCT complete datasets from lab-imaged, back-illuminated chrome on glass transparencies of a 1951 USAF resolution target (a), (b), and (c) and a grayscale Lena .... 111

Figure VII.9: (a), (b), (c), (d) Noiselet, (e), (f), (g), (h) Hadamard, and (i), (j), (k), (l) DCT reconstructed images of a 1951 USAF resolution target using 0.1, 1, 10, 100% of samples recorded in a lab setting. Error is calculated with reference to the .... 112

Figure VII.10: (a), (b), (c) Noiselet, (d), (e), (f) Hadamard, and (g), (h), (i) DCT reconstructions of a grayscale image using 1%, 10%, 100% of samples recorded in a lab setting. Error is calculated with reference to the respective 100% reconstruction. .... 113

Figure VII.11: Plot of the RMSD error versus the number of measurements used in the reconstruction of the two lab scenes using the three basis sets. Measurements are taken from Fig. VII.9 (solid line) and Fig. VII.10 (dashed line) and are normalized to the .... 114

Figure VII.12: Reconstructed images with the red, green, and blue bandpass filters were combined to form outdoor color images. Reconstructions were made using 1% of the (a) Noiselet, (b) Hadamard, and (c) DCT transforms. Daylight changed .... 115

Figure VII.13: Infrared reconstructed images using 1% of samples recorded outdoors using an IR bandpass filter. ................. 115

Figure VII.14: Comparison of the compressive imager with a conventional camera. (a) Images taken by a Canon 5D Mark II where the pixel data has been binned to 1% of its original 1 megapixel size, namely 10,404 pixels. (b) Shows the images in .......... 116
### LIST OF TABLES

Table IV.1: Statistics taken over the same white part of the images in Fig. IV.10 before and after occlusion removal show the process significantly improves image uniformity. 45  

Table VI.1: On-Trak PSD (PSM2-4) specifications 74  
Table VI.2: Time and memory requirements of the described algorithms 86  

Table VII.1: Texas Instruments high definition 0.95 1080p DMD™ Specifications 102  
Table VII.2: Speed and memory requirements of encoding algorithms for acquiring 1% of the samples (10,486 basis patterns and coefficients) of the corresponding 1 megapixel image. Basis load time is the time it takes to load the transform 106
ACKNOWLEDGEMENTS

I thank my research advisor, Professor Joseph Ford, for his expertise, and guidance throughout the course of my time at UCSD both in classroom and in his Photonics Systems Integration Laboratory (PSILAB). In addition, many thanks go out to members of the PSILAB with whom I have worked and shared ideas with for the past years, namely Eric Tremblay, Jason Karp, Kate Baker, Justin Hallas, Brett Nadler, Igor Stamenov, Ashkan Arianpour, Nojan Motamedi, William Mellette, Glenn Schuster, Salman Karbasi, and Ilya Agurok. I would like to thank Peter Ilinykh for leading the academic lab coursework I participated in and later helped teach, and for taking extra time to help support to the Photonics Society at UCSD student organization activities. Thanks go out to Yaron Rachlin, Lydia Gu and Brian Gardiner for the time and attention they provided to ensure our projects we successful. I would like to thank Nima Nikzad for taking time to contribute in the system integration of our experimental setups. Special thanks to Michal Šorel for extending his expertise, guidance, and support over the course of this thesis work.

Thanks also go to JP Laine, and Robin Dawson at the Charles Stark Draper Laboratory, Rick Morrison, Ron Stack and Adam Johnson at the Distant Focus Corporation whose skills and expertise in packaging, coding, electronics and prototyping made the demonstration cameras a reality. Many special thanks to the optics and photonics professors who provided me a path to gather lifelong skills while completing coursework at UCSD. I would also like to include the UCSD faculty who introduced me broader topics in the field of engineering. Thanks to all my teachers throughout the years who taught me some of what they know so that I could build upon, question and contribute to these ideas myself.

Thanks goes out to my family for being supportive and encouraging throughout the trajectory that ultimately led to this completed thesis.

Thank you all!
The text of Chapter IV is work in which the dissertation author was the primary investigator and author of publications. It is, in full, a reprint of the material as it appears in:


The text of Chapter V is work in which the dissertation author was the primary investigator and author of publications. It is, in full, a reprint of the material as it appears in:


The text of Chapter VI is work in which the dissertation author was the primary investigator and author of publications. It is, in full, a reprint of the material as it appears in:


Stephen J. Olivas, Michal Šorel, Nima Nikzad, and J. Ford, “Platform Mo-
tion Blur Image Restoration System”, *OSAs Computational Optical Sensing and Imaging Conference* CTu2B.3 (2012).

The text of Chapter VII is work in which the dissertation author was the primary investigator and author of publications. It is, in full, a reprint of the material as it appears in:


Research documented in Chapters IV and V of this dissertation was supported by the DARPA SCENICC program under contract W911NF-11-C-0210. Research contained in Chapters VI and VII of this dissertation was supported by the Charles Stark Draper Laboratory University Research and Development Program, contract number SC001-000000628A. Research conducted by Michal Šorel contained in VI and V was supported in part by the Czech Ministry of the Interior, Grant No. VG20102013064.
VITA

2008 Bachelor of Science
Electrical Engineering (Photonics),
University of California, San Diego

2011 Master of Science
Electrical Engineering (Photonics),
University of California, San Diego

2015 Doctor of Philosophy
Electrical Engineering (Photonics),
University of California, San Diego

PUBLICATIONS


ABSTRACT OF THE DISSERTATION

Physical Optics Based Computational Imaging Systems

by

Stephen Joseph Olivas

Doctor of Philosophy in Electrical Engineering (Photonics)

University of California, San Diego, 2015

Professor Joseph E. Ford, Chair

There is an ongoing demand on behalf of the consumer, medical and military industries to make lighter weight, higher resolution, wider field-of-view and extended depth-of-focus cameras. This leads to design trade-offs between performance and cost, be it size, weight, power, or expense. This has brought attention to finding new ways to extend the design space while adhering to cost constraints. Extending the functionality of an imager in order to achieve extraordinary performance is a common theme of computational imaging, a field of study which uses additional hardware along with tailored algorithms to formulate and solve inverse problems in imaging. This dissertation details four specific systems within this emerging field: a Fiber Bundle Relayed Imaging System, an Extended Depth-of-Focus Imaging System, a Platform Motion Blur Image Restoration System, and a
Compressive Imaging System. The Fiber Bundle Relayed Imaging System is part of a larger project, where the work presented in this thesis was to use image processing techniques to mitigate problems inherent to fiber bundle image relay and then, form high-resolution wide field-of-view panoramas captured from multiple sensors within a custom state-of-the-art imager. The Extended Depth-of-Focus System goals were to characterize the angular and depth dependence of the PSF of a focal swept imager in order to increase the acceptably focused imaged scene depth. The goal of the Platform Motion Blur Image Restoration System was to build a system that can capture a high signal-to-noise ratio (SNR), long-exposure image which is inherently blurred while at the same time capturing motion data using additional optical sensors in order to deblur the degraded images. Lastly, the objective of the Compressive Imager was to design and build a system functionally similar to the Single Pixel Camera and use it to test new sampling methods for image generation and to characterize it against a traditional camera. These computational imaging systems share a common theme in that they seek to accomplish camera designs that meet more demanding system requirements through the use of additional measurements made possible by hardware modifications, while relying on modeling and computational methods in order to provide valuable scene information.
Chapter I

Introduction

Digital sensors have replaced film in almost every imaging application, where relatively more user friendly digital image processing and storage occur in place of film’s chemical processing and physical storage. Digital cameras were originally designed to simply mimic their film camera counterparts and have been improved for the most part by increasing sensor linearity and pixel count. This method of improvement has reached many of its fundamental limits and for this reason, this approach has recently begun to be re-evaluated. In many fields, there is a need for further improved resolution, focus, dynamic range, field-of-view, and light collection while at the same time keeping cost, power consumption, size and weight at a minimum. Exploring and taking full advantage of digital image processing has only recently been a topic of in-depth investigation as computational capabilities have become more sophisticated. More fundamental changes in the way a scene is sensed, provides data which can be processed to produce images that would otherwise be unachievable. This emerging field of Computational Imaging uses the additional sensed information provided through the use of clever system illumination, optics, and sensor modification in order to formulate inverse problems that are used to computationally process improved imagery outside the capabilities of traditional cameras. Hardware and software within the computational imager is designed simultaneously to work interdependently as a single end-to-end system in order to achieve a desired system performance. Each step within the imaging pipeline is leveraged to produce the best end result, even if the intermediate results
are worse by traditional imaging standards or metrics.

A brief history of camera technology development, an overview of lens imaging and the fundamental trade-off’s in imaging functionality are presented in Chapter II. Chapter III discusses the formulation and solution of the inverse problems frequently used to process data captured by Computational Imagers. These chapters conclude the background material useful in explaining Computational Imaging systems. This is followed by chapters dedicated to specific imaging systems used to illustrate how Computational Imaging can be used to extend the functionality of traditional cameras. A discussion on the hardware incorporated into, and used to assist each Computational Imager in the sensing process is contained in each respective chapter.

Chapter IV presents a Fiber Bundle Relayed Imaging System which incorporates a fiber optic bundle to relay non-planar image surfaces to planar detectors. This frees lens designers from the constraint of having the lenses form a planar image surface, allowing for better performing and more innovative designs. Characterization of the end-to-end systems allows one to computationally remove degradations due to the fiber optic bundle. Chapter V presents an Extended Depth-of-Focus Imaging System which uses a translation stage to capture a swept focus image which can be processed to have a greater depth-of-focus than would otherwise be possible. Chapter VI presents a Platform Motion Blur Image Restoration System which incorporates Position Sensing Detectors into a standard camera in order to estimate a spatially variant Point Spread Function (PSF) in order to deblur images. This modification allows for longer integration (exposure) times that would otherwise not be feasible, allowing more versatility in camera operation. Chapter VII presents a Compressive Imaging System in which patterned portions of the scene are projected by a Spatial Light Modulator onto a single photo-diode, where the signal is integrated and sensed. The stored data, which is a set of linear projections of the scene, is then processed to form an image. This system allows very few samples, much less than the pixel count, to be used to form high resolution images. In addition to more efficient sensing, this system configuration allows for task specific sampling of the image space for applications
such as object identification.

In this thesis, the imaging systems are presented in order of how well they fall within the field of Computational Imaging. For example, many of the techniques presented in Chapter IV are standard in image processing; however, incorporating these techniques into the overall system design and characterization methodology qualify it as computational imaging, even though the unprocessed data is already stored in the form of images. Chapter VII, on the other hand, presents a system in which an image must be computationally extracted from the sensed information, and is therefore well within the field of computational imaging. The systems presented in this thesis show how: incorporating the transfer function of the end-to-end system, measuring additional information, or sensing the scene information differently can assist in improving image quality and extend the functional operating space of imaging systems. Although the systems are very different, they share a common theme. A theme which seeks to best extract scene information and use it to formulate a better conditioned inverse problem from which to computationally solve for images that are superior, according to some metric, than those produced by traditional cameras.
Chapter II

Imaging Camera Fundamentals and Trade-offs

Photography is one of the most important inventions of the 19th century which arose from cameras of the era, such as the Daguerreotype camera. Up until then, documentation of important events or scientific data was recorded in drawings, paintings, sculptures and the like. Photography allowed people in other locations and at later times to view and interpret these events and discoveries firsthand. The improved fidelity facilitated more accurate reevaluation of data, which in turn lead to more findings. A summary historical description of the technological developments leading to the traditional camera is presented followed by an overview of its functionality and limitations.

The camera obscura or pinhole camera is a basic imaging system in which an enclosure with a small pinhole sized aperture allows light to enter and form an image. The pinhole projects (maps) an image of an exterior scene onto the interior surface of the enclosure, which can then be viewed, or recorded. This type of imager maps each ray incident at a particular angle to a single point in the image and therefore boasts a large depth of field; however, the angular selectivity of the small aperture greatly limits the light collected by the camera. For this reason, the pinhole camera could be used to image bright objects (i.e. solar eclipses) and if the viewer’s eyes were adjusted to low light, other scenes could be viewed as well. By the time of the Renaissance, a lens had replaced the pinhole within the
camera, allowing it to collect and focus more light emanating from the scene. The increase in light collection enabled users to view the images without having to accommodate their eyes to the low light levels produced by the pinhole camera. This made the camera more practical and eventually became a drawing aid, where artists could trace images onto paper. Using lenses in cameras has drawbacks in the added size and weight of the glass or plastic materials, spherical, coma and chromatic aberrations, limited field of view, limited depth of field, as well as their cost in design and manufacturing. Additional mechanics are needed to accurately position the lens with respect to the image plane in order to focus a scene at a particular distance. Yet, the role of lenses is unrivaled in the design of modern cameras and many computational imagers.

The first recorded references to the phenomena of imaging with a camera obscura date back to the 5th century BCE China and 4th century BCE Greece [[1]]. The pinhole camera more formally investigated in “The Book of Optics” [[2]], which was finalized around 1021 CE by Ibn al-Haytham (Alhazen), the “Father of Modern Optics”. The Nimrud lens dates to 8th century BCE Assyria but it is not clear if it was used to concentrate or image light [[3]]. Other references alluding to magnification lenses date back to hieroglyphs of 8th century BCE Egypt, and Pliny the Elder of 1st century CE Rome [[4], [5]]. Eyeglasses (lenses used for magnification) were commonly used by the late 13th century CE Europe [[4]]. Camera lens designers of the 18th and 19th centuries revolutionized the lens, producing exceptional designs tailored for specific purposes, such as the macro lens, wide angle lens, and zoom lens. Recording images with photosensitive chemicals, such as silver chloride, began in the late 18th century but the chemicals were unstable and could not store images for long. Capturing images on film requires a shutter to control of how much light is allowed to expose the film. In the early 19th century, Joseph Nicèphore Nièce used a solution of bitumen and lavender oil to create a permanent film process. Louis Jacques Mandè Daguerre who worked with Nièce announced the first camera (the Daguerreotype) which recorded images onto metal plates. William Henry Fox Talbot shortly thereafter introduced paper and transparent recording mediums as well as negatives from which to make im-
age duplicates. In 1861, James Clarke Maxwell used color filters to capture three images and used them together to make the first (additive) color image. In 1888, George Eastman, founder of the Eastman Kodak Company, introduced the Kodak camera which used film rolls and in 1935 Kodak introduced Kodachrome film, the hugely successful color film transparency. Meanwhile flash technology had been evolving, and in the mid 18th century the magnesium flash was used, and in 1899 a electric trigger was invented. In 1929 flash bulbs were invented, and in 1931 Harold Edgerton invented the first electronic flash bulb. Since the mid 20th century, the chemistry used to produce long lasting high quality film has come to maturity.

In 1975 a semiconductor image sensor was used to replace the film in a camera creating the first analog semiconductor camera. The digital camera was first introduced in the early 1990’s. Digital sensors have allowed the users to modify and share images in quantities unimaginable in the years prior. Since then these image sensors and supporting electronics have become the standard and for this reason image-space telecentric lenses are now more preferred since sensors are more angularly selective than film. Liquid Crystal Display’s (LCD) and compact flash memory storage were first incorporated into digital cameras in the mid 1990’s.

Today virtually all cameras are digital and include infrared/ultraviolet filters, anti-reflection coated lenses with aberration correction, semiconductor optical sensors with color filters, flash illumination, electronic memory storage, and an LCD display. This design also incorporates image processing such as sensor response calibration, color interpolation, compression, and image stabilization. Additionally, many cameras have automatic ISO, aperture, shutter speed and color balance which use image data, or simple sensors to measure ambient light, to help the camera produce a pleasing image with less noise degradation. Some cameras have mechanical image stabilization that uses moving lenses to correct for image motion due to camera shake. A schematic depicting the internal components of a typical camera’s image processing pipeline (with the exception of flash illumination) is shown in Fig. II.1. Most photographs taken today are processed as described and generally the unprocessed image sensed by the detector cannot be accessed or recovered. The vast majority of cameras made today are incorporated
Figure II.1: Schematic depicting components typically found in cameras.
into cell phones, each of which has at least one camera. The camera pipeline of state-of-the-art traditional cameras is a topic of in-depth and careful investigation. Camera pipeline simulations are often used to assist in improving the overall system operation, ensuring a balance in performance among the components [[6],[7]]. All of these advances have made cameras so affordable, easy to use, and small that today they can be found almost everywhere.

In what follows we describe basic camera operation and trade-offs in imaging performance [[8], [9]]. The imaging lens (Fig. II.2) forms an image of the scene and follows the lens equation defined as

\[
\frac{1}{f} = \frac{1}{S_{\text{object}}} + \frac{1}{S_{\text{image}}},
\]

where \( f \) is the focal length of the lens, \( S_{\text{object}} \) is the distance from the lens to the object, and \( S_{\text{image}} \) is the distance from the lens to the image. The image is magnified according to

\[
m = \frac{S_{\text{image}}}{S_{\text{object}}},
\]

where \( m \) is the magnification. The F/# of the lens, denoted as \( N \), is given by

\[
N = \frac{f}{d},
\]

and describes the light collection properties of the lens, where \( d \) is the size of the aperture. For a fixed focal length lens, lowering the F/# is achieved by increasing the aperture and therefore increasing the light gathering capabilities as well.

**Figure II.2**: A simple imaging lens forms a magnified inverted image of the scene at the image plane.
Figure II.3: The Angle of Coverage and the Field of View.

In photography, the Angle of Coverage (AOC), denoted $\alpha_{AOC}$, describes the angle range the lens maps from the object plane (Field of Coverage) to the image plane (Image Circle). The latter is typically larger than the sensor’s surface. The Angle of View (AOV), denoted $\alpha_{AOV}$, describes the angle range that can be imaged instantaneously by a camera. In the object plane this corresponds to the Field of View. This is shown schematically in Fig. II.3. The AOV is given by

$$\alpha_{FOV} = 2 \arctan \frac{D}{2f}, \quad (\text{II.4})$$

where $D$ is the size of the sensor or film, and $f$ is the effective focal length. Here a vertical or horizontal FOV can be obtained using $D$ corresponding to a vertical or horizontal sensor size, respectively. If the sensor’s surface is not filled by the image then the circular image boundary will be present in the recorded image and the FOV will be limited by the AOC. Generally for large FOV cameras, the image quality degrades off-axis due to lens aberrations even for well-designed, costly lenses. For certain systems, vignetting, a reduction in the image brightness at the image periphery, can be significant and must be taken into account.

Lenses suffer from aberrations due to material properties and geometry, the former are wavelength dependent and the latter are due to the lens shape. Axial chromatic aberration occurs when different wavelengths focus to different axial positions. Transverse chromatic aberration occurs when different wavelengths focus to different lateral positions. This is shown schematically in Fig. II.4.

Five geometric aberrations are summarized in the following text and depicted schematically in Fig. II.5. Spherical Aberration is characterized by rays
passing through different parts of the lens being focused to different axial positions. Coma is characterized by off-axis illumination focusing to different transverse positions depending on if the light travels through the center or edge of the lens. Astigmatism occurs when a set of rays entering the lens along a slit have different focal positions depending on the orientation of the slit. Curvature of Field occurs when an image is formed on a curved surface rather than a plane. Distortion occurs when a set equally spaced points in the object plane are imaged to form a set of non-equally spaced points in the image plane. Frequently this non-uniformity is constant radially from the optical axis and is characterized as barrel or pincushion distortion.

Acceptable focus is defined using the circle of confusion described as

\[ c = md \frac{|S_{object2} - S_{object}|}{S_{object2}}, \]  

which is depicted schematically in Fig. II.6. As a reminder, here \( m \) is the magnification, and \( d \) is the lens diameter. A lens focused at \( S_{object} \) correctly focuses points at this distance; however, a point at \( S_{object2} \) is imaged to a disk called the circle of confusion. The Depth of Focus (DOF) describes the range of depth in the object space that is within an acceptable focus (Fig. II.6). The near and far

Figure II.4: Axial and transverse chromatic aberrations.
Figure II.5: Schematic of geometric aberrations
Figure II.6: Circle of Confusion and Depth of Focus.

Object distance limits within the DOF are given by

\[ S_{object1} = \frac{S_{object}f^2}{f^2 + Nc(S_{object} - f)}, \quad S_{object2} = \frac{S_{object}f^2}{f^2 - Nc(S_{object} - f)}. \]  

The hyperfocal distance is the nearest focal distance, \( S_{image} \), at which the DOF encompasses infinity (\( S_{object2} < \infty \)). The hyperfocal distance given by

\[ f_{hyp} = \frac{f^2}{Nc} + f \approx \frac{f^2}{Nc}, \]  

where the approximation typically holds. Using this approximation and assuming \( S_{object} < f_{hyp} \), DOF of a camera is given by

\[ DOF = \frac{2f_{hyp}S_{object}(S_{object} - f)}{f_{hyp}^2 - (S_{object} - f)^2}. \]  

The limitation effects of DOF is depicted in the photographs contained in Fig. II.7. Aside from selecting a lens with particular properties, or focusing to a particular (possibly application specific) object distance the user can only control the aperture size and therefore there is an inherent trade-off between light gathering capability and DOF. In photography, it is at times desirable to have a shallow DOF to draw attention onto certain objects. In scientific applications it is generally favorable to have more of the scene in focus, unless for a specific purpose, for example if scene depth is being inferred from the image.
(a) Open Aperture leads to a smaller DOF

(b) Closed Aperture leads to a larger DOF

**Figure II.7:** Depth of Field as a function of aperture size.

The ISO is a measure of the sensor or film's sensitivity to light. It is a legacy metric of sensitivity based on SNR as related to film speed, the more sensitive the film the lower the SNR. Manufacturers configure the digital camera to have image sensor sensitivity control that is analogous to the legacy film standards. ISO can be described as a function of luminous exposure $H$, where

$$H = \frac{qLt}{N^2}, \quad (II.9)$$

$L$ is the luminance of the scene (in candela per m), $t$ is the exposure time (in seconds), $N$ is the numerical aperture, and $q$ is an efficiency parameter given by

$$q = \frac{\pi}{4} TV(\Theta)\cos^4(\Theta), \quad (II.10)$$

which in turn depends on the transmittance $T$ of the lens, the vignetting factor $v(\Theta)$, and the angle $\Theta$ relative to the axis of the lens. One major difference is
that the noise in film produces a granular texture that some consider aesthetic, while the noise in digital image sensors is random or fixed-pattern arising from electronics and signal processing. With the exception of film, where high ISO provides an aesthetic image quality, low ISO (low noise) images are usually more desirable especially for viewing fine details or when formulating inverse problems. The latter is due to noise amplification and is discussed in the next chapter.

Film or sensors require a certain amount of photons to adequately sense the image. This is dictated by the radiant exposure $H_e$, which follows $H_e = E_e t$, here $E_e$ is the irradiance and $t$ is the exposure time. The exposure can be increased by increasing the time the film or sensor is exposed, or by increasing the amount of light entering the camera. The latter can be achieved by either increasing the amount of light entering the camera, using a flash for example, or by opening the aperture. In many applications adding light to a scene or increasing the aperture (which as discussed previously decreases the DOF) is not desired or impossible. Increasing the ISO sensitivity allows for a shorter exposure time but increases noise which degrades the image and limits the image processing that can be done on the image. If these negative effects are unacceptable then the exposure time must be kept long, and the image might be degraded by object or camera motion.

Fortunately, there are techniques that can restore blurred images and retain the quality afforded by low ISO (noise), and ample light collection. In the case a limited DOF there are techniques to produce extended DOF imagery with the desirable image quality described above. This thesis describes specific methods to capture additional image degradation information in order to recover images with specific qualities without having to give up other desirable qualities. So far, all the imaging techniques discussed here deal with images that are acquired (and possibly degraded) by measuring the image plane with densely resolved film or a densely sampled digital sensor array. This method of measuring the image can be thought of as simultaneously taking a large number of point samples at (semi) periodic spacings withing the image plane. The recorded information is essentially already an image. In this thesis, we also discuss how reevaluating the way the image space is measured leads to new exciting capabilities in image formation and information
extraction. All this is possible, in part, through the advent of computational imaging techniques and systems built upon the increasingly capable processors. A summary of some of the ideas within this continually emerging field can be found in [10],[11]. Each Computational Imaging system described in this thesis is accompanied by a detailed explanation of the system design, components, problem formulation and experimental results, which are demonstrated and discussed within their respective chapters.

Chapter IV presents a system that uses image relaying fiber optic bundles in order to be able to incorporate a lens that forms a non-planar image. The newly compatible lens allows for increased light collection, Field of View, while reducing aberrations and system size. In general, fiber optic bundles make unconventional imaging systems possible by allowing non-standard lenses to be compatible with standard sensors. The fiber optic bundle does; however, degrade image quality which can be restored using the techniques discussed in the referenced chapter.

Chapter V uses a linear motion stage to sweep an image sensor during the integration of a single exposure in order to be able to produce an image with a larger Depth of Focus than a standard camera. In one sense, this allows the camera to use a low ISO (low noise) and a large aperture (greater light gathering capabilities) to produce images with the Depth of Focus of a small aperture lens.

Chapter VI uses additional optical sensors that track the position of bright light sources to extract localized image motion. This in turn is used to restore images that have been degraded by image motion. This configuration allows imaging systems to operate at lower ISO, and collect more signal from the scene producing less noisy images that are more immune to image motion degradation (blur).

The imager in Chapter VII samples the image plane with modulated basis patterns. The modulated signal is sequentially integrated and sensed by a single photo-diode to form a set of basis coefficients, rather than sampling the image plane directly with millions of pixels. This unconventional way of probing the scene shows great promise in extending the capabilities of cameras since measurements are tailored to the application freeing-up resources for other tasks.
II Bibliography


Chapter III

Inverse Problems

Light emanating from an object through an imaging system in order to be recorded and stored is a direct (forward) problem. Using the data produced by the imaging system to infer what object caused this particular dataset to be stored is an inverse (backward) problem. In a forward problem you are given a set of causes and asked to calculate the effect. Historically the forward problem has been studied more extensively since it is used to predict future events. In an inverse problem you are given an effect and asked to calculate the cause. There is an inherent loss of information in the forward problem due to physical system characteristics, and therefore solving the inverse problem is more challenging.

Forward problems are well-posed in that their solutions exist for any input, are unique and depend continuously on the data. Forward problems based on physical systems tend to be well-posed. Inverse problems are often ill-posed, meaning that their solution does not exist, is not unique, or does not depend continuously on the input data. The domain of objects is mapped by the transfer function of the camera system into the range of images. The loss of information in this process is due to the systems transfer function which performs an imperfect and incomplete (generally bandlimited) mapping. This information loss leads to distinct objects mapping to identical images, and dissimilar objects mapping to similar images. Noise in the sensing process extends the range of possible objects the imaging system is capable of producing, from the set noise-free images to include the set of noisy images. These describes possible reasons why inverse
problems can be ill-posed. Discretized versions of these ill-posed inverse problems can be well-posed producing unique solutions that do exist; however, they suffer from large oscillations due to noise which is amplified during the inversion by the transfer function. Division by zeros in the transfer function can severely degrade even noiseless images. This deprives these solutions of physical meaning. The key to solving inverse problems is not to solve for the exact inverse but to use physical constraints known about the solution in order to find the best approximation. These constraints can include knowledge of the energy, smoothness, positivity, or bandwidth of the solution to reduce the set of possible solutions, a process called regularization. Alternatively, statistical knowledge of the solution and the noise can be used in a Bayesian framework.

Theoretical analysis and practical solutions to inverse problems can be better understood and carried out through the use of the Fourier Transform

\[ \hat{f}(\omega) = \mathcal{F}\{f(x)\} = \int \exp(-i\omega \cdot x)f(x)dx, \quad (III.1) \]

and its inverse

\[ f(x) = \mathcal{F}^{-1}\{f(\omega)\} = \frac{1}{(2\pi)^q} \int \exp(i\omega \cdot x)f(\omega)d\omega, \quad (III.2) \]

where the q-dimensional space variable is defined as \( x = \{x_1, x_2, \ldots, x_q\} \), and its derivative as \( dx = \{dx_1, dx_2, \ldots, dx_q\} \). For monochrome two-dimensional images, \( q = 2 \). Channels of color images can be treated independently and joined after processing. Space frequencies are defined as \( \omega = \{\omega_1, \omega_2, \ldots, \omega_q\} \), and similarly its derivative as \( d\omega = \{d\omega_1, d\omega_2, \ldots, d\omega_q\} \). The dot product in the equations can be expanded as \( \omega \cdot x = \{\omega_1 x_1, \omega_2 x_2, \ldots, \omega_q x_q\} \). \( f(x) \) is the image in the spatial domain, and \( \hat{f}(\omega) \) is the image in the spatial frequency domain, also called the Fourier domain. The Fourier Transform is denoted here with the notation of \( \wedge \) or \( \mathcal{F} \). The Fourier Transform projects an image described in spatial coordinates into a two-dimensional complex plane representation described in spatial frequencies, with the lowest spatial frequencies represented nearer the origin. This is extremely useful since spatial frequency content of an image can be analyzed and scaling portions of the Fourier plane can be used to filter the corresponding spatial frequencies. The Fourier Transform is standard in computation for these reasons and
since it accelerates certain mathematical computations. An object of finite extent in the image domain is termed spacelimited, while objects of finite extent in the spatial frequency domain are referred to as bandlimited. The Wintaker-Shannon Theorem states that perfect image representation can be achieved by sampling an image at equidistant points provided the separation of the points is at most half the inverse of the highest spatial frequency contained in the image. The most efficient sampling of an image with circularly shaped bandlimited frequency content is performed using a two-dimensional hexagonal Bravais lattice of sampling points.

Imaging can be described as a convolution of the object $f$ with the point spread function (PSF) or impulse response function $K$, defined as

$$g(x) = \int K(x - x') f(x') dx',$$  \hspace{1cm} (III.3)

if $K$ is spatially variant or as

$$g(x) = \int K(x, x') f(x') dx',$$  \hspace{1cm} (III.4)

if $K$ is spatially invariant. These equations are valid under the conditions that $\mathcal{F}\{K\}$ is bounded and $\mathcal{F}\{f\}$ is square-integrable. The resulting function is also square integrable. Convolution has a short form notation denoted by $*$ as in

$$g(x) = K(x) * f(x).$$  \hspace{1cm} (III.5)

The characteristics of the imaging system (such as limits in resolution) are modeled in this forward problem as the PSF, $K$. The PSF of the optical system is the image of an impulsive (diffraction limited) point source produced by the system. The PSF is generally finite in extent, integrates to unity, and can be experimentally gathered or generated analytically. For linear space invariant systems the PSF does not change as a function of the transverse location of the point source in the object plane, and is therefore said to be space invariant. When this condition does not hold the PSF is spatially variant. The later case is more general but more challenging to operate with mathematically.

Provided the PSF is spatially invariant, an equivalent operation to convolution (which is generally computationally faster for large images) can be achieved
using convolution theorem, described as

\[ g(x) = \mathcal{F}^{-1}\{\hat{g}(\omega)\} \quad \text{(III.6)} \]

\[ = \mathcal{F}^{-1}\{\mathcal{F}\{K(x)\}\mathcal{F}\{f(x)\}\}. \quad \text{(III.7)} \]

In other words, image formation can be expressed in spatial coordinates as a convolution of the system’s PSF with the object, or in spatial frequency coordinates as a scaling of the spatial frequency content of the object with the transfer function which describes the spatial frequency transmission of the system. In the former case, the PSF can be thought of as a burring operator acting on the input points, spreading the point source image. In the latter case, the transfer function can be thought of as a filter operating on the input image’s spatial frequency content. If the transfer function is bandlimited it acts as a low pass filter, limiting the resolution of the image.

A linear and bounded convolution operator, A, can be defined in L2 space (the space of all square integrable functions) as

\[ g(x) = Af \quad \text{(III.8)} \]

\[ = K(x) \ast f(x). \quad \text{(III.9)} \]

Here the convolution operator is a representation of the PSF, K. The adjoint operator given by

\[ A^*g = K^*(x) \ast g(x), \quad \text{(III.10)} \]

has the property that

\[ (Af, g) = (f, A^*g). \quad \text{(III.11)} \]

The inverse of a convolution operator is defined as

\[ (A^{-1}g)(x) = \frac{1}{(2\pi)^2} \int \frac{\mathcal{F}\{g(\omega)\}}{\mathcal{F}\{K(\omega)\}} e^{ix \cdot \omega} d\omega, \quad \text{(III.12)} \]

and is not bounded when \( \mathcal{F}\{K(\omega)\} \) approaches zero.

An image formed and recorded by a camera undergoes, not only optical image formation degradations, described by \( K \) or \( A \), but also opto-electronic acqui-
sition degradations, $w$. A more realistic model of the imaging process is therefore

\[ g_{\text{noise}}(x) = Af + w(x) \]  

(III.13)

\[ = g(x) + w(x) \]  

(III.14)

\[ = \int K(x, x')f(x')dx' + w(x), \]  

(III.15)

or in the spatially invariant case

\[ g_{\text{noise}}(x) = Af + w(x) \]  

(III.16)

\[ = g(x) + w(x) \]  

(III.17)

\[ = \int K(x - x')f(x')dx' + w(x) \]  

(III.18)

\[ = \mathcal{F}^{-1}\{\mathcal{F}(K(\omega))\mathcal{F}(f(\omega))\} + w(\omega). \]  

(III.19)

All functions in the null space of the bandlimited operator $A$ satisfy $Af = 0$, and form the space of invisible objects (objects for which the optical imaging system cannot produce images). Similarly, the range of $A$ is the set of all functions satisfying $Af = g$, and forms the set of all noise-free images. The null space of $A$ (set of all invisible objects) is always orthogonal to the range of $A$ (set of all noise-free images). Noise, which comes as a cost of sensing the image, extends the range of images an imaging system can produce to not only to include noise-free images but also noisy-images which would otherwise be invisible.

Optical image formation degradations are generally due to lens aberrations, camera motion, focusing issues or limited resolution and are discussed at the end of this chapter. The opto-electronic image capture is degraded by noise due to the thermal behavior of the semiconductor image sensor, quantization due to the read-out electronics, and rounding errors due to analog-to-digital conversions. These are statistical processes inherent to the physical operation of the semiconductor and related electronics.

Noise tends to occur randomly, at times on the order of a pixel; however, some electronic fixed pattern degradations are deterministic. The noise can be correlated or uncorrelated, additive or multiplicative, and have a specific distribution such as Gaussian or Poissonian. Gaussian noise is generally due to thermal behavior of the sensor. Salt-and-Pepper noise is generally caused by semiconductor
or electronic errors and produces impulsive, deterministic errors. Poisson (shot) noise arises from the discrete nature of photon detection and is more significant at lower intensities. Read out noise due to the electronics generally follows a normal (zero-mean Gaussian) distribution. Analog to digital conversion causes quantization error due to the finite dynamic range and follows a uniform distribution with standard deviation inversely proportional to the number of bits. Deterministic noise errors can be corrected in preprocessing using dark-field and flat-field calibration images to perform flat-field correction. If known, the statistical properties of the noise can be included in the inverse problem formulation. Depending on the degradation, the blur or noise can be the dominant factor in III.16 - III.19. When discretized and put into matrix form, III.18 can be written as

\[
g_{\text{noise}} = g + w \quad \text{(III.20)}
\]

\[
= Af + w. \quad \text{(III.21)}
\]

This is used, in the discussion that follows, to describe methods to analyze and solve inverse problems.

Singular valued decomposition (SVD) can be used to analyze the linear system and better explain the problem of inverted noise. SVD follows

\[
A = U\Sigma V^T = \sum_{i=1}^{N} \sigma_i u_i v_i^T, \quad \text{(III.22)}
\]

where the convolution operator is decomposed into a scaled set of orthonormal matrix basis functions \( U \) and \( V \). \( u_i \) and \( v_i \) are the column vectors that make up the matrices \( U \) and \( V \), respectively. The diagonal matrix \( \Sigma \) denotes the basis function weights which are its diagonal entries \( \sigma_i \) referred to as the singular values. \( N \) is the total number of pixels. These singular values are in descending order and approach zero as \( i \) increases. With increasing \( i \) the spatial frequency of the basis function increases. The inverse convolution operator can be described as

\[
A^{-1} = V\Sigma^{-1}U^T \quad \text{(III.23)}
\]

\[
= \sum_{i=1}^{N} \frac{1}{\sigma_i} v_i u_i^T. \quad \text{(III.24)}
\]
and although it is not the best method it can be used to understand and solve the inverse problem.

Estimating the unblurred image by inverting the forward problem with the convolution operator is described mathematically as

\[
\begin{align*}
\text{f}_{\text{estimate}} & = A^{-1}g_{\text{noise}} \\
& = A^{-1}g + A^{-1}w \\
& = V\Sigma^{-1}U^Tg + V\Sigma^{-1}U^Tw \\
& = \sum_{i=1}^{N} \frac{u_i^T \sigma_i v_i}{\sigma_i} + \sum_{i=1}^{N} \frac{u_i^T w v_i}{\sigma_i},
\end{align*}
\]

where \(g\) is the noise-free blurred image. The term \(A^{-1}w\) represents the inverted noise which can often dominate the solution producing unacceptable results.

Singular values, \(\sigma_i\), decay at higher spatial frequencies and cause amplification of the noise, \(w\). This can begin to dominate the solution at these higher spatial frequencies. A convenient metric to characterize this instability is the condition number, given by \(\text{cond}(A) = \frac{\sigma_1}{\sigma_N}\). When this metric is small the problem is well conditioned and an acceptable solution is more likely; however, when the condition number is large the solution is likely to be dominated by inverted high-frequency noise. Highly discretized inverse problems tend to have a higher the condition number. Statistically, natural images tend to have less signal values at higher spatial frequencies (see Chapter VII), and after passing through an imaging lens system there is often attenuation at high spatial frequencies. Noise; however, can occur at any frequency. The spatial frequency at which the singular values are small and that the SNR is unity can be used to help avoid obtaining a solution dominated by noise amplification. This is described in the two following methods, truncated SVD and Tikhonov SVD.

Truncated SVD does not allow contributions, where the amplified noise is dominant over the signal, to enter the solution. This is achieved by only allowing the signal dominant singular values to reconstruct the signal, defined by the truncated sum

\[
\text{F}_{TSVD} = \sum_{i=1}^{k} \frac{u_i^T g_{\text{noise}}}{\sigma_i} v_i, \quad k < N,
\]
and thereby selectively omitting the amplified noise dominant contributions.

The singular values of an well conditioned convolution operator, $\sigma_i$, maintain high values and decay slowly, while a degraded convolution operator produces singular values that decay quickly. This can be quantified using the condition number. The singular values of the blurred data, $u_i^T g_{\text{noise}}$, decay until approaching a noise plateau. This transition can be used to select a reasonable filter factor so as to only include values that are above the noise floor.

Filtered Singular Valued Decomposition (FSVD),

$$F_{FSVD} = \sum_{i=1}^{N} \phi_i \frac{u_i^T g_{\text{noise}}}{\sigma_i} v_i,$$  \hspace{1cm} (III.30)

uses the filter factor $\phi$ to attenuate noise dominated contributions in the solution. Too much filtering suppresses the noise but discards valid data, while too little filtering allows noise to degrade the image. Truncated SVD can also be represented as (III.30) with the filter factor

$$\phi_i = \begin{cases} 1 & i = 1, \ldots, k \\ 0 & i = k + 1, \ldots, N \end{cases}.$$  \hspace{1cm} (III.31)

Gradually decaying singular values complicates the selection of $k$ or $\phi$ that balances the filtering strength.

For Tikhonov regularization we solve (III.30) with the filter factor

$$\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \quad i = 1, \ldots, N,$$  \hspace{1cm} (III.32)

where $\alpha$ is the regularization parameter. This in effect solves the total variation minimization problem

$$\min_f \left\{ ||g_{\text{noise}} - Af||^2_2 + \alpha^2 ||f||^2_2 \right\},$$  \hspace{1cm} (III.33)

where the first term fits the data using a least squares metric and the second term is a metric of the total energy of the solution. The filter factor $\alpha$ determines the relative influence between the two terms. For a small $\alpha$ the first term dominates data and noise can dominate the solution. For a large $\alpha$ the solution contains less high spatial frequencies and the noise is filtered out.
In order to investigate the trade-offs the regularization parameter creates, it is useful to define the following errors. Regularization error is the difference in the data fitting term (first term) when computing the filtered and unfiltered solutions. Perturbation error is the difference in the total energy of the solution (regularization term, second term in (III.33)) when computing the filtered and unfiltered solutions. These metrics and others such as the expected value, noise variance, can be used to help choose a regularization parameter by using methods such as discrepancy principle, general cross validation, and L-curve criterion. For example, in the L-curve criterion method, a log-log plot of the norm of the regularization term versus the norm of the solution produces a L-shaped curve whose elbow provides a good estimate of the regularization parameter. In Total Variation (TV) the regularization term is replaced by the integral of the gradient of the signal. This imposes a smoothness which can be used to selectively limit the higher order gradients in the solution, coinciding well with the idea that natural image exhibit a decaying spatial frequency content. These filtered solutions mitigate noise effects while preserving edges in the scene.

Inverse problems can also be regularized by iterative methods used to solve linear systems of equations. Here the number of iterations is used as a form of regularization because the regularization of noisy images is semi-convergent. As the iterations progress the solution first approaches then diverges from the best image estimate of the unknown object. All the solution methods discussed so far are deterministic. Next we briefly cover some statistical methods.

Statistical methods used to solve inverse problems, in general, model the noise as a random process, and can model the object as deterministic or a random process. In the former case, problem formulation remains ill-posed and a method of maximum likelihood, such as Lucy-Richardson deconvolution, can be used. Here constraints are imposed by limiting the domain of the likelihood function. Maximum likelihood solutions for stationary Gaussian noise or Poisson noise with a large expectation value are nearly identical to the least squares solution. The Lucy-Richardson algorithm deals well with Poisson limited noise and naturally enforces positivity in the solution. In the later case, Bayesian methods, such
as the Weiner filter, are used in a well conditioned problem.

Statistical methods can also be used to solve these same problems. Depending on the selection of noise and image prior models determines whether computation is more efficient in the spatial or frequency domains.

In addition to the filtering methods and the regularization methods described we can also solve these problems using Bayesian methods, such as Lucy Richardson Deconvolution (III.34). The Lucy-Richardson algorithm can be sped up using Fourier methods when the PSF is spatially invariant. Although not included in this discussion, stopping rules such as the randomized generalized cross-validation criterion are necessary to ensure the algorithm does not diverge from an appropriate solution

\[
\hat{f}_{k+1} = \hat{f}_k \left[ h^* \ast \frac{g_{\text{noise}}}{h \ast \hat{f}_k} \right]. \tag{III.34}
\]

To conclude this chapter on inverse problems some common PSFs and transfer functions describing image degradation relevant to this thesis are discussed, namely motion blur, out-of-focus blur, and lens aberrations. Motion blur is caused by relative motion between the camera and objects during image acquisition. In the most general case, different objects and the background can move independently with respect to the camera causing object motion blur. Alternatively, camera motion blur occurs when the camera moves with respect to a scene of objects fixed against the background. Here we will discuss the second, more simplified case. We assume the blur is spatially invariant and that the image sensor responds linearly to optical power collected from the object and arriving at the sensor surface. Pixels integrate the contributions of photons collected from the object as it moves across the image sensor, therefore the resulting intensity is proportional to the duration the image points of optical power stay on a given pixel. If the motion follows the time dependent vector in

\[ a(t) = \{a_1(t), a_2(t)\}, \quad 0 \leq t \leq T, \tag{III.35} \]

where \( T \) is exposure time, then the resulting noise-free blurred image is given by

\[ g^{(0)}(x) = \frac{1}{T} \int_0^T f^{(0)}[x - a(t)]dt. \tag{III.36} \]
For linear motion of constant velocity $c$, and ignoring diffraction effects, the PSF is given by a rect function defined as

$$K(x) = \frac{1}{T} \int_0^T \delta[x - a(t)]dt,$$  \hspace{1cm} (III.37)

and the transfer function is given by the sinc function

$$\mathcal{F}\{K(\omega)\} = \exp\left(-i\frac{T}{2}c \cdot \omega\right) \text{sinc}\left(\frac{T}{2\pi}c \cdot \omega\right),$$  \hspace{1cm} (III.38)

where $s = T \cdot c$ is the pixel displacement at the sensor plane. It is therefore zero when $s \cdot \omega = 2\pi n$, where $n = \pm 1, \pm 2, \ldots$. The zeros form a set of parallel lines orthogonal to the direction of motion. This transfer function leads to an inverse problem that is ill-conditioned since it has an infinite number of zeros and tends to zero as the frequency increases. For recorded images there is only a finite spatial frequency bandwidth is represented. The Fourier transform of the noise-free blurred image is also zero at these same locations. The recording process does add noise that can make the resulting blurred images different from zero at these locations.

Out-of-focus blur is modeled next. In the previous chapter imaging with a simple lens maps a point in the scene to a point in the image provided the lens equation (II.1) is satisfied. If the equation is not satisfied then the image formed by the system is out-of-focus and a point in the scene maps to a disk in the image space called the circle of confusion. The intensity distribution is approximately uniform over the circle of confusion. This leads to a PSF described as

$$K(x) = \frac{1}{\pi R^2} \chi_{\text{coc}}(x),$$  \hspace{1cm} (III.39)

where $R$ is the radius, and $\chi_{\text{coc}}(x)$ is the characteristic function of the circle of confusion, respectively.

A scene generally contains objects at various distances and therefore requires a PSF estimate for each given distance. In this case a spatially variant PSF is needed to describe a general scene. For a given out-of-focus PSF, the transfer function is given by

$$\hat{K}(\omega) = 2 \frac{J_1(R|\omega|)}{R|\omega|},$$  \hspace{1cm} (III.40)
where a $J_1$ is a first order Bessel function.

As is the case for the sinc describing the transfer function of motion blur, here too the Bessel function has infinite zero crossings. The crossings occur when $R|\omega| = \pi x_n; \ n = 1, 2, 3\ldots$, and define a set of concentric circles where the transfer function approaches zero. These circles are centered at the origin and $\omega_n = \frac{\pi}{R} x_n; \ n = 1, 2, 3\ldots$ defines their radii. As with motion blur this transfer function also attenuates high frequencies leading to information loss.

Lastly we deal with lens aberrations. The PSF

$$K_{coherent}(x) = \int P(u) \exp(-i x \cdot u) du,$$

(III.41)

is the Fourier Transform of the pupil function $P(u)$ of the optical imaging system. The amplitude transfer function (ATF)

$$\hat{K}_{coherent}(\omega) = \int K_{coherent}(x) \exp(-i \omega \cdot x) dx,$$

(III.42)

is a scaled Fourier Transform of the PSF and describes the response of coherent illumination imaging system which is linear in amplitude. Optical aberrations due to the lens can be modeled as phase distortions, $\phi(\omega)$, of the amplitude transfer function, and can be written

$$\hat{K}_{aberrated}(\omega) = \exp[i k \phi(\omega)] \hat{K}_{coherent}(\omega).$$

(III.43)

The optical transfer function (OTF) describes the response of the incoherent illumination imaging system which is linear in intensity. The OTF is given by

$$\hat{K}_{incoherent}(\omega) = \frac{1}{(2\pi)^2} \int \hat{K}_{coherent}(\omega + \omega') \hat{K}^*_{coherent}(\omega') d\omega',$$

(III.44)

which is the normalized autocorrelation function of the ATF. Abberations in incoherent illumination imaging systems are described in terms of equations (III.41)-(III.44). The resulting OTF can be used to formulate an inverse problem in order to correct for aberrations in imaging systems. In the case of motion blur, out-of-focus blur, and blur due to aberrations; the zeros within the transfer function can be used to extrapolate information about the physical system, namely information about the motion or focus position of the system, respectively.
Any noise-free blurred image cannot contain information at frequencies where the transfer function is zero. In other words, the noise-free imaging system cannot produce “invisible” images, images outside the range of its transfer function. Noise added during the recording process can extend the range of images that the system produces to include invisible images. Inverting a degraded (blurred) noisy image or more generally solving a inverse linear set of equations is a main challenge and common theme in computational imaging. Linear inverse problems often suffer from ill-posedness which is caused by blurred images that are non-existent, non-unique, or do not depend continuously on the data. This is due to the fact that the imaging system does not transmit the all scene information equally or completely. Additionally, measuring and recording the image adds noise to this incomplete signal further compounding the problem. In order to successfully solve this class of problems, we solve for an approximate (rather than exact) less stringent solution. At the same time we enforce a more stringent constraint to include only physically realistic images.

Computational Imaging is the union of cameras, computers, hardware and algorithms, where the data acquired is not the finished product, but a means from which to formulate an inverse problem. Through the use of clever hardware modifications and tailored problem formulation, Computational Imaging can be used to better condition the imaging problems so that the overall system can surpass the limitations described in the previous chapter. As the field of Computational Imaging has progressed some advances have made sophomoric steps forward; yet, its future is extremely encouraging. One goal of the technology is to transfer the most information from the object space to the reported data. Alternatively, the objective could be to reduce complexity and cost in components within the imager while still retaining the same level of performance. The physical form Computational Imagers adhere to can be very foreign when compared to the traditional camera to which most people are akin. Incorporating machine vision technologies, beyond face or smile recognition, into daily use cameras is likely to continue to add useful functionality. The overall impact Computational Imager will have it to likely alter the way users view and use imagery.
III Bibliography


Chapter IV

Image processing for cameras with fiber bundle image relay

IV.A Introduction: Traditional Imagers and Fiber-Coupled Imagers

Traditional imaging systems form images onto flat image sensors, a difficult design constraint for wide-angle imaging in which off-axis aberrations dominate. A monocentric lens, where all optical surfaces are spherical and share a common center of curvature, images without coma or astigmatic aberration and significantly reduces imager volume (Fig. IV.1) [1, 2]. This comes at a cost of forming a spherical image surface, which is incompatible with conventional image sensor fabrication technology [3]. Fiber optic bundles, a dense array of high index optical fibers with thin low index cladding, can be used to incoherently transfer optical power between an input and output surface [4, 5].

This technique was used to couple the spherical image surface from a 60° field-of-view monocentric lens to one or more 221 kpixel CCD sensors with 23 µm pitch pixels [6]. More recently, the same basic structure has been applied to double the lens field-of-view and couple the image to high-resolution CMOS sensors [7, 8]. The lens, fiber bundle, and fiber-coupled sensor are shown in Fig. IV.2, and a more detailed view of the structure of the fiber bundle and sensor is shown in Fig.
IV.3. Signal acquisition in such fiber-coupled imagers results from a cascade of optical processes. A continuous image is formed by the lens on the fiber bundle surface, where the local signal coupling may be strongly space variant according to the quasi-periodic fiber bundle structure. This signal is transferred through the multimode fiber core, where it loses all spatial coherence then is emitted from the output face of the fiber. In general, there can be a small free space gap (an air space or thickness of isotropic adhesive) between the rear face of the fiber and the focal plane, and the signal emission into this region can depend sensitively on launch conditions and wavelength. Finally, this signal is captured by the regular sampling grid of the focal plane sensor. If the fiber pitch is much smaller than the sensor pixels, this image acquisition “cascade” might be neglected. To achieve the highest possible spatial resolution; however, both fiber bundle and pixel pitch need to be minimized. Understanding each step of the image acquisition process can enable a broader class of computational image processing techniques.

Here we present the first (to our knowledge) systematic characterization of the spatially variant impulse response at various stages within a fiber-coupled
Figure IV.2: (a) Monocentric lens (MC) comprised of a symmetric 2-glass achromat (top) and meniscus lens (bottom). (b) A fiber bundle relays a business card logo from its flat to its curved surface. (c) A fiber bundle coupled to a sensor using optical epoxy.

image (an imaging system that uses fiber bundle image relay). We use a fiber-coupled monocentric lens camera prototype to investigate computational imaging techniques applicable to any imaging system that uses fiber bundles to relay or field-flatten images onto planar image sensors as well as system-specific issues related to combining multiple sensors to form visually seamless panoramas. Section IV.B is devoted to an in-depth investigation of the local image transfer characteristics of each component within the prototype. Section IV.C describes the methods we employed to globally remove image artifacts due to fiber optic bundle image transfer and system components. Section IV.D shows the results of mapping the spherical image space to a planar space and how to form panoramas by aligning and stitching contiguous sensor data. Section IV.E explains the experimental setup under which data is collected for camera calibration and performance measurements. Section IV.F discusses our experimental results from the prototype imager and compares the performance to a much larger conventional digital SLR “benchmark” camera Fig. IV.4.
Figure IV.3: (a) The fiber bundle structure and defects lead to image artifacts. Fiber absorbers between every other 2.5 \( \mu \text{m} \) fiber intersection prevent crosstalk. (b) An OVT5653 sensor is used to sample the image exiting the fiber bundle. The two magnified inset micrographs are of the same scale.

IV.B Component and System Point Spread Function

We characterized our custom monocentric lens fabricated by Optimax (focal length \( f = 12 \text{ mm} \), \( \text{F/#} = 1.35 \)) (Fig. IV.2(a))\[9\] and compared it to a high-resolution Nikon LU Plan APO 150X microscope objective (\( f = 1.33 \text{ mm} \), numerical aperture NA = 0.9). Both of these objectives were used with the following devices under test (DUTs) to sample the image formed: a 24AS Schott fiber bundle (Figs. IV.2(b) and IV.3(a)) with 2.5 \( \mu \text{m} \) fiber pitch \[10, 11\], an Omnivision OV5653 backside illuminated 5 megapixel color CMOS sensor (Fig. IV.3(b)) with 1.75 \( \mu \text{m} \) pixel pitch and image area of 4592 \( \mu \text{m} \) wide by 3423 \( \mu \text{m} \) high \[12\], and a fiber-coupled sensor formed by bonding the fiber bundle to the sensor using UV-cured Norland optical adhesive NOA72 (Fig. IV.2(c)). The internal structure of the fiber bundle is shown in Fig. IV.3(a) and leads to artifacts in the relayed imagery. A schematic and image of the cross-section of the fiber-coupled sensor is shown in Fig. IV.5. The fiber bundle’s input face has been cut and polished with a 12 mm radius of curvature to mate with the meniscus lens. A pedestal has been cut out of its exit face to mate with an image sensor whose cover glass has been removed. The schematic illustrates that the light will pass through the lens, fiber bundle, epoxy layer, lens array, and Bayer filter before it is collected by the image sensor. The
Figure IV.4: (a) Side-by-side size comparison of the prototype with an F/4 commercial camera of similar field-of-view. (b) The F/1.35, 30 megapixel, 126° field-of-view fiber-coupled monocentric lens imager prototype.

epoxy layer must be kept to a minimum thickness in order to prevent the divergent light exiting the fiber bundle to defocus before it is collected by the image sensor. The fiber bundle’s polished surface flatness is on the order of a wavelength, and the OV5653 surface flatness variations are approximately 1.5 µm, as measured with the Veeco NT1100 Optical Profiling System. A cross-section of one of the fiber-coupled sensors confirmed a 2 µm adhesive thickness over the sensor’s 1.6 µm thick lenslet and color filter. Extremely wide-angle, spatially and temporally incoherent light emission through these surface and interface conditions do not generate high-contrast, submicron spatial patterns; therefore, negligible interference effects are detected by the sensor.

To measure the impulse response of the DUTs we built an optical setup to simulate a moving infinitely distant point source using a white LED light behind a small (3 µm) pinhole aperture at the focal plane of a long focal length Fourier-transform lens (f = 360 mm, F/# = 5.14, and diameter Ø = 76 mm) (Fig. IV.6). The pinhole and LED are both mounted on a precision translation stage and shifted in small lateral steps to characterize the highly localized spatial variations in the
Figure IV.5: (a) Schematic and (b) cross-section showing the internal structure and scale of the fiber-coupled sensor. The fiber bundle relays light incident at its surface to the CMOS image sensor though an epoxy bonding layer, lens array, and Bayer filter.

impulse response of the DUTs. The total translation corresponded to a 25 µm displacement at the DUT and was discretized into 200 steps. For each step, an image was captured, and all images were then used to form the 3D data cubes shown in Fig. IV.7 and the data projections in Fig. IV.8. To form the data cube, the image frames in $x, y$ are stacked up along $z$. The $x, y$ coordinates of the data cube are in the pixel domain of the sensor. The $z$ axis contains the individual photographs captured by the sensor for each step of the point source translation. Projections onto each coordinate axis plane are used to help interpret the data. The projection onto the bottom plane shows the integrated response of the sensor for all frames captured during translation. Projections onto the two side planes show the behavior due to the translation, where each frame is integrated in one dimension to form these side projections. We repeated this process for each DUT using both the microscope objective and monocentric lens. Figure IV.8 presents the data collected but only shows two of the three projections of the data cube, namely, the x-y pixel-pixel sensor domain and the x-z pixel-distance domain.

A Nikon LU Plan APO 150X microscope objective (1.6 µm spot size) was used to illuminate each DUT to produce the data in the top row of Fig. IV.8, while a custom monocentric lens (2.6 µm spot size) illuminated each DUT producing data shown on the bottom row of Fig. IV.8 (see schematics in Figs. IV.6(a) and
Figure IV.6: Schematic of the impulse response characterization setup where an LED, pinhole, and collimation lens ($f = 360 \text{ mm}, F/# = 5.14, \Theta = 76\text{mm}$) form a collimated point source. The setup in (a) uses a high resolution microscope objective ($f = 1.33 \text{ mm}, NA = 0.9$) to characterize the monocentric lens in setup (b). Light emitted by the LED is focused onto a DUT. A translation stage is used to move the point source so that the focused spot moves $25 \mu\text{m}$ laterally across the surface of the DUT. This characterization is done to measure the local spatial variance of the impulse response for each DUT. The DUTs are a fiber bundle, sensor, and fiber-coupled sensor (Fig. IV.8).
**Figure IV.7:** Data cube representation of the impulse response of various DUTs corresponding to the bottom row of Fig. IV.8 collected using the setup shown in Fig. IV.6(b). The x and y axes correspond to the image sensor space, while the z axis is formed by stacking up 200 images taken during the 25 μm translation of a point source.
The spot size was defined by the $e^{-2}$ intensity level. For comparison, Figs. IV.8(a) and IV.8(e) show the ideal impulse response of the microscope objective and monocentric lens, respectively (measured using a 1600 × 1200 resolution, 4.4 µm pixel pitch Keyence VHX-1000 microscope with a 20x optical magnification lens VH-Z100UW of NA < 0.48 and exposure time of 1/60th of a second). Both follow the image of the translated point source smoothly and continuously, meaning these devices are spatially invariant. The width of the line corresponds to the spot size produced by each lens, namely 1.6 µm & 2.6 µm. Figures IV.8(b) and IV.8(f) show the same impulse response measured using an Omnivision 5653 sensor without any optical magnification. The Bayer-filtered 1.75 µm pixels cause the spatial quantization in image sensor space. The difference in the slope in Fig. IV.8(f) when compared to the rest of the figure is due to the index-matching fluid-filled gap between the planar sensor and spherical image surface. Figures IV.8(c) and IV.8(g) show the same impulse response as Figs. IV.8(a) and IV.8(e) after being relayed by the fiber bundle and measured with the Keyence microscope. These images exhibit large variation in transmission depending on the location of the impulse relative to the fiber structure, but this is primarily due to the light’s higher divergence angle at the fiber’s exit face (when incident at the non-core regions), which cannot be captured because it is beyond the microscope’s numerical aperture. Figures IV.8(d) and IV.8(h) show data collected by a image sensor coupled directly to the exit face of the fiber bundle. There is some image spread but more uniform energy transmission. A similar evolution of the impulse response of the DUTs in Figs. IV.8(e)-IV.8(f)-IV.8(g)-IV.8(h) is also seen in the corresponding data in Fig. IV.7.

Figures IV.8(c,d,g,h) show the same measurements as in Figs. IV.8(a,b,e,f) except a fiber bundle is introduced into the system to relay the image. The fiber bundles used in our system were made by Schott using 24AS glass, with a five-sided core structure and a 2.5 µm pitch (for more information, see [10, 11]). Figures IV.8(c) and IV.8(g) show that the fiber bundle impulse response is extremely spatially variant on a fine spatial scale (on the order of the pixels), allowing a large signal to pass when light is incident on one of the fiber cores. Note here that the
Figure IV.8: Impulse characterization of system components where each subfigure contains (from top to bottom) the integration of the 200 images captured during a point source scan, a projection of the data onto translation-pixel space (see Fig. IV.7), and an image of the setup. These plots show the focus and linearity for each DUT. (a)-(d) and (e)-(h) were captured using the setups in Fig. IV.6(a) and IV.6(b), respectively. The difference in the slope of the data presented in (f) is due to the planar sensor being placed at the spherical image surface, which introduces an index-matching fluid-filled gap at the edges, where light propagates before reaching the sensor. Although the intermediate response is spatially variant (g), invariance is recovered by the complete prototype (h). This extends the applicable image processing techniques that can be used.
fiber bundle does transmit light incident on non-core regions but that the Keyence microscope’s numerical aperture (NA < 0.48) does not collect the highly divergent emitted light. However, this divergent light is detected by a CMOS sensor placed in contact with the fiber bundle. Spatially variant impulse response behavior severely limits the image processing that can be used on such a system. Fortunately, Figs. IV.8(d) and IV.8(h) show that a fiber bundle that has been bonded to an image sensor exhibits considerably less spatial variance than the fiber bundle alone (Figs. IV.8(c) and IV.8(g)). This study shows the evolution of the impulse response behavior at the various stages within the fiber-coupled monocentric lens imager. It reveals that, although the impulse response is highly spatially variant within the system, the end-to-end system performance is in general not highly spatially variant. This enables us to use image processing techniques such as deconvolution [13, 14] in order to deblur images or extended depth-of-focus in order to improve image quality [15, 16].

IV.C Global Artifact Removal Calibration of the Fiber-Coupled Sensor and Prototype

The fiber bundle’s internal structure produces an occlusion pattern that is superimposed on the images it transfers. Strong sampling effects unique to fiber-coupled imagers are produced when the irregular fiber lattice and periodic pixel lattice are of comparable pitch, which leads to misalignment between the lattice sites, causing moiré patterns. There have been efforts to remove the occlusion effects related to fiber bundle image relay using techniques such as spatial filtering [17], Gaussian smoothing [18], or a Bayesian framework [19]. However, all of these techniques reduce image resolution. The objective here is similar to work presented in [18, 19] with more general obscuration/occlusion removal techniques documented in [20, 21, 22, 23, 24].

Modern CMOS focal planes achieve high apparent uniformity using flat-field calibration and analog processing that substantially masks unavoidable point defects, local variations in sensitivity, and fixed pattern noise. Flat-fielding has
been used to correct fiber relayed images [25] and spectra [26]. We applied a similar process to calibrate the fiber-coupled sensor by experimentally acquiring a flat-field obscuration map for a rigidly attached fiber-coupled sensor or fiber-coupled monocentric lens imager. This map quantifies artifacts that affect the system across the overall image space, whether introduced by the fiber structure or coupling between lens, fiber bundle, and sensor. It also allows correction of field nonuniformity due to radial vignetting due to the angle of incident light on the fiber’s input face as well as lens vignetting. The flat-field obscuration map is essentially an image of an evenly illuminated white planar target object acquired by the imaging system. The target should provide uniform illumination free of any structure and does not require the target be in focus. The dark-field image is used to account for the dark current. These artifacts due to the lens, fiber bundle, adhesive, and sensor can have higher order dependence on system temperature (due to differential thermal expansion) or sensor gain and exposure settings (due to nonlinear sensor response), leading to the use of either multiple calibration images or a parametrized response function. This idea was presented in [27, 28] which used multiple calibration images to deal with nonlinear intensity response. This is the topic of ongoing research, but preliminary results indicate the approach is robust to moderate temperature variations (20 °C) and a range of exposures.

The flat-field calibration images show that the finely structured global attenuation moiré pattern that overlaid the image is deterministic (Fig. IV.9). The most attenuated values fall to a level of about 65% of the peak intensity on-axis (Fig. IV.9(a)). Fig. IV.9(b) shows the vignetting effects that lead to the significant attenuation (see the left side of the calibration image) and shifting of the histogram. Even in this case, the signal level is above 10% of the peak intensity. In both cases, the distribution that centered below the 10% value corresponds to a region of the sensor, which is not coupled to the fiber bundle (dark regions at the edge of the calibration images) and can be ignored since it does not contain image data. For these reasons, the flat-field calibration image can be used to amplify the signal of subsequent images. Figure IV.10 shows the calibration image (a) taken of a flat white scene, an image taken of an indoor laboratory scene (b), and the
(a) On-axis flat-field calibration image and histogram. Note: there is no fiber coupling at the right and left edges.

(b) Off-axis flat-field calibration image and histogram. The histogram is shifted due to vignetting on the image’s left edge.

Figure IV.9: Two typical flat-field calibration images (left) and their histograms (right) captured using the fiber bundle relayed imager. The lowest intensities transmitted by the bundle are above 175 (65%) (a) on-axis and 25 (10%) (b) off-axis on an 8-bit scale. Off-axis intensity transmission suffers from vignetting. The intensity values below 10% correspond to a region of the sensor where there is no fiber bundle image relay and can, therefore, be ignored. The large scale structure in (b) is due to delamination in the optical adhesive; yet, this calibration is still effective.
(a) Calibration image of a flat white scene shows obscuration effects.

(b) Image of a lab scene with obscuration effects present.

(c) Image IV.10(b) with obscuration removal by amplifying using the calibration image.

**Figure IV.10:** A calibration image of a flat white scene (a) shows artifacts due to broken fibers, scratches, and other imperfections as well as the fiber bundle’s internal structure, all of which lead to obscuration and color artifacts (b). These artifacts inherent to fiber bundle imaging are corrected using a calibration image (c). The center white region in the magnified views of (b) and (c) are used to quantify the calibration’s effectiveness, see Table IV.1.
Table IV.1: Statistics taken over the same white part of the images in Fig. IV.10 before and after occlusion removal show the process significantly improves image uniformity.

<table>
<thead>
<tr>
<th>Color Channel</th>
<th>Uncalibrated Image Fig. IV.10(b)</th>
<th>Calibrated Image Fig. IV.10(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Red</td>
<td>178.35</td>
<td>20.59</td>
</tr>
<tr>
<td>Green1</td>
<td>175.34</td>
<td>23.88</td>
</tr>
<tr>
<td>Green2</td>
<td>174.04</td>
<td>24.52</td>
</tr>
<tr>
<td>Blue</td>
<td>159.86</td>
<td>17.73</td>
</tr>
</tbody>
</table>

corrected image (c). An example of the same correction applied to an outdoor image is shown in Figure IV.11, where (a) is the calibration image taken of a flat white scene, (b) is an image taken of an outdoor scene, and (c) is the corrected image. The correction was made by element-wise division, which compensated for attenuation of the signal. The normalization factor for each pixel was obtained from the corresponding pixel within the calibration image. In the most attenuated regions of the image, this process was still extremely effective. Noise in the flat-field calibration image is amplified and can be significant in darker regions of the image, which require strong amplification. Although the effects are not significant in our results, they can be reduced by averaging multiple dark-field and flat-field images, respectively. Areas that may not be remedied by this process, regions of the calibration image where the signal was below a certain threshold, could have been further corrected by interpolating from nearby regions. Table IV.1 quantifies the improvement in uniformity over a uniform white area in Figs. IV.10(b) and IV.10(c). It was shown in [29] that the flat-fielded images still contain remnants of moiré “beat patterns” that can be used to measure fiber bundle distortion. This residual variation is evident in the calibrated image in Fig. IV.10(c) and its corresponding statistics in Table IV.1.
(a) Flat-field calibration image (b) Image taken outdoor with artifacts and obscuration effects present. (c) Corrected image restored by amplification using the calibration image.

**Figure IV.11:** (a) A flat-field calibration image of a uniform white scene shows artifacts due to broken fibers, scratches, and other imperfections as well as the fiber bundle's internal structure, all of which lead to artifacts in the acquired images. (b) A raw outdoor image exhibits degradation due to fiber bundle transfer. (c) Corrected image showing that the artifacts inherent to fiber bundle imaging can be corrected using a calibration image.
IV.D  Spherical to Planar Mapping, Image Alignment, and Stitching Together Image Seams

Even with an ideal lens, there is an intrinsic distortion associated with mapping a portion of a spherical image onto a flat image sensor, and the distortion is increasingly apparent as the field-of-view recorded with a single sensor increases. This distortion can be corrected using one of several general mathematical transformations [30]. The mapping of a direct spherical projection onto a plane are given by

\[ s = 2r - \frac{R r}{\sqrt{R^2 - r^2}}, \]

where the radial distance \( r \) is transformed into a scaled form \( s \). \( R \) is the radial distance from the point where the spherical surface tangent plane is orthogonal to the sensor plane. Here Cartesian coordinates are transformed into azimuth-elevation global coordinates. Each sensor is mapped radially in this way to correct for radial pincushion distortion and to aid in stitching individual sensor data to form panoramas.

Assembly of the imager involves affixing the fiber-coupled sensor modules to the lens system. In addition to cutting the input and output faces of the fiber bundles, the sides are also cut such that multiple fiber-coupled sensors can fit together closely to form a contiguous sphere of the same radius as the meniscus lens so the two can interface. Physical alignment of the modules within the system is subject
Figure IV.13: Schematic showing the fiber-coupled sensors in the prototype and their respective field-of-view. The six sensors gather image data across a 126° field-of-view.

to error but, once assembled, is constant for the lifetime of the system. The width of the seams is limited by fiber shaping methods, which may ultimately be comparable or smaller than that of the fiber cores. But, even with near-perfect alignment, there will be some information loss at the seams (if only spreading signals between conjoined fibers), and the current prototype imager is far from this ideal. Overlapping the image boundaries would greatly facilitate stitching the image boundaries but is not possible with the current design and would require greatly increasing camera size and complexity. Here, we describe the part of the overall processing flow (Fig. IV.12) that follows occlusion removal and radial vignetting correction. We separate the occlusion and radial vignetting corrected images from the various sensors into their four Bayer color channels, process them to form four panorama color channels, and then demosaic them into a color panorama. The active area of the individual sensors is not fully illuminated by the image being relayed by the fiber bundles. The images from the different sensors must be distortion corrected, rotated, cropped, and intensity balanced prior to the individual image data
alignment needed to form a contiguous panorama for each color channel. One di-

mensional interpolation from each boundary edge of valid data is used to restore
lost data at the image seams. Vignetting effects are compensated using a modified
ramp function to amplify the lower signals that appear at the extreme field an-
gles. Gamma correction is then performed on the four color channel panoramas.
After the four color channels are processed as described to form panorama color
channels, they are interpolated to form a single white balanced color panoramic
image.

The highly localized structural and color artifacts due to the fiber bundle
limit the effectiveness of demosaic interpolation techniques that rely on texture,
edges, or gradient image content. After evaluating several interpolation techniques,
we found that bilinear interpolation produced the best resulting images. A global
saturation adjustment was used to compensate for the sensor’s color response.
The steps in the methodology mentioned here were selected out of a large class
of techniques and can be modified or replaced. For example, the stitching process
documented in [31] is effective at balancing the content across seams that exhibit
discontinuous content due to flawed image performance, producing a much less
noticeable image boundary.

**IV.E Experimental Setup**

Experimental data was collected in the lab with the monocentric lens us-
ing two adjacent fiber-coupled sensors corresponding to Sensors 4 and 5 in Figs.
IV.1(b) or IV.13. One sensor imaged the scene from 0° on axis to 20°, and the
other sensor imaged from 20° to 40°. This image boundary was chosen because it
well represented an arbitrary image seam since it was not along the optical axis.
The planar scene was positioned at an orthogonal distance of 1 m from the cam-
era along the optical axis. The scene was a flat printed grid (designed to be 1
m from the camera) with centimeter markings and concentric circles denoting the
angle deviation from the optical axis. Laboratory data (Fig. IV.14) enabled us
to quantify distortion correction and the amount of information loss at the image
(a) Raw images gathered from two adjacent fiber-coupled sensors (Sensors 4 and 5 in Figs. IV.1(b) and IV.13) in the monocentric imager show barrel distortion and information loss at the sensor adjacent boundary.

(b) Images from (a) with distortion correction and interpolation to stitch the images together. The ground-truth mapping angles markings are shown in red to indicate residual error in distortion correction.

Figure IV.14: Distortion correction and image stitching.
Figure IV.15: An image taken using a conventional wide field-of-view camera (Canon 5D mark II with EF 8-15mm F/4 fisheye lens) with similar characteristics as the fiber-coupled monocentric lens imager. Magnified regions show the image resolution.

boundary between closely tiled adjacent sensors, which corresponded to 2 mm in object space at a distance of 1 m, or $0.2^\circ$. From this characterization, we can accurately stitch the scene together. The calibration described here needs to be calculated only once since the monocentric lens imager parts are held fixed by adhesive, with the exception of color balance which is affected by the ambient illumination spectrum. This sufficiently characterizes the imager so that significantly improved panoramas can be processed from the individual raw image sensor data. The complete systematic methodology we used to process the fiber-coupled monocentric lens image data is documented in Fig. IV.12. Partially and completely processed images resulting from this process are shown in Fig. IV.14. In what follows, we demonstrate that this process enables the fiber-coupled monocentric lens imager to produce artifact-limited panoramas that suffer from less aberrations and distortions than traditional wide FOV lenses.

We compared the performance of the prototype against a commercially available camera (Canon 5D mark II with EF 8-15 mm F/4 fisheye lens) by taking
(a) Unprocessed fiber-coupled monocentric lens individual sensor data

(b) Image data from (a) processed to form a continuous panorama

**Figure IV.16:** Prototype image showing significant image improvement due to distortion correction, image stitching, and interpolation to form panorama. Artifacts due to bubbles in the optical adhesive, broken fibers, and fiber bundle scratches are effectively removed by flat-field calibration. The flat-field calibration images for Sensors 1 and 4 (from right) are shown in Fig. IV.9. Magnified regions show that the prototypes resolution is superior to the benchmark camera images shown in Fig. IV.15.
images of the same outdoor scene under identical conditions. The outdoor scene was imaged using all six fiber-coupled sensors within the F/1.35 fiber-coupled monocentric lens imager to form a panorama (Fig. IV.13). The prototype has a 126° x 16° field-of-view and resolution of 30 megapixel that can be increased by incorporating more fiber-coupled sensors. The ground-truth comparison image is shown in Fig. IV.15. This comparison camera was chosen, since it was designed to meet similar performance specifications without being limited to a small form factor. Figure IV.16 exemplifies the improvement our methodology provides to fiber relayed imagery. Of particular interest is the removal of artifacts due to bubbles (in the optical adhesive used to mate the lens with the fiber), broken fibers, and scratches. This first prototype suffered from more of these effects than subsequent prototypes; however, we chose to include this imagery, since it best illustrates the efficacy of the techniques described. Additionally, the significant vignetting effects at the edges of the wide angle panorama are corrected for using an amplification map. Distortion correction and interpolation of missing data at the boundaries allows the formation of panoramic imagery. The composite image demonstrated in Fig. IV.16(b) shows that, by using fiber bundle image relay along with the techniques described here, we are now able to use lenses (that form non-planar image surfaces and are otherwise incompatible with current technology) in order to form high-resolution images with less distortion than the commercial alternative. Magnified views (enclosed in red) inset Figs. IV.15 and IV.16 show a portion of the scene that is about 50° left of center and 150 m downrange for both comparison cameras. This comparison shows the prototype exhibits less distortion, less chromatic aberration, and a large boost in resolution.

IV.F Discussion and Conclusion

This paper summarizes the characterization and image processing for imaging systems that use fiber bundle image relay, including a component-by-component transfer function characterization of the fiber-coupled imager. We found that the high spatial variance of some of the individual components is reduced when used in
conjunction with other components, allowing for more flexibility when implement-
ing image processing techniques. We experimentally demonstrated how flat-field
correction proves beneficial for fiber-coupled imaging systems by correcting for the
response of the lens, fiber bundle, and sensor, all of which produce fixed pattern ar-
tifacts. We showed how to stitch together panoramas taking into account artifacts,
image transfer uniformity, mapping, and the registration of individual sensor data
prior to interpolation of image information not captured between adjacent sensors.
Using this methodology, we formed a 30 megapixel 126° field-of-view fiber-coupled
panoramic image, which compares favorably in image resolution and distortion
to the benchmark commercial camera whose volume (for lens and sensor only) is
more than 10 times the fiber-coupled prototype. A future direction of research to
improve the technology can be achieved through custom fiber bundle fabrication
to achieve alternate fiber orientation and transfer properties.

Acknowledgments

Support was provided by the DARPA SCENICC program contract W911NF-
11-C-0210. The authors thank Glenn Schuster, William M. Mellette, and Nojan
Motamedi for technical discussion and laboratory support.

Chapter IV, contains material published in the following, of which the dis-
sertation author was the primary investigator and author:

Stephen J. Olivas, Ashkan Arianpour, Igor Stamenov, Rick Morrison, Ron
A. Stack, Adam R. Johnson, Ilya P. Agurok, and Joseph E. Ford “Image Processing
for Cameras with Fiber Bundle Image Relay”, Applied Optics, 54 (5) 1124-1137
IV Bibliography


[10] Schott® “Fiber Optic Faceplates,”


Chapter V

Digital Image Processing for Wide-Angle Highly Spatially-Variant Imagers

V.A Introduction - Traditional and Fiber Coupled Imagers

We use the system presented, along with relevant references, in the previous Chapter to show a spatially variant deconvolution process [1, 2] that uses focal swept images to extend the depth-of-focus of a wide field-of-view imager. Fig. IV.4(a) shows a side-by-side size comparison of the prototype camera and commercial camera.

Section V.B explains the experimental setup used to measure the static depth-of-focus performance and to collect focal swept PSFs and images for extended depth-of-focus (EDOF) processing. Section V.C discusses the spatially variant EDOF algorithm and the resulting images the current EDOF system produces. Section V.D summarizes the fiber bundle artifact removal, the angle and depth dependent focal swept PSF measurements, and the spatially variant EDOF imaging process.
V.B Experimental Setup

In order to produce data to perform and test the spatially variant extended depth-of-focus algorithm we configured the imager to take images both on and off axis using two fiber coupled sensors corresponding to Sensors 4 and 6 in Fig. IV.1(b), also shown in green in Fig. V.1. One sensor imaged the scene from 0° on-axis to 20° and the other sensor imaged from 40° to 60°. The experimental lab scene we tailored to test the EDOF algorithm was setup to include targets placed at various distances (65 cm, 83 cm, 1 m, 1.33 m and 2 m) perpendicular to the optical axis for each of the two angle ranges (Fig. V.1 and V.2). Ground truth images were taken with the camera focused to an on-axis orthogonal distance of 1 m from the camera (Fig. V.3). These images show that the depth-of-focus of the imager is limited to a certain depth surrounding the focus.

In order to extend the depth-of-focus we employ a swept focus method presented by Nayar et. al. [3], which relies on PSF depth independence; however, we formulate the problem in a spatially variant framework in order to deal with the significant PSF dependence on field angle. The monocentric lens was mounted on a linear on-axis translation stage such that the focus could be swept during the capture of a single exposure (see inset of Fig. V.1). The focal sweep corresponded to an object distance from 0.5 m to infinite conjugate. The actuated system was also used to capture single exposures of point sources placed at various angles.
Figure V.2: EDOF imaging system and experimental lab setup.

(a) Actuated system  (b) Experimental setup

Figure V.3: Image taken with lens focused to 1 m. Five targets are placed in the scene at depths of 0.65 m, 0.83 m, 1 m, 1.33 m, and 2 m orthogonal to the optical axis. Magnified views show the limited depth-of-focus.
Figure V.4: Single exposure focal swept images of a point source placed at 1 m orthogonal along the optical axis for various field angles, and placed at $0^\circ$ and $60^\circ$ for a set of distances orthogonal to the optical axis. This shows the PSF dependence on angle is significant and varies continuously. This also shows that the on-axis PSFs are approximately independent of depth while the off-axis PSFs are highly dependent on depth.
Figure V.5: The spatially variant PSF changes depending on object distance and angle (note PSF scale). Many exposures were taken as the EDOF system’s focus was incrementally stepped. Profiles of the captured images as a function of focal position (intensity shown using false color plots) show that the behavior is different for the various depths. Integrated PSFs are shown in grayscale. However, the integrated on-axis PSFs (a)-(e) are substantially independent on scene depth; whereas, the off-axis PSFs (f)-(j) are greatly influenced by the distance of the point source.
Figure V.6: Single exposure focal swept images and PSFs. The focus is swept from 0.5 m to infinite conjugate. Five targets are placed in the scene at depths of 0.65 m, 0.83 m, 1 m, 1.33 m, and 2 m orthogonal to the optical axis. Magnified views show the extent of blur. Also shown are measured PSFs which exhibit more than 140 pixels of blur at high angles.
across the scene (fixed orthogonal distance of 1 m) (Fig. V.4). This was done to measure PSFs as a function of field angle, which is later used in the spatially variant EDOF algorithm. That is that the spatially variant EDOF algorithm is based on angular dependent PSFs measured at a fixed orthogonal distance. Additionally, we explored the system response to point sources at various object distances (field angle fixed to 0° and 60°) (Fig. V.4). This data reveals that the idea that the PSF is relatively independent of depth is reasonable for on-axis images taken with the prototype, as was the case for the entire image collected by the substantially telecentric lens used in reference [3]. In other words, as the focus is swept, the distribution of light on the detector varies notably depending on the distance of the point source; however, the integrated swept intensity is approximately the same for point sources at different distances (see 0° PSF data in Fig. V.4). Since the on-axis PSF is substantially depth independent, this allows conventional space invariant image processing techniques be used to extend the depth-of-focus of the image scene in the paraxial region. This is in stark contrast to the behavior of the integrated off-axis focal swept PSFs which exhibit significant depth dependence. This behavior is evident in the 60° data shown in Fig. V.4. Even though the off-axis PSF is depth dependent, there is a range at which the EDOF algorithm can improve the system performance. In addition to the single exposure focal swept PSF measurements shown in Fig. V.4, Multiple PSF exposures were taken as the focus stepped incrementally to show the PSF dependence on focal position during the focal sweep. This stepped PSF measurement investigation is summarized in Fig. V.5. Although not experimentally explored here, this raises the possibility of performing depth dependent deblurring based on a priori scene information or a depth map captured using a low resolution stereo pair imager.

The fiber coupled monocentric lens prototype has the property that it focuses to an orthogonal object plane. In our experiments, we used the 1 m orthogonal distance to the camera as our central distance (Fig. V.3). To extend the EDOF framework to wide angle images we must take into account the PSF variation as a function of field angle (Fig. V.4). Raw focal swept images and spatially variant PSF data (angle variant, fixed orthogonal depth) captured using the actuated
EDOF are shown in Fig. V.6. Using these PSFs, the raw focal swept image, the flat-field calibration image, and the dark-field image we implemented a spatially variant EDOF image processing algorithm. This spatially variant deconvolution algorithm is described in the following section.

V.C Spatially Variant Extended Depth-of-Focus Algorithm

Here we present the spatially variant EDOF algorithm that was used to process the mechanically actuated focal swept imagery. This relies on solving

\[
    z(x, y) = [Hu](x, y) = \int u(x - s, y - t)h(x - s, y - t; s, t)dsdt,
\]

using a degraded image and measured spatially-variant PSF data (taken at various field angles with fixed orthogonal distance of 1 m). The inset of Fig. V.1 shows a schematic of the system along with a photo of the prototype.

Reconstruction of the original sharp image can be formulated in the Bayesian sense as seeking the maximally probable solution given certain assumptions on distributions of the noise and image \(u\). A common choice of the image prior distribution is proportional to the total variation of image gradient \(\int |\nabla u| \, dx \, dz\). Assuming the Gaussian noise distribution, this is equivalent to minimizing the functional

\[
    \min_u \frac{1}{2} \|z - Qu\|^2 + \lambda \int |\nabla u| \, dy = \arg \min_u \frac{1}{2} \|z - Qu\|^2 + \frac{|\nabla u|^2}{2 \max(\varepsilon, |\nabla u_n|)} \, dy.\]

where \(Q\) is a diagonal operator \(Q : R^2 \to [0, 1]\) representing multiplication by an attenuation factor in each pixel, \(H\) the operator of space-variant convolution\([1, 2]\) \(H : R^2 \to R^2\), and \(z : R^2 \to R^+\) is the observed image we obtain from the image sensor. The attenuation factor \(Q\) is used here if the focal swept image has not been flat-field corrected by the method described in section IV.C.

There are many numerical methods to minimize (V.2). We have chosen an iterative solution, where in each iteration, the functional is replaced by its quadratic majorant

\[
    u_{n+1} = \arg \min_u \frac{1}{2} \|z - Qu\|^2 + \lambda \int \frac{|\nabla u|^2}{2 \max(\varepsilon, |\nabla u_n|)} \, dy.\]

(V.3)
The absolute value in the denominator of (V.3) is constrained to be above $2\varepsilon$ to avoid division by zero. This measure can be also interpreted as replacing the non-differentiable absolute value in (V.2) by the Huber function.

The EDOF image resulting from the implementation of this algorithm on the experimentally obtained laboratory data is shown in Fig. V.7. These preliminary results show the algorithm is mathematically operational and produces deblurred reconstructions of the focal swept images in Fig. V.6. When compared to the swept focal images, the near-axis image depth-of-focus is improved and the off-axis image shows improved lateral resolution at a broader range of depths. However, the objective is to demonstrate an improved depth-of-focus when compared to the static images in Fig. V.3. The $0^\circ$-$20^\circ$ reconstructed image does seem to have sharper edges, but upon closer inspection there is no resolution gain over the range of depths. The $40^\circ$-$60^\circ$ reconstructed image is much worse than the image taken with a static focus. Three possible causes as to why the current system is not performing up to its expectations include the following reasons. Repeatability of the actuated system can directly influence the reconstructions because the PSF measurements must coincide with the focal swept data. Since the two are not measured simultaneously, this can be a cause for concern. A second possible cause is the large number of pixels over which the PSF extends in detector space. The size of the PSF is over 40 pixels for the on-axis image and over 120 pixels for the off-axis image. This leads to complications when implementing the algorithm. A third possible cause can be the amount of noise in the imagery. As with a large PSF, noise leads to severe errors in the solution of inverse problems such as the one formulated here. The noise introduced into the system is caused by the need to capture long exposure images in order for the actuation system to produce repeatable motion. The shortest feasible exposure time in our current system is on the order of 2 seconds. This produces imagery with substantial noise causing the EDOF results to suffer. A future prototype design involves the use of a voice coil actuator in order to induce a sinusoidal oscillation of lens so that the integration time of the sensor can be brought down to tenths of a second or less. During this short exposure time the focus would oscillate multiple times enabling
V.D Discussion & Conclusion

We demonstrated how flat-field correction proves exceptionally beneficial for fiber coupled imaging systems since it corrects for obscurcation and non-uniformity in the response of the lens, fiber bundle, and sensor. More generally, we have described a systematic process for reducing image degradation for fiber bundle image transfer between non-planar surfaces, a technique useful for multiple applications that in some cases would otherwise be impractical. We supported this idea by correcting experimentally collected outdoor images, and showing that this step is essential in removing fiber bundle artifacts and correcting vignetting effects. We used this methodology to generate wide field-of-view fiber coupled images with less distortion than those produced by traditional retro-telephoto “fish-eye” lens imagers. We went on to establish a framework to create spatially variant extended depth-of-focus images using single exposure images taken synchronously with a spatially variant EDOF system that is less susceptible to noise.
mechanical focal sweep. In order to investigate the problem and process the raw focal swept images, we performed an in-depth study of the dynamic behavior the PSF exhibits as a function of both scene angle and depth. We found that near the optical axis (which is along the direction of the focal sweep) the PSFs were depth independent; off-axis this was depth invariance did not hold. We presented a mathematical framework to process spatially variant extended depth-of-focus images using focal swept image and PSF data. We presented preliminary results which remains a topic of ongoing research, possibly improved through more accurate motion control and lower SNR. A more sophisticated system could incorporate a depthmap captured using a primary or secondary system in order to process the EDOF imagery.

**Acknowledgements**

This work was supported by the DARPA SCENICC program under contract W911NF-11-C-0210. Michal Šorel was partially supported by the Grant Agency of the Czech Republic under project 13-29225S.

Chapter V, contains material published in the following, of which the dissertation author was the primary investigator and author:

V Bibliography


Chapter VI

Platform Motion Blur Image Restoration System

VI.A Introduction

Image degradation due to platform motion blur is a common problem for airborne and space-based imagers, where object motion is often relatively negligible [1, 2]. It is also a problem when using camera’s in low light conditions [3], especially when taking hand-held pictures from cell phones, moving vehicles, or point-and-shoot cameras. Motion blur is accentuated by long exposure times or more pronounced motion. Optical sensors fundamentally restrict light intensity and integration time to adhere to fundamental exposure requirements. In general, the motion that causes blur, and therefore the point spread function (PSF), is unknown, and varies across the image.

Deconvolution methods that deblur images without being provided the PSF are called blind deconvolution methods. Deblurring images in software using these methods [4] can reduce both spatially invariant [5, 6, 7, 8] and spatially variant (SV) [9, 10, 11, 12] motion blur. Multiple images with different exposures can be used for this purpose, again for both space-invariant [13, 14] and space-variant blur [15]. In recent years, many approaches were proposed that facilitate blind deconvolution using various alternative optical designs, such as coded aperture [16, 17]
Figure VI.1: Platform motion deblur system schematic. A lens and beamsplitter form identical image planes for the image sensor and position sensing detector (PSD) array to provide data for image restoration.

or wavefront coding [18]. Prototypes that linearly accelerate the image sensors mechanically were built to successfully leverage one- [19] and two-dimensional [20] parabolic coded exposures; however, they only apply to invariant blur in predetermined directions of motion. For an introduction to the restoration of blurred images, see [21]. For other methods in the fast expanding field of computational photography and light-field imaging, see survey papers [22, 23].

Limitations of blind deconvolution methods can be eliminated by measuring the PSF during image acquisition using additional hardware. Camera movement has been measured using accelerometers to produce an SV PSF [24], with a disadvantage that the resolution of the accelerometer does not scale with the focal length of the lens. Another solution is to estimate the PSF using an additional low spatial resolution, high temporal resolution video camera [25, 26]; however, this increases processing costs arising from the need to compute a PSF estimate from the video data. The approach proposed in this paper uses alternative hardware to achieve the same goals.

Platform motion blur varies across the image field in a structured fashion, constrained by only six degrees of freedom corresponding to translation and rotation of the camera. This structured nature of platform motion blur makes it
sufficient to detect motion from the image itself at a few locations in order to generate an SV PSF. A position sensing detector (PSD) placed in the image field for this purpose can maintain high motion tracking accuracy provided a bright feature on a dark background remains incident on its surface. Simulation and testing of the PSD motivated the construction of a prototype, Fig. VI.A, which simultaneously acquires images while tracking motion at specific image locations using a PSD array. These image motion measurements are used to form an SV PSF from which to significantly restore blurred images. By leveraging its larger photon energy collection area, the PSD far exceeds the temporal resolution of image sensors, allowing for the fast collection of image motion information. Since the PSD in the prototype acts on the image itself, its resolution is independent of the object distance.

VI.B Measuring Image Motion Using PSDs

A lateral effect PSD is essentially a single photodiode capable of tracking intensity centroid motion using the lateral photo-effect [27, 28, 29]. Light incident on the PSD generates a photo-current (shown in green) that flows across the uniformly resistive sensor surface (Fig. VI.2). The amount of current that is drawn from each bottom electrode (blue) or that flows to each top electrode (red) is inversely proportional to the distance between the incident light and the particular electrode. The \((X, Y)\) position of a spot of light on the sensor surface is given by

![Lateral photoeffect of a PSD.](image)

**Figure VI.2**: Lateral photoeffect of a PSD. Differential currents from input (blue) and output (red) electrodes produce sum (VI.2) and position (VI.1) data.
(VI.1), where \( L \) is the distance between input or output electrodes and \( i \) are the currents at the respective electrodes. A PSD position output value of zero volts references to the sensor center and can be calibrated using manufacturer circuitry. For more complex light projections incident on the sensor surface, the lateral effect acts on the superposition of the points that make up the projection. This results in the device tracking the centroid of the illumination. The sensors also provide an \( S \) intensity (VI.2) output, which is proportional to the net optical intensity at the sensor’s surface. Tracking is not possible once the feature moves off a detection region. This event is revealed by the PSD outputs as position reaching the sensor edge or as a change in intensity. A pathological case, which was simulated but does not appear in our experiments, occurs when one bright feature moves on the sensor while another moves off. This produces a position tracking error value whose magnitude can be twice the size of the sensor; unfortunately, the PSD outputs would not indicate invalid data for this situation. We address this issue in section VI.H.

\[
X_{PSD} = \frac{i_1 - i_2}{i_1 + i_2} \frac{L}{2} \quad Y_{PSD} = \frac{i_3 - i_4}{i_3 + i_4} \frac{L}{2}
\]  

(VI.1)

\[
S_{PSD} = i_1 + i_2 = i_3 + i_4
\]

(VI.2)

The commercially available lateral effect PSD modules used in our experiments were made by the On-Trak Photonics Corporation (module part number PSM2-4), which contains a Sitek Electro Optics PSD (part number 2L4SP). Each PSD has an active area of 16 mm\(^2\) which corresponds to an area of more than 200,000 pixels in the image sensor. This affords the sensor more energy to make tracking proportionally faster. The lateral effect PSD linearity, sensitivity [30], and transfer function [31] have been characterized. Its electronic and mechanical behavior and its application to spectroscopic optical systems [32] as well as tracking and displacement sensing [33] has been studied in depth. Specifications of the experimental lateral effect PSDs are given in Table VI.B and were verified by the manufacturer and as well as experimentally. It is possible that for custom PSDs be tailored to operate with high sensitivity or outside the visible band.
Table VI.1: On-Trak PSD (PSM2-4) specifications

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling rate</td>
<td>15 kHz</td>
</tr>
<tr>
<td>Position accuracy</td>
<td>100 − 200 nm</td>
</tr>
<tr>
<td>Position linearity</td>
<td>99.7%</td>
</tr>
<tr>
<td>Size</td>
<td>4 mm × 4 mm</td>
</tr>
<tr>
<td>Spectral response</td>
<td>400 − 1100 nm</td>
</tr>
<tr>
<td>Required incident light</td>
<td>2.5 µW − 2.5 mW</td>
</tr>
<tr>
<td>Position output</td>
<td>−10 − 10 V</td>
</tr>
<tr>
<td>Sum output</td>
<td>0 − 6 V</td>
</tr>
<tr>
<td>Valid sum output</td>
<td>1 − 6 V</td>
</tr>
</tbody>
</table>

VI.C PSD Image Motion Tracking Simulation

We modeled lateral effect PSD behavior as images move across its surface in order to find how well it can measure the PSF. We used the definition of centroid position \( (X_{\text{centroid}}, Y_{\text{centroid}}) \) given by (VI.3) to find the true centroid of images prior to simulated motion. Then this fixed centroid underwent the same SV movement as the images. This served as a benchmark for accurate centroid tracking. The total intensity \( (S_{\text{centroid}}) \) is given by (VI.4). Here \( I(X,Y) \) is the optical intensity distribution at the sensor surface:

\[
X_{\text{centroid}} = \frac{\int X I(X,Y) \, dX \, dY}{\int I(X,Y) \, dX \, dY}, \quad Y_{\text{centroid}} = \frac{\int Y I(X,Y) \, dX \, dY}{\int I(X,Y) \, dX \, dY},
\]

\[
S_{\text{centroid}} = \int I(X,Y) \, dX \, dY.
\]

In simulation, we then created frame-by-frame motion of the images using a known rotation and translation trajectory. A highly localized image (Fig. VI.3(a)) was blurred in simulation using SV movement. A video depicting the motion blur was used to calculate tracking error by comparing accurate motion (VI.3) with a PSD tracking model (VI.1). Gaussian noise with a positive mean value representing background illumination was included in the input images. Images used in simulation are shown in Figs. VI.3(b) and VI.3(c) without and with noise, respectively. In this way we quantified how well the PSD tracks images under varying levels of background illumination. The simulation showed that localized features on a dark
Figure VI.3: Scenes input into PSD image motion tracking simulation to characterize the effects of background illumination. (a) Ground truth input image. (b) SV blurred image. (c) SV blurred image with Gaussian noise ($\sigma = 0.1$) with a positive mean value representing background illumination (5% of the maximum brightness).

The motion blurred image $z$ can be modeled by the linear operation

$$z(x, y) = [Hu](x, y) = \int u(s, t)h(x - s, y - t; s, t)dsdt,$$

where $u$ is the unblurred image scene and $h$ the SV PSF. We can think of this operation as a convolution with a PSF that varies depending on pixel location $(x, y)$.
Figure VI.4: Simulation results: Plot of PSD tracking error for various levels of background illumination showing the sensor is only capable of tracking image features much brighter than the background. Percentages represent the ratio between the brightness of background and that of the tracked spots.

Figure VI.5: PSD centroid tracking of identical image movement is skewed toward the sensor’s center as background illumination increases. (a) Motion estimate from Fig. VI.3(b) with no background illumination and (b) from Fig. VI.3(c) with background illumination at 5% of the tracked spots.
within the image. We denote the operation (VI.5) in short form as $Hu$. Note that the PSF is a more general construct that can also represent other complex image degradations that depend on spatial or temporal coordinates, such as motion blur, lens distortions, out-of-focus blur, and atmospheric blur.

In the previous sections, we explained how the PSD can be used to estimate local image motion. The PSD output voltage signal represents the point location of the centroid of optical intensity at one instant in time. We can rapidly track these elementary shifts in time to get the PSF. This is valid for the pixel whose location in the image plane coincides with the PSD. In our experiments, we used three PSDs to give us the PSF for three different image coordinates. However, to be able to apply our model (VI.5), we need to know the PSF $h$ everywhere, for an arbitrary pixel coordinate.

Fortunately, there are sufficiently precise simplifications that allow PSD measurements to estimate an SV PSF for platform motion blur in most practical situations. One approach would be to consider only camera rotation, which was shown to be a good approximation for pictures taken by hand [12, 34, 10, 24]. In this case, the PSF does not depend on the depth map and we can recover the change of the camera orientation from the motion of only two points [35]. Information from more than two PSDs can be used to make the estimate more robust.

In our experiments, we adopt an alternate approach, which assumes that the image motion within one elementary step can be modeled by the affine transform

$$
\begin{bmatrix}
x' \\
y'
\end{bmatrix} = \begin{bmatrix}
a(t) & b(t) \\
d(t) & e(t)
\end{bmatrix} \begin{bmatrix}
x \\
y
\end{bmatrix} + \begin{bmatrix}
c(t) \\
f(t)
\end{bmatrix}.
$$

(VI.6)

Motion blur depends on the spatial location $(x, y)$ within the image as well as on time $t$. Rotation, scale, and shear manifest itself in the coefficients $a$, $b$, $d$ and $e$, while translation is manifested in the $c$ and $f$ coefficients. The use of the affine transform can be justified by the fact that it includes as a subset in-plane translation (translation in x and y-axes) and rotation about the optical axis, which is appropriate for many real scenes [11]. Computation can be made more robust by enforcing the transform to be rigid using singular value decomposition [36].

Coefficients of the transform, for each elementary time step $t$, are calculated
using multiple regression solved by the method of least squares. This calculation is
based on consecutive (in time) voltage information \((V_x(t), V_y(t))\) (Fig. VI.6) from
PSDs placed in known locations within the image plane. At one time instance,
we collect two position voltage samples \((X, Y)\) from each of the three PSDs. Each
sample pair of consecutive PSD voltage data is mapped into pixel space to cre-
ate its corresponding pixel information \((P_x(t), P_y(t))\); see Appendix VI.J. Each
sample pair of consecutive PSD pixel data is used to fit the affine model (VI.6)
forming a set of affine model solutions, one for each sample pair of consecutive
data. Each solution describes the motion that occurred from one time step to the
next. These affine models are in turn is used to generate the SV PSF (Fig. VI.7).
The generation of the SV PSF is recursive in that as it is generated from the affine
model, the current solution in time depends on the affine model of the previous
time step. Three PSDs are sufficient for an unambiguous estimate of transform
coefficients. Our prototype system operated in the low noise regime since the en-
ergy collection area of the PSD was large and illumination for tracking was ideal
in our experimental setup; see section VI.F. For this reason, we did not include a
noise model in our coefficient estimation task.

VI.E System Prototype Design

The design space requires the imager and PSDs be in the same image plane
(Fig. VI.A). It requires a long back focal distance lens in order for the beam splitter
to occupy part of the optical path. Commercial cameras don’t meet this require-
ment without modification and therefore we chose to build a system that uses two
lenses that have short back focal distances. Similarly, two-dimensional closely tiled
PSD arrays are not commercially available, so we met system requirements using
three discrete PSD’s.

We introduce a custom cube assembly fabricated to house the beam splitter,
image sensor and PSDs (Fig. VI.8). Two identical lenses are used to form identical
image planes. A motion stage creates reproducible motion. A custom plate mount
holding the PSDs is supported by an adjustable sliding rod configuration that
Figure VI.6: PSD data at specific locations in the image field after calibration (Appendix VI.J). These local PSF measurements are used to construct an SV PSF. Their relative location is given in Figs. VI.10 and VI.9.

Figure VI.7: PSF specific to pixel \((x, y)\) is recursively generated using affine model coefficients \(a(t) - f(t)\) as the elementary time step \(t\) is integrated over the sampling time.
allows for focusing of the PSDs. The computer controlled system simultaneously triggers the motion stage, imaging camera, and PSD voltage acquisition, enabling the collection of blurry images and PSD data. Input parameters include aperture, ISO and exposure time for the camera, sampling rate, and number of samples for the PSD array and position, velocity, and acceleration for the motion stages. Misalignment due to the physical mounting of the PSDs to the plate is accounted for in the calibration method (Appendix VI.J).

VI.F Experimental Setup

The laboratory experiments we conducted involve a starfield and a color image scene (Figs. VI.9 and VI.10). The images exhibit significant SV motion blur that exceeds 100 pixels at some image locations (Fig. VI.9). Both figures show the region of the scene that is to be restored; it is depicted enclosed in a red dashed line. The starfield scene is made up of dimly lit LEDs, randomly located within this region. Both scenes contain three bright LED sources focused onto the region of the image field where the PSDs are located. In this way we provide the
Figure VI.9: Experimentally blurred color image. Magnified sections of the image confirm SV blur. The calculated SV PSF is shown superimposed (yellow lines) and is consistent with the image motion blur. The PSD’s relative locations are shown in cyan. The portion of the image that is to be deblurred is enclosed in red.

PSD’s localized features with relatively little background illumination for tracking purposes. The physical scale and location of the PSDs in relation to a commercial full frame image sensor in our experiments are shown superimposed (in cyan) in Figs. VI.9 and VI.10. For all experiments performed, 30,000 sample voltage pairs \((X, Y)\) are taken from each PSD during the two second acquisition of a single image. Overlaying the blurred image in Fig. VI.9 is the SV PSF (yellow lines) describing the calculated motion blur. Similarly, Fig. VI.10 has super imposed green lines which describe the calculated SV PSF. Both figures show that the SV PSF is consistent with the image blur.

VI.G Image Deblurring: Prototype Results

As soon as we have a reliable estimate of the SV PSF, we can deblur the image using one of many known image restoration algorithms. The modern theory of image restoration is based on Bayesian statistics and algorithms usually seek the maximally probable solution, which is equivalently formulated as a minimization of certain functionals. We used a de facto standard solution, which can be expressed as the minimum

\[
\min_{u} \left[ \frac{1}{2} \|z - Hu\|^2 + \lambda \int |\nabla u| dx dy \right],
\]  

(VI.7)
Figure VI.10: Experimentally blurred starfield image (contrast reversed). The calculated SV PSF is shown superimposed in green and is consistent with the SV image motion blur. The PSD’s relative locations are shown in cyan. The portion of the image that is to be deblurred is enclosed in red.

where $z$ is the blurred observed image, $H$ is the SV PSF, $u$ is the unknown sharp image, and $\lambda$ a positive regularization constant, usually set empirically. The left term penalizes discrepancy between model and measurements, the right term is a so called regularization term, which serves in the Bayesian framework as a statistical prior. In our case we utilize total variation, a regularization technique that exploits the sparsity of image gradients in natural images. Minimizing the convex functional (VI.7) is now considered a standard way to achieve close to state-of-the-art quality of restoration without excessive time requirements [4]. There are a number of methods to minimize the functional, many of them mathematically quite involved. We refer interested readers to a report that summarizes most of the latest developments [37]. We used an efficient method [38] solving (VI.7) iteratively as a sequence of quadratic functionals,

$$ u_{i+1} = \arg \min_u \left[ \frac{1}{2} \| z - Hu \|^2 + \lambda \int \frac{|
abla u|^2}{2|\nabla u_i|} + \frac{|
abla u_i|}{2} dxdy \right]. $$

(VI.8)

Functional (VI.8) bounds the original function (VI.7) from above and has the same value and gradient in the current estimate $u_i$, which leads to provable convergence.
to the global minimum of (VI.7). To solve (VI.8), we used the conjugate gradient method [39]. For details, see Section 3 of [38]. The description is relatively accessible as this method does not use more complex results from convex analysis, such as Fenchel’s duality and Moreau’s theorem, which are necessary in primal-dual methods [37]. One example result of the procedure (VI.8) can be seen in Fig. VI.12(c).

The main problem of any algorithm working in our scenario with the SV PSF (VI.5) directly is that its computation is very time consuming and takes a lot of memory. To get an idea, consider our example; a 0.25 megapixel image, 2 second exposure time with the PSD sampling at 15 kHz. Generation of the PSF at each and every pixel requires us to fit an affine or rigid transform \(7.5 \times 10^9\) times. Each model fit is a nontrivial operation in itself that involves inversion of a small matrix or an singular value decomposition. If the size of the PSF is about \(25 \times 25\) pixels and it is stored in 4-byte floating-point format, the PSF requires \(25 \times 25 \times 4 \times 250000 = 625\) MB. Even if it is stored in a sparse form, it takes at least 60MB of memory. Common present-day cameras produce 10-20 megapixel images, which gives us 40-80 times larger figures.

To make the problem tractable, we used bilinear interpolation to both save memory and speedup computation of the PSF [40]. This idea was used in practical deblurring algorithms [15, 10, 41] and recently for super-resolution in [42]. The interpolation relies on the fact that blur is caused by camera motion, where the PSF changes very slowly with position in the field of view. The PSFs are computed only on a regular grid of positions and the values of the PSF at intermediate positions are estimated with reasonable precision by bilinear interpolation of four adjacent known PSFs [40]. In our example in Fig. VI.12(c) we considered \(15 \times 15 = 225\) positions and so we computed this number of PSFs. This number was chosen so that using more positions did not improve the quality of restoration significantly. Indexing any four adjacent grid points as \(i = 1\ldots 4\) (starting in the top-left corner and going clockwise), the SV PSF in the rectangle among them is defined as

\[
h(s, t; x, y) = \sum_{i=1}^{4} \alpha_i(x, y) h_i(s, t), \tag{VI.9}
\]
where $\alpha_i$ are the coefficients of bilinear interpolation. Let us denote $x_1$ and $x_2$ minimum and maximum x-coordinates of the rectangle, respectively, and analogously $y_1$ and $y_2$ in the y-coordinates. Using auxiliary quantities

$$t_x = \frac{x - x_1}{x_2 - x_1}, \quad \text{and} \quad t_y = \frac{y - y_1}{y_2 - y_1},$$

(VI.10)

the bilinear coefficients are

$$\alpha_1 = (1 - t_y)(1 - t_x), \quad \alpha_2 = (1 - t_y)t_x, \quad \alpha_3 = t_y(1 - t_x), \quad \alpha_4 = t_yt_x.$$  

(VI.11)

Space-variant convolution can be then computed as a sum of four convolutions of the image weighted by coefficients $\alpha_i(x, y)$

$$[Hu](x, y) = \int u(s, t)h(x - s, y - t; s, t)dsdt$$

(VI.12)

$$= \int u(s, t)\sum_{i=1}^{4} \alpha_i(s, t)h_i(x - s, y - t)dsdt$$

(VI.13)

$$= \sum_{i=1}^{4} \int (\alpha_i(s, t)u(s, t)) h_i(x - s, y - t)dsdt$$

(VI.14)

$$= \left[ \sum_{i=1}^{4} \alpha_i u \ast h_i \right](x, y).$$

(VI.15)

All first-order minimization algorithms also need the operator adjoint to $H$ (space-variant counterpart of correlation)

$$[H^*u](x, y) = \int u(s, t)h(s - x, t - y; x, y)dsdt$$

(VI.16)

$$= \int u(s, t)\sum_{i=1}^{4} \alpha_i(x, y)h_i(s - x, t - y)dsdt$$

(VI.17)

$$= \sum_{i=1}^{4} \alpha_i(x, y) \int u(s, t)h_i(s - x, t - y)dsdt$$

(VI.18)

$$= \sum_{i=1}^{4} \alpha_i(x, y)[u \otimes h_i](x, y).$$

(VI.19)

In the method described above, the adjoint operator is used in the conjugate gradient method in the gradient of data term $\frac{\partial}{\partial u} \frac{1}{2} \| z - Hu \|^2 = H^*(Hu - z)$. Using the operators $H$ and $H^*$, in forms (VI.15) and (VI.19), for large PSFs can be sped
Figure VI.11: Pixel-by-pixel deconvolution. A PSF is generated for the blurred pixel of interest (green). A section in the neighborhood of the pixel is deblurred using the PSF. Only the center pixel (yellow) of the deblurred section is used in the final reconstruction.

up significantly by computing convolutions and correlations using the Fast Fourier transform.

Consider the huge savings we achieved in both time and memory. Memory consumption is now just $4 \times 25 \times 25 \times 225 \approx 560\text{kB}$. While the PSF computation for all pixels took around 12 hours, now it is slightly more than 2 minutes, more than 250 times less time (all algorithms implemented in Matlab running on an ordinary personal computer).

We finish this section with a few remarks on the Lucy-Richardson algorithm [43, 44, 45], which is a well known iterative procedure that can be described as

$$u_{i+1} = u_i H^* \left( \frac{z}{Hu} \right). \quad (VI.20)$$

Because of its simplicity of implementation and speed, it is still occasionally in use for rapid prototyping or for extremely large data in general. Even in this project, we used it initially to prove the concept of the proposed system. It can be applied in several ways. First, pixel-by-pixel in its space-invariant version, where each pixel value of the restored image is acquired by deconvolving in its square neighborhood and taking the central value (Fig. VI.11). Second, in the space-variant version, as proposed initially in [45] with PSF computed in each pixel, and finally using the bilinear interpolation as we did in the total variation approach above. In the first two cases, the computation of the PSF totally dominates (above mentioned
10 hours); the third case is fast, even faster than the total variation approach with bilinear interpolation.

Figures VI.12(d) and VI.12(e) show the Lucy-Richardson approach in its two forms, in comparison with the total variation reconstruction in Fig. VI.12(c). We can see that the pixel-by-pixel Lucy-Richardson algorithm gives slightly better results than with the bilinear interpolation at the expense of computation time. The main reason is not interpolation but that it better suppresses boundary artifacts. The result of the total variation deblurring contains much fewer artifacts than both versions of the Lucy-Richardson algorithm. To better see the differences, readers are encouraged to zoom the results in electronic form of this article. One important property of regularization is that it partially suppresses artifacts caused by an imprecise PSF. Lucy-Richardson algorithm, on the other hand, produces many artifacts even for ideal an PSF and ideal photon noise for which it was derived. One reason is that it converges to a maximum likelihood estimate from only one observation, which is known make the algorithm highly unstable. Table VI.G summarizes the algorithms in terms of speed, memory consumption, and reconstruction quality measured by the root mean square error with respect to a ground truth taken from a motionless camera. Notice that the difference between the total variation deblurring and pixel-by-pixel Lucy-Richardson does not seem so large in terms of root mean square error, but the visual difference is well noticeable. The reason is that visual perception cannot be expressed in such simple form. Alternatives for measuring image quality can be found in [46] and references therein. An additional example of pixel-by-pixel Lucy Richardson image reconstruction is provided in Fig. VI.13(c). This shows that in the simpler case of starfield images the algorithm works very well.

<table>
<thead>
<tr>
<th>PSF Estimate</th>
<th>Deblur Algorithm</th>
<th>Speed</th>
<th>Memory</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blurred image</td>
<td></td>
<td>0.109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pixel-by-pixel</td>
<td>Lucy-Richardson</td>
<td>12 hours</td>
<td>625 MB</td>
<td>0.055</td>
</tr>
<tr>
<td>Bilinear Interpolation</td>
<td>Lucy-Richardson</td>
<td>15 seconds</td>
<td>560 kB</td>
<td>0.082</td>
</tr>
<tr>
<td>Bilinear Interpolation</td>
<td>Total variation</td>
<td>140 seconds</td>
<td>560 kB</td>
<td>0.044</td>
</tr>
</tbody>
</table>
Figure VI.12: The region enclosed in red of the blurred image in Fig. VI.9 is deblurred using (c) bilinearly interpolated total variation regularization, (d) pixel-by-pixel Lucy-Richardson techniques, and (e) bilinearly interpolated Lucy-Richardson. Magnified sections of the images are also shown. (a) A ground truth image is shown for comparison.
Figure VI.13: The region of Fig. VI.10 enclosed in red is deblurred using a pixel-by-pixel method with the iterative spatially invariant Lucy-Richardson algorithm. The (a) ground truth, (b) blurred, and (c) five-iteration reconstructed images are shown for comparison along with magnified sections of the images. Superimposed green lines in (b) show calculated SV PSF.

VI.H Future Directions

Although detecting compromised data using PSDs is not guaranteed, a PSD array could increase the probability of capturing valid data. A PSD array is not a new idea [47, 48], but here we propose forming PSD aggregate groups to improve tracking capabilities. We did not explore this experimentally and would also require a custom fabricated device, which, to the best of our knowledge, has yet to made. A PSD array schematic is shown in Fig. VI.14(a), where valid and invalid tracking of bright features is denoted by green and red arrows, respectively. Motion tracking of image features whose trajectory produces any invalid data is undesirable. Invalid data could be monitored and used to trigger a camera shutter in order to prevent excessive blur. Invalid data from individual sensors may be recovered by joining adjacent PSDs in the array using (VI.21) as shown in Fig. VI.14(b). Here \((X_{\text{aggregate}}, Y_{\text{aggregate}})\) denotes the centroid location over the closely tiled PSD aggregate array. \((X_{\text{PSD}}, Y_{\text{PSD}})\) is defined in (VI.1) and is the centroid position output of a single PSD in the array. \(S_{\text{PSD}}\) is defined in (VI.2).
Figure VI.14: A PSD can track the moving centroid of features that remains on the sensor under low background illumination. They provide position (VI.1) and intensity (VI.2) outputs, which can reveal invalid data. (a) A PSD array provides SV information and increases probability of valid data (green) capture. (b) Invalid data (red) can be recovered by aggregating closely tiled PSDs using (VI.21) to a form sensor group.

$L$ is the distance between electrodes in a single PSD, $(u, v)$ denote the index of PSDs in the array. The pairs $\{l, m\}$ and $\{o, p\}$ index the beginning and ending of the sensor group in the $x$ and $y$ directions, respectively (Fig. VI.14(a)). Piecewise concatenation of the position signal can be used when $l = m$ or $o = p$. As opposed to the discrete PSDs available for the experimental demonstration, an aggregate group requires a two-dimensional array of closely tiled PSDs to create a contiguous measurement space

\[
X_{aggregate} = \frac{\sum_{u=l}^{m} \sum_{v=o}^{p} \left[ (L(u - 1) + X_{PSD}(u, v)) S_{PSD}(u, v) \right]}{\sum_{u=l}^{m} \sum_{v=o}^{p} S_{PSD}(u, v)}, \tag{VI.21}
\]

\[
Y_{aggregate} = \frac{\sum_{u=l}^{m} \sum_{v=o}^{p} \left[ (L(v - 1) + Y_{PSD}(u, v)) S_{PSD}(u, v) \right]}{\sum_{u=l}^{m} \sum_{v=o}^{p} S_{PSD}(u, v)}.
\]

Compensation of background illumination to recover accurate PSF esti-
mates could be made by using image data in the regions where the PSDs are located. This could be computationally expensive because it requires adaptive signal thresholding in order to estimate background intensity levels using the image pixel data. The reason arises from the fact that it is difficult to discern the moving signal from background illumination. Overcoming the need for negligible background illumination may be achieve through the use of gradient centroid sensors, which would operate independent of ambient background light.

Finally, a radial distortion model to compensate for the radial pincushion or barrel distortions can be included along with the affine model. While it is in principle possible to fit a distortion model directly to PSD data, it would require an excessive number PSDs in the measurement space to sufficiently constrain this higher-order model. A better option is to measure distortions in advance (pp. 189-193 in [35]).

VI.I Conclusion

We introduced a proof-of-concept computational imaging system prototype that incorporates optical position sensing detectors (PSD), a conventional camera and a method to reconstruct images significantly degraded by SV platform motion blur. The system exploits the large energy collection area of the PSDs to make fast analog centroid position measurements to track light distributions on its surface. It leverages more energy collection than a single pixel since it has a larger area making it proportionally faster. This affords it high temporal resolution as it measures the PSF at a specific location in the image field. Data from a few PSDs at known locations in the image field is mapped into pixel space and then used to fit an affine model valid for platform motion. Using the information from the model, we recursively generate an SV PSF for any pixel in the image enabling us to deblur images using a pixel-by-pixel deconvolution. We also show an alternate approach that uses a bilinearly interpolated PSF in conjunction with total variation to improve restoration quality and speed up computation.

The image sensor and PSD array both operate directly on the image it-
self, making the system resolution independent of object distance. The system hardware and algorithm functionality can be incorporated into the form factor of modern cameras and serves as a new direction in computational imaging. To improve the accuracy of the PSF estimate a spherical lens distortion model can be used alongside the affine model during its calculation. The PSD is designed to track laser beams and is therefore insensitive to light. This limits the system applications to sparse imager such as where bright features of specific size appear on a dark background. Although not explored experimentally, we suggest that closely tiled PSD arrays used along with grouping algorithms can make the PSD technology more compatible with general scenes. Interlacing PSD technology with the pixels of a custom image sensor has been achieved and can make these systems more compact and practical. As an alternative way to resolve the limitation of PSF estimation error, we propose the use of a not commercially available gradient centroid sensor.

VI.J Appendix: Calibration - Mapping of Voltage to Pixels

To calibrate the PSD module, data is taken from the system in the absence of motion with an LED focused on the PSD and imager. The PSD voltage output and the acquired image pixel information correspond to the centroid of the LED in the scene. In this way PSD voltages are paired up with image pixel data (Fig. VI.15(a)). This is done at two image locations. The voltages and pixel values are used to create the vectors, shown schematically in Fig. VI.15(b). These vectors are not necessarily collinear due to misalignment during the mounting of the sensors. The magnitudes of the two vectors and the angle between them are used to serve as calibration parameters, which are valid across the image plane (VI.22).

\[
\begin{bmatrix}
P_x \\
P_y
\end{bmatrix}
= \frac{|P|}{|V|} \begin{bmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix}
\begin{bmatrix}
V_x \\
V_y
\end{bmatrix}.
\]  

(VI.22)
(a) Sensors at image plane.  
(b) Calibration vectors.

**Figure VI.15:** (a) Depiction of a PSD and imager’s physical location in the image plane. PSD voltage and imager pixel information is paired at two separate image locations and used to form calibration vectors. (b) Depiction showing vectors whose angle and magnitude are used to map voltage data into pixel data.

**Acknowledgments**

The authors thank Robin Dawson for technical discussion. This work was supported by the Charles Stark Draper Laboratory, contract SC001-000000628A. Work done on this project by Michal Šorel was partially supported by the Czech Ministry of the Interior, Grant No. VG20102013064.

Chapter VI, contains material published in the following, of which the dissertation author was the primary investigator and author:


VI Bibliography


Chapter VII

Characterization of a Compressive Imaging System

VII.A Introduction

Traditional cameras capture a full resolution image, then process and compress this information for storage, typically retaining 1% of the full (raw) image data. Even without a priori knowledge of the scene, perfect reconstruction can be possible with sampling below the Nyquist limit [1, 2, 3]. Compressive imaging (CI) systems acquire images, or other optical scene information through a series of measurements of the total energy transmitted through image plane spatial filters (basis functions), using the minimum number of measurements required to achieve the desired image quality or, more generally, the feature-based scene information [4, 5, 6, 7]. CI systems are of great potential interest for specific applications where focal planes with the desired spectral or spatial resolution are unavailable, or where adaptive feature-based sensing may operate with few measurements. Before undertaking these exotic applications; however, it is useful to investigate the performance of CI systems, where they can be directly compared to images acquired by conventional focal planes. This paper provides the first, to our knowledge, experimental comparison test of a CI system, where we explore the impact of alternative basis functions (digital, continuous, and random) on binary, grayscale, and natural light
scenes, and compare the resulting images to those acquired by a conventional focal plane imager.

The paper is organized as follows. In Section VII.B, we briefly summarize CI theory and discuss camera image capture in relation to sampling theory. In Section VII.C we describe the experimental apparatus and methods to reconstruct the images using experimentally collected data. Section VII.D explains the hardware configuration for grayscale filtering. In Section VII.E we discuss system operation and the experimental setup under which the compressive images were taken. We then present results and conclude with a discussion of their significance.

VII.B  CI and Image Sampling Theory

Any signal (or image), \( x \), can be expressed as a weighted sum of basis functions following the equation

\[
x = \sum_{n=1}^{N} [\theta(n) \psi_n].
\]

Here \( \psi_n \) are the basis functions, \( \theta(n) \) are the weighing coefficients, and \( N \) is the total number of pixels. The number of basis functions needed to completely describe the image space is necessarily equal to the number of pixels in the image. Nyquist theory states that sampling at a rate faster than twice the highest frequency contained in the signal guarantees perfect reconstruction and proves beneficial when sampling band-limited signals. However, many images are “sparse” and can be completely represented using fewer samples, provided an appropriate basis set is used. For this reason, transformation between basis sets is useful and has become commonplace [8].

Compression algorithms decorrelate image signals into a small set of orthonormal basis coefficients leveraging the sparse representation of images. This is the premise of lossless image compression used in many cameras today, where

\[
x = \sum_{l=1}^{K} [\theta(l) \psi_l]
\]

even though \( K < N \). Karhunen-Loève expansions completely decorrelate a signal into its most sparse basis representation, maximally compacting the energy (information) contained in the signal. Reduction in dimensionality allows for more efficient use of memory and data transfer. It is possible to compress the information content further at the cost of less accurate reconstruction. For example, JPEG uses discrete cosine transforms (DCT) and JPEG2000 uses
wavelet transforms as efficient basis representations for lossy or lossless compression. In most applications; however, there is no a priori knowledge about the image scene and therefore its bandwidth or decorrelation, which in turn limits the implementation of this rich theory to post processing.

Modern cameras compress images by first acquiring all the pixel values (basis coefficients), then transforming them into another basis set, discarding the smaller coefficients and finally encoding the remaining coefficients. This is inefficient since samples and analog to digital (A/D) conversions are performed for coefficients that are ultimately unused. It would be ideal to sample only the values needed for perfect reconstruction. The basis set these traditional cameras use to sample the image space is a bandlimited two dimensional comb function, with one impulse per pixel. This sampling method does not statistically meet the needs of the ensemble of natural images, whose spectra follows a distribution inversely proportional to the spatial frequency \([9, 10, 11, 12]\). This spectral distribution follows \(S(f) = f^{-\gamma}\), where \(f\) is the spatial frequency and \(\gamma\) is an empirical constant satisfying \(1.8 < \gamma < 2.3\). The imager can be tailored to the statistical distribution of most images by sampling lower spatial frequencies more densely.

Donoho, Candès, Romberg and Tao \([1, 2, 3]\) have shown that if a signal is sparse (compressible) in one basis (called the reconstruction basis), perfect reconstruction is guaranteed provided at least \(\Omega > \frac{T \cdot (\log N)}{C_M} \cdot \log(\log N)\) samples are measured using a basis incoherent to this reconstruction basis. Here \(\Omega\) is the number of measured samples, \(T\) are the sparse basis coefficients, \(N\) is the total number of pixels, \(\frac{N-T}{N}\) is the degree of compressibility, and \(C_M\) is a constant which depends on an accuracy parameter \(M\). If the measurement basis is random, then with overwhelming probability it is incoherent with arbitrary basis sets, satisfying the conditions above. In other words, CI does not seek to measure the \(T\) sparse basis coefficients directly; instead, it seeks to measure \(\Omega\) random basis coefficients. These random measurements are democratic in that all samples have an equal probability of capturing significant data \([13]\). Since the number of measurements does not describe the image space completely, image recovery is an ill-posed problem. Yet, if the image is sparse and the measurement basis satisfies the restricted isometry property (RIP),
then an $L_1$ solution to the basis pursuit problem satisfies the reconstruction as well as the $L_0$ solution [4]. The $L_1$ minimization is less complex than $L_0$ minimization; it arrives at the solution faster and provides flexibility in numerical optimization [14, 15].

Sampling with both a tailored spatial frequency distribution and compressive sensing (CS) techniques is fundamentally different than sampling in modern cameras. They all sample image projections on a basis set, but the basis sets used in CI need not be impulsive, and can extend across the image space to collect a spatial superposition of image intensity of the scene. This idea of measuring intensity superposition in the image space makes compressive imagers advantageous when samples are scarce or expensive, or when the image data can be accurately represented by a relatively small number of measurements in any basis set.

VII.C System Design and Configuration

CI requires that the intensity values of a scene be spatially multiplexed and summed, using specific basis patterns, to form a corresponding set of basis coefficients. The measurable coefficients are formed by the spatial dot product of the scene intensity distribution with each basis function. Several hardware architectures for CI have been proposed [16]. The basis projections can be sampled by imaging the scene onto the spatial light modulator (SLM), then condensing the resulting energy onto a single detector (optical summing). Alternatively, a two-dimensional array image sensor with spatially weighted sensitivity and analog summing could implement the same operation (electronic summing) [17, 18]. Creating an image sensor (focal plane array) that can directly operate as a compressive imager is possible; however, fabricating one with high resolution, optical fill-factor and sensitivity is challenging, while high resolution SLMs are commercially available. The benefits of single photodetector compressive imagers were enumerated in [4]. We chose to use this system configuration since it demonstrates the extreme case of CI, which should provide the most informative performance comparison.

To characterize single photodetector CI system performance as a function
Figure VII.1: Schematic of compressive imaging system. A scene is imaged onto a digital micromirror device (DMD\textsuperscript{TM}), which serves as a spatial light modulator (SLM). A sequence of specific basis patterns are emulated by the DMD\textsuperscript{TM} mirrors, and the filtered signals are collected by a second lens on to a single photo-detector. A sink blocks undesired deflected stray light.

of encoding basis sets, we built the test-bed system similar to those of previous works [4, 5, 6, 7] (shown schematically in Fig. VII.1). The input scene is imaged onto a 1080p Texas Instruments digital micro-mirror device (DMD\textsuperscript{TM}) [19], so that each mirror element can direct the light from the scene into one of two direction paths, separated by $\pm 12^\circ$. The DMD\textsuperscript{TM} specifications are given in Table VII.C. The positive portion of the transform is projected in the $+12^\circ$ direction and the negative portion in the $-12^\circ$ direction. We use a condenser lens to collect all the optical energy from the positive path onto a single-element photodetector (with noise equivalent power $NEP = 0.2 \, pW/\sqrt{Hz}$). The $12^\circ$ deflection angle of the DMD\textsuperscript{TM} mirrors imposes a requirement that the lenses have an $F/# \geq 2.573$, following the relation $F/# = \frac{1}{2\tan \theta} = \frac{f}{d}$, where $\theta$ is the maximum collection angle, $f$ is the focal length of the lens, and $d$ is the diameter of the lens. Optomechanics must also be taken into account when satisfying this angular requirement. We chose to use a Fujinon F/9 double Gauss Copal lens with a focal length of 180 mm. The system was operated both in the lab with a back illuminated object
(transparency) and outdoors with a natural daylit scene.

**Table VII.1:** Texas Instruments high definition 0.95 1080p DMD™ Specifications.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mirror Array Size</td>
<td>1920 × 1080</td>
</tr>
<tr>
<td>Number of Mirrors</td>
<td>2,073,600</td>
</tr>
<tr>
<td>Mirror Pitch</td>
<td>10.8 µm</td>
</tr>
<tr>
<td>Fill Factor</td>
<td>91%</td>
</tr>
<tr>
<td>Active Area</td>
<td>20.736 mm × 11.664 mm</td>
</tr>
<tr>
<td>Diagonal</td>
<td>23.8 mm</td>
</tr>
<tr>
<td>Deflection Angle</td>
<td>±12°</td>
</tr>
<tr>
<td>Settling Time</td>
<td>12 µs</td>
</tr>
<tr>
<td>Mirror Overshoot</td>
<td>1.5 µm</td>
</tr>
<tr>
<td>Rebound Deflection</td>
<td>2°</td>
</tr>
<tr>
<td>Data Rate</td>
<td>51.2 Gb/s</td>
</tr>
<tr>
<td>Binary Patterns per Second</td>
<td>24,690</td>
</tr>
<tr>
<td>Clock Rate</td>
<td>400 Mhz (2× LVDS)</td>
</tr>
</tbody>
</table>

Since we only sample from one output path of the DMD™, it is necessary to split each pattern into a sequence of two complementary patterns, one containing the negative portion and one the positive portion. For each basis pattern, two PNG image files are generated and stored to the system computer’s hard-drive. Photo-diode measurements are taken as the DMD™ displays the patterns from the corresponding basis set. In order to obtain each coefficient, measurements taken from the negative portion of the basis are subtracted from measurements taken from the corresponding positive portion of the basis. In this way, the coefficients of the basis set are collected to later reconstruct the image scene. Each PNG file’s resolution is 1920 × 1080, corresponding to the DMD™ specifications, with the actual transforms occupying the center 1024 × 1024 mirrors for a total of 1,048,576 mirrors. This determines the maximum resolution of the system, which is about 1 megapixel. A photo of the experimental system is shown in Fig. VII.2. The DMD™ and photodetector were configured in the system as follows. A computer running the Digital Light Innovation’s Accessory Light-modulator Package (ALP) software was used to upload previously generated PNG transform basis image files to the Discovery D4100 controller board via a USB 2.0 link, which in turn controls the Texas Instruments DMD™ chip. A Newport Optics Corporation
photodetector 918D-SL was used to measure light intensity collected from the DMD™ and send it to a Newport power meter 1936-C, which was sampled using a National Instruments PCIe-6363 data acquisition card in a desktop computer. We synchronized the sampling of the basis coefficients with the displaying of each basis projection by feeding the trigger from the D4100 controller board to the data acquisition card.

Using the data obtained from the CI system, we formulate an underconstrained problem which has an infinite number of solutions, $b = Ax + z$. Here $A$ is the sampling basis set, $z$ is the noise term, $b$ are the measured basis coefficients and $x$ is the signal (image) we reconstruct. A solution can be chosen by enforcing constraints, such as sparsity. In other words, we assume the solution has many zero or near-zero coefficients in a representation basis, which is the case for compressible signals. We also enforce smoothness in the solution, which is reasonable for most natural images. Linear programing can be used to find this solution; however, in the presence of noise, basis pursuit techniques prove to be better suited. We chose
to use the NESTA software package due to its fast and accurate recovery method for solving basis pursuit problems [20]. The code converges to a solution quickly, and includes total variation smoothing parameters which can be used to deal with different amounts of noise within the data.

VII.D DMD™ Time Division Multiplexing

Displaying two-level patterns on the DMD™ is a far simpler process than displaying a discretized continuous signal. The DMD™ operates in two states, lending itself to easily display binary patterns. Having the DMD™ display continuous (in magnitude) patterns; however, requires time division multiplexing (TDM) over the duty cycle of the display time. See [21] for details on DMD™ TDM operation. When loading 8 bit PNG image files into the D4100 controller board, the onboard software automatically generates eight 1 bit files corresponding to its binary representation. The D4100 then uses these 1 bit files for pulse width modulation of the DMD™. This time sequence of patterns impacts the sampling of the compressive imager and were investigated using the experimental data plotted in Fig. VII.3. To obtain this data, we input an 8 bit PNG representing the pattern where every pixel value was set to 85 decimal. This value was selected because it’s binary representation, 01010101 binary, requires every transition in the TDM representation. To generate this dataset 5120 samples were taken during the display time of the PNG. This corresponds to 2560 samples for the most significant bit, and 20 samples for the least significant bit. Some samples are taken during the transition state of the mirror, when they are moving between the +12° and the −12° positions, as can be seen in Fig. VII.3.

To characterize mirror transitions as a possible source of error, we used a uniform white light input source and displayed patterns where every pixel was set to the same constant value. We input 256 PNG files where the constant values ranged from 0 to 255, so the measured intensities should ideally also ramp from 0 to 255. We then compared two sampling modalities. For the first, we sample within each subframe time interval for the respective bits, then average and scale
Figure VII.3: Grayscale representation through the use of time division multiplexing of the DMD™. An 8 bit PNG image loaded into the D4100 controller board is converted into eight 1 bit image files, each of which is displayed for a different time duration. The dataset of a single frame shows a test case where all the values of the input 8 bit PNG are set to 85 decimal which, when converted to binary, alternate between zeros and ones. Some samples are taken during the mirror transition reset time.

them by their respective place value. This sampling technique was used to selectively ignore samples taken during mirror transitions. For the second sampling modality, we sample and average all 5120 values taken during the entire TDM 8 bit frame. The comparison between these two sampling modalities for TDM of 8 bit DMD™ patterns is shown in Fig. VII.4. From the plots in Figs. VII.4(c) and VII.4(d) it is evident that there is error associated with the transitions required for TDM grayscale representation. We operated the system, taking an average over the entire TDM PNG display time, without attempting to avoid sampling mirror transitions since the resulting error was less significant. The system operation parameters the authors used for the 1 bit and 8 bit encoding are summarized in Table VII.D for the particular case of capturing 1% of complete basis measurements. Basis load time is the time it takes to load the transform patterns onto the D4100 controller board, image acquisition is the duration of image capture, basis set memory is the memory needed to store the basis set, and measurement memory is the memory needed to store the measurements of the basis set coefficients. Many of the limitations are due to the memory size on the D4100 and the USB 2.0
data transfer rates. These are not fundamental limits; a dedicated ASIC controller
would improve performance.

**Table VII.2**: Speed and memory requirements of encoding algorithms for acquiring 1% of the samples (10,486 basis patterns and coefficients) of the corresponding 1 megapixel image. Basis load time is the time it takes to load the transform patterns onto the D4100 controller board, image acquisition is the duration of image capture, basis set memory is the memory needed to store the basis set and measurement memory is the memory needed to store the measurements of the basis set coefficients.

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Basis Load Time</th>
<th>Image Acquisition</th>
<th>Basis Set Memory</th>
<th>Measurement Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-bit</td>
<td>2.22 min</td>
<td>1.4 min</td>
<td>250 Mb</td>
<td>163 kb</td>
</tr>
<tr>
<td>8-bit</td>
<td>16.83 min</td>
<td>1.4 min</td>
<td>7.4 Gb</td>
<td>163 kb</td>
</tr>
</tbody>
</table>

**VII.E Experimental Setup**

Two measurement/reconstruction approaches were utilized in the experiments. The first relies on sampling techniques presented in the original formulation of CS theory [1, 2, 3], where the measurement basis is known to be incoherent with the sparse representation basis, and where sparsity pursuit reconstruction algorithms are used to reconstruct the image. The second approach uses a sparse measurement basis and relies on *a priori* knowledge (discussed in Section VII.B) that the statistical distribution of spatial frequencies in natural images, as an ensemble, follows a decaying exponential. Noiselet transforms were utilized as the incoherent measurement basis for the former, and Hadamard and DCT were used as the sparse basis for the latter.

The system’s optical components were enclosed to prevent stray light from affecting the measurements. We took 5120 samples per basis pattern, each of which was displayed for 4 ms, with the data acquisition card operating at 200,000 samples per second. Although not the maximum DMD™ speed, this display time accommodated for more measurements. The DMD™ mode of operation depended on whether the basis transform sets were two level or discrete. Our experiments tested three basis transform sets: the Noiselet, Hadamard, and the DCT shown
Figure VII.4: Two methods of sampling the CI system patterns are compared in order to improve 8 bit system performance. To perform the comparison, 256 PNG files are run through the system, each containing uniform intensity ranging from 0 to 255. (a) Data taken by averaging samples within a subframe and scaling by their place value. (b) Data taken by averaging all values within the entire frame. Error associated with (a) and (b) are shown in (c) and (d), respectively. Periodic fluctuations in (c) are dominated by the $2^5-2^4$ place value transition. Averaging over the entire frame introduces less significant error.
Figure VII.5: Sixteen of the basis patterns depicting the (b) Noiselet (c) Hadamard and (d) DCT basis transform sets. The colorbar is shown in (a).

schematically in Fig. VII.5. The first two patterns are two level and the last is a discretized continuous signal. For reasons discussed in Section VII.D, the samples for 1 bit and 8 bit transforms were averaged over the entire pattern display time. The experimental conditions are discussed in the following.

Two chrome on glass transparencies were back illuminated using a 10 W white light four-element (Seoul Semiconductor model number P7) LED and diffusive screen in a lab setting. A USAF 1951 resolution target, whose ground truth image is shown in Fig. VII.6(a), served as a binary transparency containing sharp edges. A portrait of Lena, a standard test image whose ground truth image is shown in Fig. VII.6(b), served as a grayscale image with more texture and features. A camera with a sensor pixel pitch and resolution comparable to the mirror pitch of the DMD™ were unavailable, so in order to form a ground truth image we used a commercial digital SLR camera with resolution superior to the CI system. Both of these indoor lab images were taken using a Canon DSLR 5D Mark II using a Sigma 50 mm focal length lens set to f/14 with an exposure of 1/125 and 1/500 seconds, respectively. The sensor in the Canon Mark II has $5616 \times 3744$ pixels with a pitch of 6.4 $\mu m$, of which we used a $1024 \times 1024$ pixel region to image the scene. For the lab images there was 18.3 lux of illuminance and 11.91 mW of power at the camera lens. No optical filters were used. During these indoor tests, stray light was well controlled and its effects negligible. Contrast was limited by the DMD™.

For an outdoor field test of the system in a daylit setting, the input scene
Figure VII.6: Ground truth images taken with a Canon 5D Mark II DSLR camera. The scenes include (a) and (b) back-illuminated chrome on glass transparencies in a lab setting, and (c) a daylight illuminated building.

was of a University of California San Diego engineering building. A ground truth image taken using a Canon DSLR 5D Mark II with a Canon 70-300 mm focal length zoom lens set to $f = 100$ mm, f/9 and with an exposure of 1/200 seconds is shown in Fig. VII.6(c). In addition to taking natural environment experimental data, we sought to form a multispectral image using optical filters at the photodetector to create RGB color and infrared images. The RGB color filters had center wavelengths of 450, 550 and 650 nm, each with a FWHM bandwidth of 70 nm. The infrared filter passed wavelengths longer than 700 nm. In this outdoor setting there was 665 lux incident on the system, with 22, 227, 76 and 11 lux passing through the four respective filters. The outdoor power reading measured using the Newport power meter was 8.9 W and with the respective filters present was 484 mW, 623 mW, 969 mW and 5.19 W. A plot of the outdoor spectra (shown in Fig. VII.7) taken during the data acquisition was measured using an Ocean Optics Spectrometer USB 4000-VIS-NIR-ES in the absence and presence of these filters. This plot shows the relative energy levels the CI system performed under
Figure VII.7: Spectral data of the outdoor image scene Fig. VII.6(c) taken using an Ocean Optics Spectrometer with and without color filters in place.

during image acquisition. In this setting, stray light increased noise to nearly twice the noise floor of the photodetector. For example, with the blue filter in place, the noise floor was 4 nW while the stray light increased this value to 6.8 nW; the average value of the signal present during imaging was approximately 460 nW.

VII.F Experimental Results

The six complete 1,048,576 sample datasets captured for the two laboratory images are shown in Fig. VII.8. The data shown in Figs. VII.8(a) and VII.8(d) supports the idea that the Noiselet sampling technique, is "democratic" (see Section VII.B), since the coefficient values are nearly uniform for all samples. When using a spatially structured pattern such as the Hadamard (Figs. VII.8(b) and VII.8(e)) or DCT (Figs. VII.8(c) and VII.8(f)) transforms, the measured coefficient values stand out for certain spatial frequencies. This is due to interactions between the structure of the basis patterns and the image. In general, magnitudes of the coefficients are inversely proportional to spatial frequency. This is supported by both datasets but more so by the data from the resolution target than from the Lena portrait. This can be expected since the statistical model cited in Section VII.B acts upon the ensemble of natural images and not the few specific images.
Figure VII.8: Experimentally acquired (a), (d) Noiselet, (b), (e) Hadamard, and (c), (f) DCT complete datasets from lab-imaged, back-illuminated chrome on glass transparencies of a 1951 USAF resolution target (a), (b), and (c) and a grayscale Lena portrait (d), (e), and (f). Each sample corresponds to a basis pattern and produces a basis coefficient. The ground truth of the corresponding images are shown in Figs. VII.6(a) and VII.6(b). As opposed to Noiselet transform, highly spatially structured patterns, such as Hadamard and DCT, interact with the image, making certain coefficients stand out.

with which we tested the functionality of the compressive imager. The image reconstruction computation took less than 1 minute in every case. The laboratory reconstructions of the resolution target and the Lena image are shown in Figs. VII.9 and VII.10, respectively. The root mean squared difference (RMSD) values were calculated between reconstructions made using some and all samples. The plot in Fig. VII.11 summarizes the CI system error for the two images using the three measurement transforms.

Figures VII.12 and VII.13 show reconstructions of an outdoor scene when only 1% of the samples are taken, namely 10,486 samples. It is important to note that the Noiselet and Hadamard images took about 7 minutes per channel, while
Figure VII.9: (a), (b), (c), (d) Noiselet, (e), (f), (g), (h) Hadamard, and (i), (j), (k), (l) DCT reconstructed images of a 1951 USAF resolution target using 0.1, 1, 10, 100% of samples recorded in a lab setting. Error is calculated with reference to the respective 100% reconstruction.
Figure VII.10: (a), (b), (c) Noiselet, (d), (e), (f) Hadamard, and (g), (h), (i) DCT reconstructions of a grayscale image using 1%, 10%, 100% of samples recorded in a lab setting. Error is calculated with reference to the respective 100% reconstruction.
Figure VII.11: Plot of the RMSD error versus the number of measurements used in the reconstruction of the two lab scenes using the three basis sets. Measurements are taken from Fig. VII.9 (solid line) and Fig. VII.10 (dashed line) and are normalized to the number of functions in the complete basis set. The Hadamard transform has the least error followed closely by the Noiselet and lastly the DCT.

the DCT images took about 20 minutes per channel due to the larger memory needed to store and transfer the 8 bit DCT patterns. The D4100 controller board has limited memory, making it necessary to load and run the patterns several times, imposing a limit on the operation of the compressive imager. Since the 1 bit and 8 bit RGB color images shown took about 21 and 60 minutes to be captured, respectively, changes in the daylit scene became significant. Additionally, since different color filters were used at different times, there are color artifacts in these regions of changing illumination.

Fig. VII.14 shows the comparison of the compressive images made using 1% of the total samples (10,486 samples) with Canon 5D images where the pixel data has been binned to 1% of the pixel count (10,486 pixels). Images in Fig. VII.14(a) are compressed to 1% of their original size by binning the pixels. Images in Fig. VII.14(b) are upsampled versions of those in Fig. VII.14(a) to show a less pixelated result for comparison. Images in Fig. VII.14(c) show the resulting compressive images made using the 10,486 lowest spatial frequency samples measured with the Hadamard transform.
Figure VII.12: Reconstructed images with the red, green, and blue bandpass filters were combined to form outdoor color images. Reconstructions were made using 1% of the (a) Noiselet, (b) Hadamard, and (c) DCT transforms. Daylight changed during the image acquisition resulting in variations in color and illumination.

Figure VII.13: Infrared reconstructed images using 1% of samples recorded outdoors using an IR bandpass filter.
Figure VII.14: Comparison of the compressive imager with a conventional camera. (a) Images taken by a Canon 5D Mark II where the pixel data has been binned to 1% of its original 1 megapixel size, namely 10,404 pixels. (b) Shows the images in (a) upsampled to the same resolution (1 megapixel) as the compressive imager. (c) Show the compressive imaging result using 1% (10,404 samples) of the lowest spatial frequency Hadamard data.
VII.G Conclusion

Here we report the first (to our knowledge) systematic performance comparison of a CI system to a conventional focal plane imager for binary, grayscale, and natural light (visible color and infrared) scenes. We generate \(1024 \times 1024\) images from a range of measurements (0.1% to 100%) acquired using digital (Hadamard), grayscale (discrete cosine transform) and random (Noiselet) CI basis sets. In general, the CI system images acquired with different basis set representations conformed to expectations. Images taken with the Hadamard and DCT transforms increase in resolution as more samples acquired using higher spatial frequency patterns are included. Images acquired with the Noiselet basis set, which targets all spatial frequencies equally, exhibit a different behavior. With the Noiselet basis, higher spatial frequency information is present but more samples are needed to converge to a visually appealing solution. This is apparent from the 1951 USAF resolution target (Fig. VII.9), as well as from the grayscale "Lena" portrait (Fig. VII.10). The DCT reconstruction deals with contour edges less effectively than the Hadamard reconstruction due in part to the DMD\textsuperscript{TM}'s TDM encoding (see Fig. VII.10(g)). The system performance is reflected in the RMSD error; however, the error values do not reveal subtle differences in the reconstruction quality. The outdoor color images (Fig. VII.12) and outdoor infrared images (Fig. VII.13) are similar in that the Hadamard transforms perform the best overall, the Noiselet reconstructions are data starved for low sample counts, and the DCT reconstructions exhibits artifacts at the image edges and color artifacts due to the longer exposure times. For this reason, we concluded that the Hadamard transform was the best representation for general-purpose compressive imaging, and used it in our comparison to the images acquired by a conventional (but equal resolution) focal plane imager (Fig. VII.14). The comparison shows clearly that, as the theory predicts, the CI images with 1% of full sampling did obtain substantially identical resolution when compared to a focal plane with 1% of the original 1 megapixel image resolution. Given the potential SLM and detector hardware limitations, this was not a foregone conclusion. The 20 dB dynamic range of the SLM clearly limited the contrast for the outdoor images (comparing outdoor images in Figs.
In general; however, the results confirm that a general-purpose CI system can use a standard basis representation to acquire images with a wide range of scenes and lighting conditions. As expected, the CI imager’s exposure requirements make it impractical when compared to a conventional imager; however, here we used it as a test bed to quantify and validate performance of this emerging technology.

There are several hardware configurations that could improve performance of the compressive imaging system. The most obvious is to sample both paths of light reflected from the DMD by replacing the light sink with a second photodetector (Fig. VII.1). We successfully reconstructed images using this configuration in lab experiments but crosstalk between the transform and the surrounding mirrors of the DMD worsened performance since the DMD resolution did not match that of our transforms. This implementation doubles the acquisition speed by capturing the positive and negative portions of each pattern simultaneously. Color artifacts in the color images could be mitigated using a rapidly switching color filter to acquire spectral data for each pattern sequentially, rather than taking three separate images. Increasing DRAM memory in the D4100 controller board, such that all the patterns could be loaded prior to image capture, would significantly speed up data acquisition to a 1% sampling runtime of 1.4 minutes or less. Even faster acquisition could be achieved, and dynamic range improved, by multiplexing parts of the spatially filtered signals onto multiple parallel detectors, as proposed by Ke, Shankar and Neifeld [22], providing a continuum between conventional focal planes and compressive imagers. In the longer term, ASIC image sensors with high pixel (sensor) counts and programmable on-chip signal aggregation will be able to integrate the pattern encoding directly into the image sensor itself, eliminating the need for an external SLM and enabling CI systems to function with high-performance image formation optics [17, 18]. Finally, the most significant performance improvement in CI will be enabled by making use of the intrinsic architectural flexibility for feature specific imaging [23, 24], including face recognition [25, 26], to dramatically decrease the basis set size required to acquire a conclusive measurement. In all, CI offers a versatile approach to optical sens-
ing which can improve SWaP for multi-spectral imaging or feature-based optical sensing.

**Acknowledgments**

This work was supported by the Charles Stark Draper Laboratory, University R & D program contract SC001-00000628A.

Chapter VII, contains material published in the following, of which the dissertation author was the primary investigator and author:


VII Bibliography


Chapter VIII

Conclusion

Specific systems are presented in this thesis to provide a synopsis of the field of Computational Imaging. This is preceded by a brief overview of traditional imaging and inverse problem theory used to assist as a background and foundation of this emerging field. The original contribution of this thesis is to present the design and experimental characterization of these novel Computational Imaging systems. These systems are then used to demonstrate how Computational Imaging can extend the trade space of traditional cameras by changing in the way a scene is measured and how the data is processed.

After reviewing some of the history, operation, and constraints of traditional cameras; a discussion followed on the degradations commonly suffered by these imagers. A mathematical construct of the imaging process and typical degradation models were presented. Inverse problem theory was then introduced and its ill-posedness, due to inversion and noise, was given as a reason as to why its elementary solutions are generally unacceptable. More sophisticated methods used to analyze, better condition, and solve these inverse problems are briefly summarized. It is then explained how Computational Imaging uses hardware modifications and tailored problem formulation to make a well-posed inverse problem describing the system degradation. Solving this, now tractable, problem can produce exceptional images beyond the capability of traditional cameras. Computational Imaging is the union of cameras, computers, hardware and algorithms; and is designed in unison as a single end-to-end system. The data acquired is, in general, not the fin-
ished product but instead a set of measurements from which to process a superior image, or more effectively extract scene information.

Chapter II reviews the history, operation, and fundamental trade-offs of traditional cameras. Chapter III discusses inverse problem theory, methods to solve these problems, and complications that arise due to ill-posedness. This establishes a rudimentary background upon which many Computational imaging Systems are built. Chapter IV shows how fiber bundle image relay used in conjunction with computational imaging techniques allows for non-conventional high-performance lenses to be incorporated into cameras. This makes for innovative and possibly peculiar designs that can enhance performance and expand the application space of cameras. Further performance enhancement may be achieved through custom fiber bundle fabrication with alternate fiber orientation and transfer properties. Chapter V formulates and implements a more general method of extended depth of focus imaging. The theory is explored and supported through the characterization of the angular and depth dependence of the imaging system’s impulse response. More accurate motion control and lower SNR may improve the system’s current results. A greater improvement would likely occur if a depthmap captured using a primary or secondary system is used along with the depth dependent PSF measurements in order to process the EDOF imagery. Chapter VI introduced a camera system that uses optical position sensors to estimate the PSFs at various location within the image in order to calculate a spatially variant PSF that is valid across the image. This was used to deblur images degraded by platform motion blur. The optical position sensors are designed to track laser beams incident on their surface, making the system (in its current state) only useful for sparse image scenes. A mathematical framework upon which to re-validate closely tiled PSD arrays was presented as a way to make the system more compatible with more general images. Interlacing PSD technology with the pixels of a custom image sensor would also make these systems more compact and practical. Chapter VII demonstrates a system which does not follow the traditional imaging method of sampling the image space as a set of contiguous area elements whose intensity measurements are captured simultaneously using pixelated image sensors. Rather this prototype
system captures data from the scene as a sequential set of intensity measurements made up of a superposition of structured regions of the image that adhere to a basis set encoded using a DMD. An image can be recovered from the data collected even if the measurement count is far below the resolution of the modulator allowing for front-end compression. The system can also be used to target specific spatial frequency image content with only the necessary measurements. The sampling modality of the Compressive Imager enables a powerful and useful image sensing technique unique to this imager. Rather than producing imagery, the Compressive Imager can measure correlation between the scene and a tailored basis set in order to report statistical information about what is present in a scene; an application of great interest in machine vision. Adaptively determining which measurements to make based on previously captured measurements may also serve to converge to a conclusive measurement more quickly. There are many applications where imaging is a intermediate step in an overall process. Computational Imaging can greatly aid in such applications since the end-to-end process can be engineered as a single system, where the intermediate image is no longer needed. This might occur in the same way that human perception or machine learning uses pattern recognition techniques to simplify the problem being solved in order to better perform the given task.

The four Computational Imaging system prototypes presented in this thesis sought to: extend the freedom of lens design, depth-of-focus, blur immunity, and compression capability of traditional imagers, respectively. The showcased results demonstrate these capabilities to a varying degree of success. They laid the general design methodology Computational Imaging uses to circumvent limitations of traditional cameras. This is to modify or configure a system in order to achieve a certain end result. Most importantly, they question the way imagers are designed and how imaging is currently performed. A major shift away from traditional imaging leads to extremely promising technologies; but yet, it is difficult to imagine in what forms they will manifest. As the field of Computational Imaging develops it is finding its way into more mainstream applications, this is especially true because its benefits are increasing along side improvements in processor capabilities and
system integration. It is intriguing to imagine what traditional techniques will be replaced and what impressive technologies will arise during the future development of Computational Imaging.