COMPRESSIVE CODED-APERTURE MULTIMODAL IMAGING SYSTEMS

by

Hoover F. Rueda-Chacon

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering

Fall 2017

COMPRESSIVE CODED-APERTURE MULTIMODAL IMAGING SYSTEMS

by

Hoover F. Rueda-Chacon

Approved: ______________________________

Kenneth E. Barner, Ph.D.
Chair of the Department of Electrical and Computer Engineering

Approved: ______________________________

Babatunde Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: ______________________________

Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____________________________________
Gonzalo R. Arce, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____________________________________
Kenneth E. Barner, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____________________________________
Javier Garcia-Frias, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____________________________________
Henry Arguello-Fuentes, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Clare W. Lau, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Sean Wang, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Daniel L. Lau, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

First of all, I would like to express my appreciation to my advisor Professor Gonzalo R. Arce for the continuous support of my Ph.D. study and for the outstanding research projects I had the opportunity to participate. I have been very fortunate to always have all the resources and guidance for every project I worked on. Also I want to thank my co-advisor Professor Henry Arguello-Fuentes, he has been a fantastic mentor in many aspects, always enlightening me with amazing ideas. Beside my advisors, I would like to thank my dissertation committee: Dr. Javier Garcia-Frias, Dr. Kenneth E. Barner, Dr. Clare W. Lau, Dr. Dan L. Lau, and Dr. Sean Wang, for their insightful comments and discussions that helped me to improve my research from various perspectives.

I also appreciate the support from my fellow research group members: Chen Fu, Carlos Mendoza, Laura Galvis, Angela Cuadros, Juan Florez and Edgar Salazar. They are all outstanding researchers which have helped me to grow and perfect my research with their suggestions.

I have to thank the funding from the Office of Naval Research, the Army Research Laboratory, the Army Research Office, Intel Corporation and the National Science Foundation; without their support my research goals would have never been accomplished. Specially, I want to thank Fulbright and Colciencias for the granted scholarship to pursue my Ph.D. studies in the United States.

I would like to thank my wife, "*Cato*". Her patience, loyalty and never-failing love have been the principal source of strength and motivation to complete this professional journey together. Her constant feedback and support have primarily driven the accomplishment of many goals in my life.

Finally, I would like to thank my family for their understanding of being far from home all this time. Specially, my mother who has always encouraged me to pursue my dreams and to become the person I am today.

TABLE OF CONTENTS

LIST OF TABLES . . . . **x**
LIST OF FIGURES . . . . **xi**
ABSTRACT . . . . **xix**

Chapter

1 INTRODUCTION **1**

1.1 Motivation of the Thesis 4
1.2 Organization and Overview of the Thesis 5

2 DMD-BASED IMPLEMENTATION OF PATTERNED OPTICAL FILTER ARRAYS FOR COMPRESSIVE SPECTRAL IMAGING . **8**

2.1 Introduction 8
2.2 System Model 11

2.2.1 Discrete Model 14
2.2.2 Inverse Model 17

2.3 Simulations 17
2.4 Experimental Setup 22
2.5 Conclusions 28

3 COMPRESSIVE SPECTRAL TESTBED IMAGING SYSTEM BASED ON THIN-FILM COLOR-PATTERNED FILTER ARRAYS **29**

3.1 Introduction 29
3.2 Compressive optical sensing image model 30

3.2.1 Multiple-snapshot matrix model 32

3.2.2 Input data cube reconstruction 34

3.3 Design and fabrication of a patterned filter array wafer 36

3.3.1 Patterned filter array design 36

3.3.1.1 Random distributed filters 37
3.3.1.2 Boolean distributed filters 38
3.3.1.3 Optimized spatial distribution of the filters 38

3.3.2 Thin-film-based patterned filter array fabrication 40

3.4 Optical testbed implementation and experiments 41
3.5 Discussion 48
3.6 Conclusions 49

4 OPTIMIZATION OF COLOR CODED APERTURES FOR COMPRESSIVE SPECTRAL IMAGING VIA 3D SPREADING . . 50

4.1 Introduction 51

4.1.1 Motivation 51
4.1.2 Related work 52
4.1.3 Contributions 52

4.2 Higher-order Compressive Spectral Imaging Model with Colored Coded Apertures 53

4.2.1 C-CASSI Continuous Model 54
4.2.2 C-CASSI Higher-Order Discrete Model 55
4.2.3 C-CASSI Matrix Representation 58

4.3 Sensing Matrix Analysis 59

4.3.1 Transfer Function (Matrix **H**) Entries 59
4.3.2 Analysis of the Compressive Sensing Matrix **A** 61
4.3.3 Grammian Matrix Analysis 62

4.4 Higher order color coded aperture optimization 63

4.4.1 Optimization Algorithm 64
4.4.2 Euclidean Filter Design 65

4.5 Simulation Results 66

4.5.1 Coded Aperture Optimization Comparison 67

4.5.2 Testing the CCAs in the CSI system 68

4.6 Conclusions 71
4.7 How to obtain Eq. (4.30) 73
4.8 Algorithm subroutines 74

5 MULTI-SPECTRAL COMPRESSIVE SNAPSHOT IMAGING USING RGB IMAGE SENSORS 77

5.1 Introduction 77
5.2 Camera model 78
5.3 Compressive sensing model 80
5.4 Color sensing model 84
5.5 Simulations and experiments 86

5.5.1 Simulations 86
5.5.2 Experiments 90

5.6 Discussion 92
5.7 Conclusions 94

6 DUAL-ARM VIS/NIR COMPRESSIVE SPECTRAL IMAGER . . . 96

6.1 Introduction 96
6.2 System Model 97

6.2.1 Optical Model 97
6.2.2 Discrete Model 99
6.2.3 Image Reconstruction 100
6.2.4 Joint Image Reconstruction 102

6.3 Experimental Setup 102
6.4 Conclusions 104

7 SINGLE APERTURE SPECTRAL+TOF COMPRESSIVE CAMERA: TOWARDS HYPERSPECTRAL+DEPTH IMAGERY . 107

7.1 Introduction 107

7.1.1 Related work 109
7.1.2 Contribution 109

7.2 Snapshot Single Aperture Compressive ToF+Multispectral Camera . . . 110

7.2.1 Time-of-Flight Imaging (ToF) 112
7.2.2 Compressive Spectral Imaging 117

7.2.3 Joint Compressive Spectral+ToF Imaging 120

7.3 Compressive ToF+Multispectral Testbed Imaging System 122

7.3.1 Calibration of the System 124
7.3.2 Capturing Compressive CSI and ToF Measurements 125
7.3.3 Reconstruction Results for CSI 128
7.3.4 Reconstruction Results for ToF 135

7.4 Comparison with alternative solutions and limitations of the current system 135
7.5 Conclusions and Future work 137

BIBLIOGRAPHY 139

Appendix

A DISCLAIMER 147

LIST OF TABLES

4.1 PSNR and SSIM results for the 6 spectral images with $L = 8$ 68

LIST OF FIGURES

1.1 (a) CASSI optical architecture proposed in [1]. (b) Light propagation through the CASSI architecture. 3

2.1 Colored coded aperture-based compressive spectral imager (C-CASSI). The 3D spectral scene $f_0(x, y, \lambda)$ is 3D-coded by the CCA $T(x, y, \lambda)$ and horizontally sheared by the dispersive element with dispersion function $S(\lambda)$, before being integrated by the focal plane array (FPA) detector $g(x, y)$. Notice the 3D coding scheme entailed by the CCA. 9

2.2 Synthesis of a 3-color CCA. The CCA is indistinctly modeled by 3 different CA $\overline{T}_0(x, y)$ (filtered by a low-pass filter), $\overline{T}_1(x, y)$ (filtered by a band-pass filter), and $\overline{T}_2(x, y)$ (filtered by a high-pass filter). 11

2.3 Synthesis of the CCA effect on the $y = q^{th}$ slice of the data cube illustrated in Fig. 1. (a) Input-output relationship of the CCA synthesis. (b) 2D coding and filtering process. The slice is first coded by the q^{th} row of the $\overline{T}_0(x, y)$ CA, and then filtered by the $\xi_0(\lambda)$ optical filter; this process is realized for every r^{th} CA/filter pair, and the resulting spectra are combined. 13

2.4 **H** matrices for a $4 \times 4 \times 4$ data cube. (a) Traditional CASSI, and (b) C-CASSI. The diagonal entries correspond to the coding pattern, which remain invariant with the wavelength in CASSI, but are wavelength dependent in C-CASSI. 16

2.5 RGB targets along with their corresponding 8-band spectral data cubes. (Left) A scene exhibiting green and blue colors (Database 1). (Right) A scene exhibiting red and yellow colors (Database 2). 18

2.6 Spectral responses of the different filter configurations. (a) All-pass filter for the traditional CASSI, and (b) $V = 2$, (c) $V = 4$, and (d) $V = 8$, for C-CASSI. 19

2.7 Impact of the transmittance (R) in the reconstruction PSNR as a function of the number of snapshots (K). 19

2.8 Impact of the number of different colors in the reconstruction PSNR for different number of snapshots and (Left) $R = 0.2$, (Right) $R = 0.3$. . . 21

2.9 Comparison of the best achieved reconstruction PSNR between CASSI and C-CASSI (with $V = 2, 4, 8$) as a function of the snapshots, for fixed transmittances (Left) $R = 0.2$, (Right) $R = 0.3$. 21

2.10 Impact of the transmittance in the reconstructions, using $K = 3$ and varying $R = 0.1, 0.2, 0.3$ from left to right. (a) Database 1, (b) Database 2. 22

2.11 Impact of the number of snapshots in the reconstructions, fixing $R = 0.3$ and varying $K = 1, 4, 8$, from left to right. (a) Database 1, (b) Database 2. 23

2.12 Spectral signatures from P1 and P2 in Fig. 2.5, with $R = 0.3$, $K = 3$ and $V = 4$. (a) Database 1. (b) Database 2. 23

2.13 Optical scheme of the DMD/Filter-wheel-based C-CASSI. The first lens focuses the scene onto the DMD image plane, which spatially modulates the scene, and a second objective lens focus the coded scene into the image plane of the FPA, which has been first filtered by the rotation wheel, and sheared by the prism. The DMD and filter wheel change rates are faster than the integration time from the FPA. 24

2.14 Snapshot synchronization control interface tasks using $V = 2$ filters. At t_0, the first DMD pattern/filter pair is set, and it is being measured between t_0 and t_1. At t_1, the control changes the DMD pattern, and rotates the filter wheel to the second filter. The DMD sets an all-black pattern while the filter wheel ends its rotation, to reduce distortion. At t_2 the second DMD pattern/filter pair is set, and it is measured during $t_3 - t_2$ ms. At t_3 the FPA ends its capturing process, and both the DMD and the filter wheel restarts. 25

2.15 Comparison of the compressive measurements using (a) CASSI (b) C-CASSI. In CASSI the intensity appears more clustered in neighboring pixels, rather than in C-CASSI where the intensity is more uniformly distributed. 26

2.16 Reconstructed data cubes mapped to an RGB profile for: (a) $K = 1$, and (b) $K = 4$. 27

2.17 Experimental reconstructions of the 3rd, 5th and 7th spectral bands. (a) $K = 1$ (b) $K = 4$ 27

2.18 Spectral signatures of 2 different points using $K = 1$ and $K = 4$, respectively. (Top row) Helmet blue point, (Bottom row) Arm orange point. 28

3.1 Comparison of the coding strategies. (a) Photomask coding. (b) Patterned filter array coding. The filter array enables 3D coding due to the spectral response of the pixelated filters, whereas the photomask performs 2D coding over the data cube. 31

3.2 Sketch of transfer function matrix $\mathbf{H}$ for a $4 \times 4 \times L$ spectral data cube, using $K = 2$ snapshots, and a patterned filter array with L filters. $Q = N^2L$ is the total number of data cube voxels, and $U = N(N + L - 1)$ is the total number of FPA pixels. Note that the coding pattern is wavelength-dependent, and a different distribution of its entries is used in each snapshot. 35

3.3 3-color patterned filter array design using 3 different spatial distributions. The blue pixels are low-pass filters, the green are high-pass filters, and the black are all-stop filters. 36

3.4 Fabricated wafer of thin-film patterned filter arrays. (a) A microscope was used to capture zoomed versions of the wafer shown from left to right. (b) Analysis of the transition band of the filters. Due to the cut-off wavelength of the filters ($\lambda_c = 550$ nm), both HP and LP filters show-up attenuated when the patterned filter is illuminated at that wavelength. 37

3.5 Thin-film-based compressive spectral imaging testbed. The objective lens images the captured scene onto the thin-film patterned filter wafer; then the relay lens transmit the coded light through the Amici prism before it is integrated by the CCD sensor. The X-Y motion system moves the wafer precisely to access different coding patterns. 39

3.6 Full data cube reconstructions when random, Boolean and optimized patterned filter arrays are used, and $K = 2$ snapshots are captured. The 14-bands data cubes are mapped to false-colored RGB profile to visually compare the color fidelity with the original target scene. 43

3.7 Full data cube reconstructions when random, Boolean and optimized patterned filter arrays are used, and $K = 8$ snapshots are captured.The 14-bands data cubes are mapped to false-colored RGB profile to visually compare the color fidelity with the original target scene. 44

3.8 Comparison of the spectrum reconstructions when $K = 2$ snapshots are captured. Six different points (P1, P2, P3, P4, P5 and P6) from the target scene (Fig. 7.5) were measured by a spectrometer and compared against the reconstructed data cubes. 45

3.9 Comparison of the spectrum reconstructions when $K = 8$ snapshots are captured. Six different points (P1, P2, P3, P4, P5 and P6) from the target scene (Fig. 7.5) were measured by a spectrometer and compared against the reconstructed data cubes. 45

3.10 Comparison of the reconstructions at monochrome level. The 2^{th}, 4^{th}, 6^{th}, 8^{th}, 10^{th}, 12^{th}, and 14^{th} spectral bands for the 3 different distributions are shown for $K = 2$ snapshots. (First column) Random, (Second column) Boolean, (Third column) Optimized distribution. . . . 46

3.11 Comparison of the reconstructions at monochrome level. The 2^{th}, 4^{th}, 6^{th}, 8^{th}, 10^{th}, 12^{th}, and 14^{th} spectral bands for the 3 different distributions are shown for $K = 8$ snapshots. (First column) Random, (Second column) Boolean, (Third column) Optimized distribution. . . . 47

4.1 CSI systems of interest. (Left) Coded aperture snapshot spectral imager (CASSI). (Right) CASSI with color-coded apertures (C-CASSI) 54

4.2 Propagation of a single voxel through the CSI system 57

4.3 Sketch of the higher-order sensing matrix $\mathbf{H}^{\ell}$ for a datacube of size $16 \times 16 \times 4$. The diagonals refer to the wavelength-dependent coding pattern $\left(\mathbf{t}_k^{\ell}\right)$ weighted by the corresponding dispersion distribution $(\mathbf{w}_k^r)$. The N off-set between the diagonals correspond to the column-wise representation of the dispersion. 60

4.4 Sketch of the 3D spreading Euclidean filter $\boldsymbol{\Xi}$. 67

4.5 Analysis of the singular value spread of the sensing matrices $\mathbf{A}$ for the 3 different color coded apertures, with $N^2 = 256^2$ spatial pixels, $L = 8$ spectral bands, and (a)$K = 2$ snapshots, (b) $K = 4$ snapshots. 68

4.6 Comparison of color coded apertures with $L = 24$ and $N = 128$. (a) Front and top-view of the 3D black-and-white representation to easily identify the 3D spreading. (b) Single realization for $K = 2, 4$ and 8 snapshots. 69

4.7 Comparison of the reconstructions in terms of PSNR (first row) and SSIM (second row) for different snapshots using the four kinds of coded apertures for the 6 databases with $L = 24$ bands. 70

4.8 Reconstructed datacubes mapped to RGB profile when $K = 4$ snapshots ($\approx 16\%$ of the data) are captured. From left to right there is shown the 6 databases: First, the Lego database, which was captured in our lab, and then, the other 5 (beads, feathers, flowers, pompoms and thread-spools), respectively, downloaded from [2]. The insets show 2 zoomed regions of each scene for easy comparison. From top to bottom we have the ground-truth image, and then the results attained with Random, Boolean, Row-wise and the Proposed color coded apertures. . 72

4.9 Zoomed version of the $3^{rd}, 7^{th}, 11^{th}, 15^{th}, 19^{th}$ and 23^{th} reconstructed spectral bands of the $L = 24$ Lego database, using $K = 4$ snapshots ($\approx 16\%$ of the data). (First row) Ground truth, (Second row) Boolean, (Third row) Row-wise optimization, and (Fourth row) Proposed optimization. 73

4.10 Zoomed version of the $3^{rd}, 7^{th}, 11^{th}, 15^{th}, 19^{th}$ and 23^{th} reconstructed spectral bands of the $L = 24$ Beads database, using $K = 4$ snapshots ($\approx 16\%$ of the data). (First row) Ground truth, (Second row) Boolean, (Third row) Row-wise optimization, and (Fourth row) Proposed optimization. 74

4.11 Zoomed version of the $3^{rd}, 7^{th}, 11^{th}, 15^{th}, 19^{th}$ and 23^{th} reconstructed spectral bands of the $L = 24$ Feathers database, using $K = 4$ snapshots ($\approx 16\%$ of the data). (First row) Ground truth, (Second row) Boolean, (Third row) Row-wise optimization, and (Fourth row) Proposed optimization. 76

5.1 Illustration of the multispectral imaging principles used by CASSI. . . 79

5.2 Structure of the system transfer functions $\mathbf{H}$ for a $4 \times 4 \times 6$ data cube, using a monochromatic sensor, where (top) shows the transfer function for a typical monochrome camera, (center) shows the same camera but with the incoming light divided into spectral bins, and (bottom) models the effects of the dispersive element by offsetting the diagonal structure from band to band. Note that the non-zero diagonals correspond to the entries of the $M \times N$ coded aperture set to be all ones. 81

5.3 FPA detector architectures along with their traditional spectral responses. (left) Monochromatic (Source: Stingray F-033B), (center) RGB-Bayer (Source: TMC) and (right) Foveon (Source: Foveon Quattro sensor). 83

5.4 Non-zero elements of the system transfer functions, $\mathbf{H}^{\ell}$, for a $4 \times 4 \times 6$ data cube, using (left) RGB-Bayer and (right) Foveon sensors. 85

5.5 Simulation results with first target scene. (a) Target scene for the simulations. (b) Averaged PSNR of the reconstructed data cubes as function of the compression ratio (R). (c) Reconstructed datacubes, using (first row) $R = 1/8$ and (second row) $R = 1/2$. 87

5.6 Simulation results with second target scene. (a) Target scene for the simulations. (b) Averaged PSNR of the reconstructed data cubes as function of the compression ratio (R). (c) Reconstructed datacubes, using (first row) $R = 1/8$ and (second row) $R = 1/2$. Note the color aliasing in the zoomed version of the RGB-Bayer reconstruction. 88

5.7 Absolute errors of the reconstructions of the $4^{th}, 8^{th}, 12^{th}, 16^{th}, 20^{th}, 24^{th}$ spectral bands, for $R = 1/8$, using (First row) Monochrome sensor, (Second row) Bayer sensor, (Third row) Foveon sensor. Notice that the Foveon sensor attains the best reconstructions between the three sensors, despite the high frequencies (borders) are almost unrecoverable. 89

5.8 DMD-based CASSI testbed setup used in the experiments. The illuminated target scene is imaged onto the image plane of the DMD which plays the role of the coded aperture. Subsequently, the relay lens transmits the coded light through the Amici prism which disperses it onto the image plane of the CCD array. 91

5.9 Real measurements for a single snapshot of the monochrome, the Bayer and the Foveon sensor. Zoomed versions highlight the sensors differences. 92

5.10 Experimental reconstruction results. (a) Target scene for the experiments. (b) Real reconstructed datacubes using $R = 1/8$ and $R = 1/2$. (c) Spectral signatures of two reference points measured from the $R = 1/8$ reconstructions. 93

5.11 Real reconstruction of the 6^{th}, 12^{th}, 18^{th} and 24^{th} spectral bands using just a single snapshot of the (First row) Monochrome sensor, (Second row) Bayer sensor, (Third row) Foveon sensor. 95

6.1 Dual-Arm optical design 98

6.2 Sensing phenomenon performed in each imaging arm 100

6.3 Dual-arm implementation. (a) Full optical prototype with the two imaging arms. (b) Zoomed version. 103

6.4 Original scene and reconstructed data cubes. (a) Lego toy with 3 LEDs with peak emission wavelengths at 1050 nm, 1200 nm, and 1300 nm. (b) Visible data cube reconstruction, and (c) NIR data cube reconstruction. 105

6.5 Monochome reconstructed spectral bands. (a) Visible spectral bands. (b) NIR spectral bands 106

7.1 Optical design sketch of the single aperture snapshot ToF+Spectral camera. The LED laser diode illumination board is attached to the front lens of the camera which images the 4D scene onto the DMD. Depending on the mode, the DMD will emulate an all-on or a random coded aperture, for the ToF and ambient light capture, respectively. The relay lens transmits the DMD coded and modulated light through a dispersive element, so that monochrome coded, modulated and sheared light is integrated on the ToF sensor. 111

7.2 Light propagation of the q^{th} slice through the proposed system. (a) In ToF mode, the reflected NIR modulated light is captured after ambient light is subtracted. The single red spectral band represents the reflected NIR light of the LED illumination. The DMD emulates an all-transmissive pattern, such that, no coding is performed. Therefore, propagation through the system causes no multiplexing but shifting in the captured image. (b) In ambient light mode (CSI mode), the incoming light is encoded by the coded aperture pattern $T(x, y)$, then the dispersive element decomposes the light in the corresponding wavelengths, and multiplexing takes place at the sensor. 113

7.3 ToF sensor raw data g_0 to g_3 are captured in every single $(m, n)^{th}$ pixel. The corresponding phase delays θ_0 to θ_3 are then calculated to obtain the depth map. 115

7.4 Ambient image is recorded in grayscale mode. Four successive frames g_0, g_1, g_2 and g_3 are captured for the depth map and point cloud estimation. 117

7.5 Compressive single-aperture ToF+CSI testbed. An objective lens with an attached LED illumination board images the light field of the target scene onto the DMD. After the DMD, a VIS-NIR relay lens transports the coded light through the dispersive element, which decomposes the light field in the corresponding wavelengths being integrated in the ToF sensor device. Side information can be obtained by incorporating a beam-splitter and a secondary sensor as proposed in [3]. 123

7.6 Coded aperture calibration patterns as seen by the ToF sensor when illuminated with monochrome light at the given wavelengths. The coded aperture is the result of a binary random variable with 50% transmittance. 125

7.7 Target scenes used in the experiments. The first scene is placed around 40 centimeters away from the objective lens, and the second scene is placed around 50 centimeters away. Spectral signatures of the four highlighted points (P1-P4) will be compared in Figs. 7.14 and 7.15. . . 126

7.8 Compressive measurements as seen by the ToF sensor when the target scenes are illuminated with broadband fluorescent white light and the DMD exhibit a random coded pattern. 127

7.9 Sequence of four images g_0, g_1, g_2, g_3 captured with the testbed, in ToF mode using $f = 12$ MHz, for (a) Scene 1, and (b) Scene 2. 127

7.10 Reconstructed data cubes mapped to RGB, when $K = 1, 2$ and 4 CSI snapshots are captured, using none, grayscale or RGB side information for Scene 1. 129

7.11 Reconstructed data cubes mapped to RGB, when $K = 1, 2$ and 4 CSI snapshots are captured, using none, grayscale or RGB side information for Scene 2. 130

7.12 Reconstruction results of 4 out of the 8 spectral bands, for a single snapshot, using none, grayscale or RGB side information for Scene 1. . 131

7.13 Reconstruction results of 4 out of the 8 spectral bands, for a single snapshot, using none, grayscale or RGB side information for Scene 2. . 132

7.14 Comparison of the spectrum reconstructions when $K = 1$ snapshots are captured with no side information, and with side information from ToF, grayscale or RGB camera. Four different points (P1 to P4) from the first target scene in Fig. 7.7(a) were measured by an spectrometer and compared against the reconstructed data cubes. 133

7.15 Comparison of the spectrum reconstructions when $K = 1$ snapshots are captured with no side information, and with side information from ToF, grayscale or RGB camera. Four different points (P1 to P4) from the second target scene in Fig. 7.7(b) were measured by an spectrometer and compared against the reconstructed data cubes. 134

7.16 Estimated depth maps using frequencies $f = 6$ MHz, $f = 12$ MHz, and $f = 24$ MHz. (a) Scene 1. (b) Scene 2. Units are in centimeters. 136

ABSTRACT

Multimodal imaging refers to the framework of capturing images that span different physical domains such as space, spectrum, depth, time, polarization, and others. For instance, spectral images are modeled as 3D cubes with two spatial and one spectral coordinate. Three-dimensional cubes spanning just the space domain, are referred as depth volumes. Imaging cubes varying in time, spectra or depth, are referred as 4D-images. Nature itself spans different physical domains, thus imaging our real world demands capturing information in at least 6 different domains simultaneously, giving turn to 3D-spatial+spectral+polarized dynamic sequences. Conventional imaging devices, however, can capture dynamic sequences with up-to 3 spectral channels, in real-time, by the use of color sensors. Capturing multiple spectral channels require scanning methodologies, which demand long time. In general, to-date multimodal imaging requires a sequence of different imaging sensors, placed in tandem, to simultaneously capture the different physical properties of a scene. Then, different fusion techniques are employed to mix all the individual information into a single image. Therefore, new ways to efficiently capture more than 3 spectral channels of 3D time-varying spatial information, in a single or few sensors, are of high interest.

Compressive spectral imaging (CSI) is an imaging framework that seeks to optimally capture spectral imagery (tens of spectral channels of 2D spatial information), using fewer measurements than that required by traditional sensing procedures which follows the Shannon-Nyquist sampling. Instead of capturing direct one-to-one representations of natural scenes, CSI systems acquire linear random projections of the scene and then solve an optimization algorithm to estimate the 3D spatio-spectral data cube by exploiting the theory of compressive sensing (CS). To date, the coding procedure in CSI has been realized through the use of "block-unblock" coded apertures, commonly implemented as chrome-on-quartz photomasks. These apertures block or permit to pass the entire spectrum from the scene at given spatial locations, thus modulating the spatial characteristics

of the scene. In the first part, this thesis aims to expand the framework of CSI by replacing the traditional block-unblock coded apertures by patterned optical filter arrays, referred as "color" coded apertures. These apertures are formed by tiny pixelated optical filters, which in turn, allow the input image to be modulated not only spatially but spectrally as well, entailing more powerful coding strategies. The proposed colored coded apertures are either synthesized through linear combinations of low-pass, high-pass and band-pass filters, paired with binary pattern ensembles realized by a digital-micromirror-device (DMD), or experimentally realized through thin-film color-patterned filter arrays. The optical forward model of the proposed CSI architectures will be presented along with the design and proof-of-concept implementations, which achieve noticeable improvements in the quality of the reconstructions compared with conventional block-unblock coded aperture-based CSI architectures.

On another front, due to the rich information contained in the infrared spectrum as well as the depth domain, this thesis aims to explore multimodal imaging by extending the range sensitivity of current CSI systems to a dual-band visible+near-infrared spectral domain, and also, it proposes, for the first time, a new imaging device that captures simultaneously 4D data cubes (2D spatial+1D spectral+depth imaging) with as few as a single snapshot. Due to the snapshot advantage of this camera, video sequences are possible, thus enabling the joint capture of 5D imagery. It aims to create super-human sensing that will enable the perception of our world in new and exciting ways. With this, we intend to advance in the state of the art in compressive sensing systems to extract depth while accurately capturing spatial and spectral material properties. The applications of such a sensor are self-evident in fields such as computer/robotic vision because they would allow an artificial intelligence to make informed decisions about not only the location of objects within a scene but also their material properties.

Chapter 1

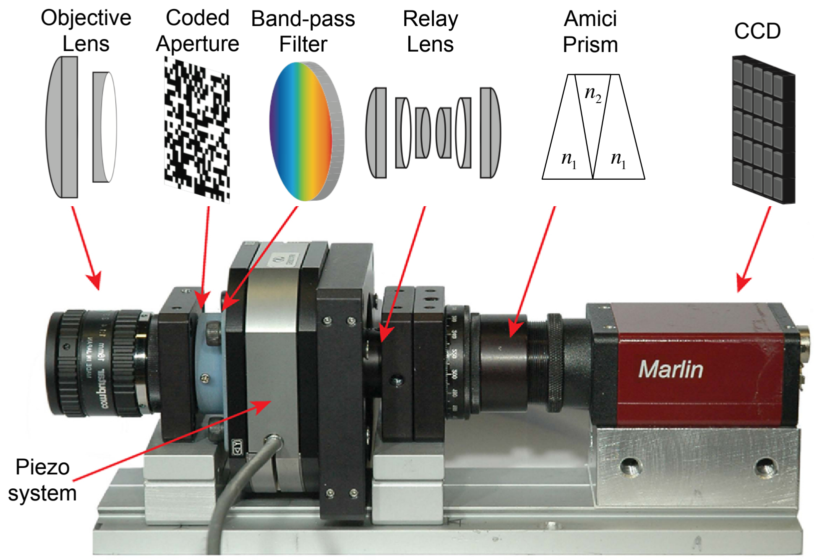
INTRODUCTION

A spectral image is a three-dimensional (3D) image cube comprised of a collection of two-dimensional (2D) images, where each 2D image is captured at a specific wavelength. Imaging spectroscopy, commonly referred to as multispectral, hyperspectral or simple as spectral imaging, is the process of sensing this kind of images. Spectral images permit to analyze spectral information about each spatial point in a scene, and thus can be valuable to identify different materials that appear in the scene [4]. Therefore, spectral imaging has applications in different areas such as medical imaging [5, 6], remote sensing [7, 8, 9, 10, 11], geology [12], and astronomy [13]. Conventional approaches to spectral imaging scan adjacent zones of the underlying spectral scene and merge the results to construct a spectral data cube. Examples of these spectral imagers include whisk-broom scanners, push-broom scanners [7], and filter-based spectrometers [14]. In whisk-broom scanners, a mirror reflects light onto a single line detector, so that one pixel of data is collected at a time; in push-broom scanners, an image cube is captured with one focal plane array (FPA) measurement per spatial line of the scene; and in filter-based spectrometers, a set of optical bandpass filters are tuned in steps in order to scan the scene band-per-band. The disadvantages of these techniques are that the data acquisition takes a long time, because they require scanning a number of zones linearly in proportion to the desired spatial and spectral resolution, and at the same time, large amounts of data are acquired and must be stored, processed or transmitted. For instance, for a megapixel camera (10^6 pixels) that captures a few hundred spectral bands (> 100 spectral channels) at 8 or 16 bits per frame, conventional spectral imagers demand roughly 10 megabytes per raw spectral image, and thus require space on the order of gigabytes for transmission or storage, which exceeds existing streaming capabilities.

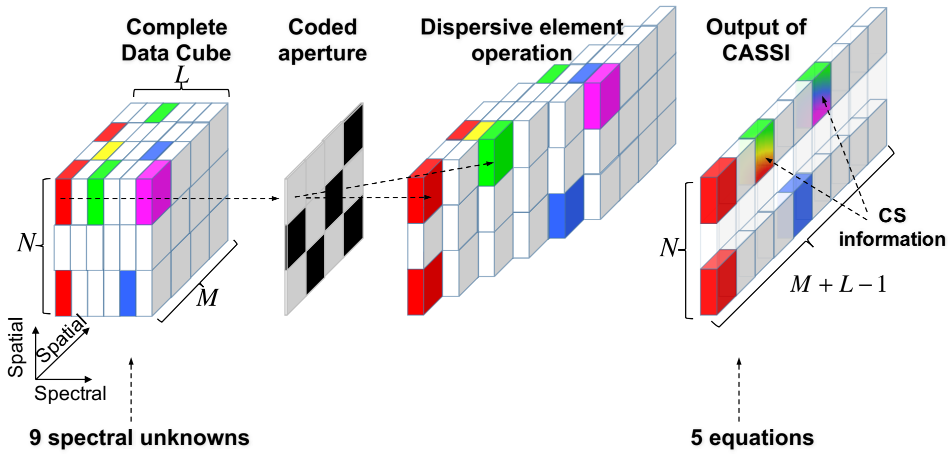
To address the limitations of conventional spectral imaging, non-scanning optical devices were proposed and summarized in [15, 16]. Although some of them require bulky and fancy equipment, the schemes based on compressive sensing (CS) [17, 18, 19, 20, 21] stand out [1, 22, 23]. The latter schemes have been called compressive spectral imagers (CSI). CS dictates that one can recover spectral scenes from far fewer measurements than that required by conventional linear scanning spectral sensors. To make this possible, CS relies on two principles: sparsity, which characterizes the spectral scenes of interest, and incoherence, which shapes the sensing structure [17, 19]. Sparsity indicates that spectral images found in nature can be concisely represented in some basis with just a small number of coefficients. This is indeed the case in spectral imaging where natural scenes exhibit correlation among adjacent pixels and also across spectral bands [7]. Incoherence refers to the structure of the sampling waveforms used in CS, which, unlike the signals of interest, have a dense representation in the chosen basis [19]. The remarkable discovery behind CS is that it is possible to design sensing protocols capable of capturing the essential information content in sparse signals with just a small number of compressive measurements. The signals of interest are then accurately reconstructed from the small number of compressive measurements by numerical optimization [17, 19, 24, 25, 26, 27, 28, 29].

The remarkable advantage of CS-based spectral imagers is that the entire spectral data cube is captured with few measurements and in some cases with as little as a single snapshot. One of the first proposed CSI architectures is the coded aperture snapshot spectral imager (CASSI). The CASSI instrument is a popular compressive spectral imager that acquires image data from different wavelengths simultaneously. It was first developed in [1] and it is shown in Fig. 1.1(a). The sensing physical phenomena in CASSI is strikingly simple, yet it adheres to the principles required by CS. CASSI measurements are realized optically by a coded aperture, a dispersive element and a focal plane array (FPA) detector [1, 30]. The coding is applied to the (spatial-spectral) image source density by means of a coded aperture as depicted in Fig. 1.1(b). The resulting coded field is subsequently decomposed by a dispersive element before it impinges onto the sensor. The compressive measurements across the sensor are realized by the integration of the coded and dispersed field.

In particular, the CASSI sensing mechanism is illustrated by the discretized model shown in Fig. 1.1(b), where the spectral data cube having L spectral bands and $N \times M$



(a)



(b)

Figure 1.1: (a) CASSI optical architecture proposed in [1]. (b) Light propagation through the CASSI architecture.

spatial pixels is first amplitude modulated by an $N \times M$ coded aperture. In this case, the coded aperture is a black-and-white coded aperture such that the energy along an entire row of the data cube is "punched out" when a "black" coded aperture element is encountered. As the coded field transverses the prism, it is then spatially sheared along one spatial axis. In essence, each coded image plane is shifted along the x-axis where the amount of shifting increases with the wavelength. Finally, the coded and dispersed field is "collapsed" in the spectral dimension by the integration of the energy impinging on each detector element over its spectral range sensitivity. The integrated field is then measured by just $N \times M + L - 1$ sensor elements. Note that conventional spectral imagers will collect the full NML measurements, that is, $\sim L$ times more data. Therefore, the acquisition time, storage space, and required bandwidth for transmission are reduced by CASSI.

To obtain an estimate of the spectral image from the compressed measurements is challenging because reconstructing 3D image cubes requires high processing. Fortunately, it is possible to reconstruct the 3D cube from the 2D measurements according to CS, because the 2D images from different wavelengths are highly correlated, and thus, the 3D image cube is sparse in an appropriate transform domain. The latter means that only a small portion of the transform coefficients have large values, and recovering them is enough to attain a good reconstructed image. Several numerical algorithms have been proposed to solve the inverse problem entailed by CSI. These can be grouped into one of six computational approaches: greedy pursuit iterative algorithms, such as Orthogonal Matching Pursuit and CoSaMP [31]; convex optimization methods, such as l1-magic software, SpaRSA [32], and GPSR [29]; Bayesian methods [26]; approximate message passing (AMP) [33] and non-convex optimization algorithms which attempts to find the solution by trying all possible support sets.

1.1 Motivation of the Thesis

Computational imaging systems comprise the new advent of imaging devices. These new systems require the integration of optical devices, electronic interfaces, physical and mathematical modeling, and computational algorithms for inverting complex sensing models. Compressive spectral imaging is a specific case of computational imaging systems which seeks to efficiently capture spectral imagery. However, by applying proper coding strategies, CSI can be further extended to multiplex additional spectral information in the ultraviolet and infrared domain, as well as to provide new multidimensional capabilities such as depth estimation. This thesis builds over these opportunities by developing new optical coding strategies, including the modeling, simulation, implementation and testing of new optical devices, that permit to optimally capture wide-band visible-and-near-infrared spectral images, as well as provide new multidimensional sensing capabilities such as spectral+depth imaging in a single imaging device.

1.2 Organization and Overview of the Thesis

This thesis contains six chapters, in addition to the introductory chapter. Each chapter details the corresponding system design, proper coding strategy used, mathematical forward model for the sensing process and inverse model for the reconstruction, system calibration, experimental setup, results, error analysis, conclusions and outcomes of each proposed coding or imaging device.

Chapters 2 to 4 present a new family of coding devices to be used in CSI architectures, called color coded apertures. These new color coded apertures replace the traditional block-unblock coded apertures in CASSI. Color coded apertures are build as 2D arrays of pixelated optical filters. The use of these new apertures entails richer coding strategies as not only spatial but also spectral coding is now performed in a single step, thus leading to less ill-posed problems which in turn provide better spectral image estimations. In particular, in Chapter 2 colored coded apertures are synthesized through linear combinations of low-pass, high-pass and band-pass filters, paired with binary pattern ensembles realized by a digital-micromirror-device (DMD). The DMD-based implementation rises as an important way to alleviate the costly fabrication of the color coded apertures.

In Chapter 3, a real color coded aperture is fabricated with lithographic thin-film-based procedures, and the testbed developed in Chapter 2 is updated by incorporating this new device. The color coded aperture, so-called patterned filter array, is designed and fabricated at micrometer pitch size thin-films. In this chapter we also realize that the number of different filters to be used and the filters spectral profile have to be carefully designed. We note that, the higher the number of different filters, the higher the fabrication cost. However, we observed in practice that the gains gradually decrease as the number of filters increase, as it is shown in Chapter 2. In fact, only a few different type of optical filters are typically needed to obtain significant improvement over imaging methods that use conventional block-unblock coded apertures.

Given that simple random realizations of the entries of the color coded apertures suffice to satisfy the CS requirements, but do not fully exploit the correlations withing spectral images, in Chapter 4, an optimization strategy for the correct distribution of the entries of a color coded apertures is presented. The proposed optimization seeks for a better conditioned sensing matrix, so that it satisfies tightly the restricted isometry property (RIP) required by CS. In that chapter we show that the selection of the spectral characteristics of each pixel, as well as the spatial distribution of them within the 2D matrix, control the quality of the captured compressive measurements, which in turn determine the quality of the spectral image to be estimated. To do this, a higher-order discretization model of the continuous phenomena within color coded aperture CSI systems is first proposed. This model entails a better approximation of the real phenomena, which in turn leads to a more precise, still structured, sparse sensing matrix. An iterative random-walk optimization algorithm is then developed to carefully design the color coded apertures, which minimizes the shifted cross-correlation between the columns of the, more precise, sparse sensing matrix. Exploiting the fact that a color coded aperture can be modeled as a 3D black-and-white coded aperture, the optimization algorithm thus seeks to spread the translucent elements (band-pass filters of the color coded aperture) along the 3D cube extent. The latter is achieved by 3D filtering the coded aperture with a 3D Euclidean filter, in each iteration of the random-walk.

On the other hand, Chapter 5 proposes the use of RGB-patterned sensors as the sensing device for CSI systems. In particular, in that chapter we expand the theoretical framework of CASSI to include the spectral sensitivity of the image sensor pixels to account for color and then investigate the impact on image quality using either a traditional color Bayer-patterned image sensor that spatially multiplexes red, green, and blue light filters or a novel Foveon-X3 image sensor which stacks red, green, and blue pixels on top of one another. We established both theoretical and experimental evaluations of how using a readily available color sensor improves upon the so far published works that rely on monochrome image sensors.

In Chapter 6, we expand the range sensitivity of the CASSI system from the visible electromagnetic spectrum to the near-infrared domain, due to the rich information contained in the infrared spectrum. This novel architecture aims to cover a wider-band of the electromagnetic spectrum in order to facilitate the identification, detection, and classification of objects. Particularly, we mathematically model and demonstrate the implementation of a broadband CASSI system covering the visible and the near infrared spectra between 448 nm to 1436 nm. This system uses a VIS-NIR dichroic mirror to split the underlying incoming coded energy into two imaging arms, each one accounting for an independent compressive spectral imager.

Finally, Chapter 7 presents a new imaging device that extends the multimodal sensing capabilities of CSI to jointly capture spectral+depth imagery. In particular, that chapter presents the development of a compressive spectral + depth imaging camera that employs a commodity 3D range time-of-flight (ToF) sensor as the sensing device of a coded-aperture-based compressive spectral imager. The proposed system uses a single aperture/single sensor, thus representing a significant improvement over existing RGB+D cameras that integrate two separate image sensors, one for RGB and another for depth. This new imaging device is made possible after the realization that ToF sensors can exploit both ambient light and modulated light, in order to recover pixelated depth and multispectral information.

Chapter 2

DMD-BASED IMPLEMENTATION OF PATTERNED OPTICAL FILTER ARRAYS FOR COMPRESSIVE SPECTRAL IMAGING

Compressive spectral imaging (CSI) systems acquire coded and dispersed random projections of the scene rather than direct measurements of the voxels. To date, the coding procedure in CSI has been realized through the use of "block-unblock" coded apertures, commonly implemented as chrome-on-quartz photomasks. These apertures block or permit to pass the entire spectrum from the scene at given spatial locations, thus modulating the spatial characteristics of the scene. This chapter extends the framework of CSI by replacing the traditional block-unblock photomasks by patterned optical filter arrays, referred as "colored" coded apertures. These in turn, allow the source to be modulated not only spatially but spectrally as well entailing more powerful coding strategies. The proposed colored coded apertures are synthesized through linear combinations of low-pass, high-pass and band-pass filters, paired with binary pattern ensembles realized by a digital-micromirror-device (DMD). The optical forward model of the proposed CSI architecture is presented along with a proof-of-concept implementation which achieves noticeable improvements in the quality of the reconstructions.

2.1 Introduction

Spectral imaging (SI) techniques sense the two-dimensional (x, y) spatial information across a range of spectral wavelengths (λ) of a scene. Knowledge of the spectral content at various spatial locations from a scene can be valuable in identifying the composition and structure of objects of interest in the scene. SI has therefore been widely used in areas such as remote sensing [9], artwork conservation [34], and biomedical imaging [5]. Conventional SI sensors use temporal scanning either spectrally or spatially and merge the results to construct a spatio-spectral datacube [10]. These techniques are suitable for

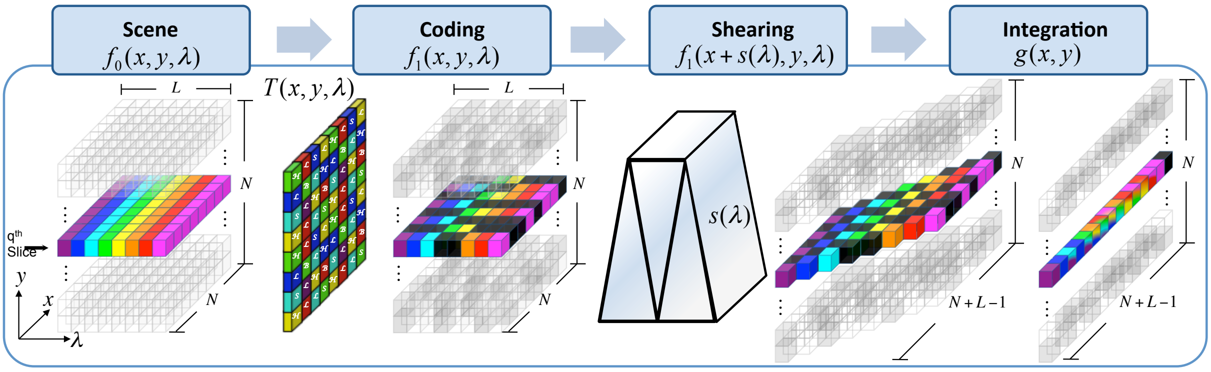


Figure 2.1: Colored coded aperture-based compressive spectral imager (C-CASSI). The 3D spectral scene $f_0(x, y, \lambda)$ is 3D-coded by the CCA $T(x, y, \lambda)$ and horizontally sheared by the dispersive element with dispersion function $S(\lambda)$, before being integrated by the focal plane array (FPA) detector $g(x, y)$. Notice the 3D coding scheme entailed by the CCA.

static scenes, however, it complicates and limits subsequent image processing and analysis of dynamic scenes due to the artifacts induced by the overlapping of the scanning operation. Their principal disadvantage is that they require scanning a number of regions that grows linearly in proportion to the desired spatial or spectral resolution. In contrast, compressive spectral imaging (CSI) techniques [30], first capture 2D coded projections of the underlying scene, and then, it recovers an estimate of the 3D datacube exploiting the fact that spectral images are highly correlated and admit sparse representations.

The Coded Aperture Snapshot Spectral Imager (CASSI) has been proposed recently as a CSI sensor [1, 35, 36], which utilizes a binary coded aperture and a dispersive element to modulate the optical field from a scene. It encodes the 3D spectral information into a 2D representation. Coded apertures in CASSI have been fabricated using materials such as chrome-on-quartz, rendering coded aperture elements that are either opaque or translucent to the whole wavelengths of interest. These coded apertures are referred to as photomasks, binary, or "block-unblock" coded apertures (CAs). Recent advances in micro-lithography and coating technologies [37, 38] have allowed the design of patterned optical coatings which are used to create multi-patterned arrays of different optical filters, here referred to as "colored" coded apertures (CCAs). CCAs have been used in applications such as deblurring and matting [39], multi-focusing and depth estimation [40]. Further, their use in CSI has been recently introduced in [41]. The CCAs entail richer coding strategies, as low-pass ($\mathcal{L}$), high-pass ($\mathcal{H}$), band-pass ($\mathcal{B}$) or any kind of filter allows modulating the scene not only spatially but spectrally as well. In this chapter a colored-CASSI (C-CASSI) architecture is proposed, where the CA in CASSI is replaced by a CCA. In consequence, every pixel from the aperture permit to pass just a desired set of wavelengths, given by the filter wavelength transmission range. Incorporating a wavelength-dependent coding procedure produces a spectrally richer optical system. The sensing richness can be mathematically understood as the increasing the diversity of unknowns components (voxels) per equation (detector pixel). Figure 2.1 presents the C-CASSI system, and its effect on an input spectral data cube.

Although the use of CCAs offers significant benefits, their fabrication requires new lithographic and thin film technology. A CCA is fabricated by attaching different optical patterned thin films, each one accounting for a different optical filter. Increasing the number of different type of filters in a CCA entails richer spatio-spectral structures, however, the associate cost increases as well. In order to exploit the advantages of the CCA, without incurring in higher fabrication costs, this chapter describes a method to synthesize a CCA, by use of a digital-micromirror-device (DMD) synchronized with a set of conventional optical filters attached to a rotation wheel. As explained shortly, a CCA can be modeled as the sum of several CA paired with different ($\mathcal{L}$), ($\mathcal{H}$) or ($\mathcal{B}$) filters as depicted in Fig. 2.2. Multiple coded-and-filtered projections will thus be multiplexed in a single snapshot.

In C-CASSI, a snapshot measurement compresses a scene by $\frac{1}{L}$, where L is the number of spectral bands. This compression ratio is high for certain applications, such that multiple snapshots are employed to reduce the compression ratio to $\frac{K}{L}$, where K is the number of snapshots, and $K < L$. Each snapshot uses a different CCA, which remains fixed during the integration time of the detector. Traditionally, multiple snapshots are captured by moving a thin-film-based CCA back and forth using a piezo-electric device. This nanopositioning action requires a high-precision realignment procedure. In contrast, the DMD-based CCA has tilting capabilities up to frame rates, without requiring re-alignment. Therefore, the DMD-based CCA offers a more robust and reliable multiple

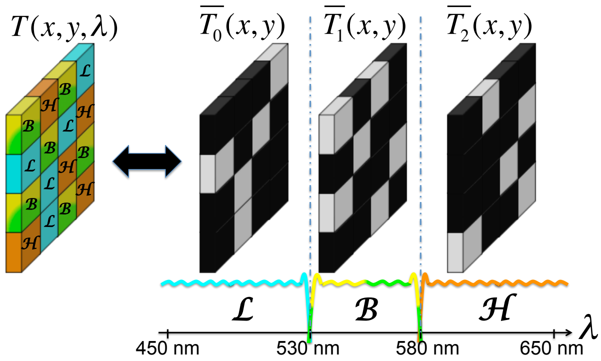


Figure 2.2: Synthesis of a 3-color CCA. The CCA is indistinctly modeled by 3 different CA $\overline{T}_0(x, y)$ (filtered by a low-pass filter), $\overline{T}_1(x, y)$ (filtered by a band-pass filter), and $\overline{T}_2(x, y)$ (filtered by a high-pass filter).

snapshot CSI system, as well as a reduced cost.

In the following, the proposed C-CASSI optical model is presented along with its mathematical discretization. Afterwards, the system matrix model and the structure of its transfer function are presented. Next, the multiple snapshot extension is introduced along with its inverse problem formulation. The simulations section presents an analysis of the impact of the number of snapshots, as well as, the number of colors in the CCA, in the attained reconstructions. At the end, a proof-of-concept DMD-based C-CASSI experiment is described and compared against the traditional block-unblock CASSI.

2.2 System Model

The C-CASSI, depicted in Fig. 2.1, is composed by five optical elements: an objective lens that focuses the scene onto the CCA, which performs the coding, an imaging lenses which relays the colored-coded scene onto the FPA image plane, a dispersive element (usually a prism) that shears horizontally the coded field, and an FPA detector which integrates the colored-coded and sheared spectral scene. The coding transfer function exhibited by the CCA is denoted as

$$T(x, y, \lambda) = \sum_{i,j,k} T_{i,j,k} \text{rect}\left(\frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j, \frac{\lambda}{\Delta_d} - k\right) \tag{2.1}$$

where $T_{i,j,k} \in \{0, 1\}$ represents the filtering performed on the $(i, j, k)^{th}$ data cube voxel; $i, j \in [0, \ldots, N-1]$, and $k \in [0, \ldots, L-1]$ index the coordinates of an $N \times N \times L$ spectral data cube, and Δ_c, Δ_d account for the pixel sizes of the CCA and the FPA detector, respectively. Notice that the spatial resolution of the resolvable scene is determined by the resolution of the CCA (Δ_c), while the spectral resolution depends on the resolution of the FPA (Δ_d) and the dispersion efficiency of the prism. Defining the k^{th} 2D matrix T^k having entries $T^k_{i,j} = T_{i,j,k}$, and exploiting the properties of the rectangular function (rect()), Eq. (3.2) can be alternatively expressed as,

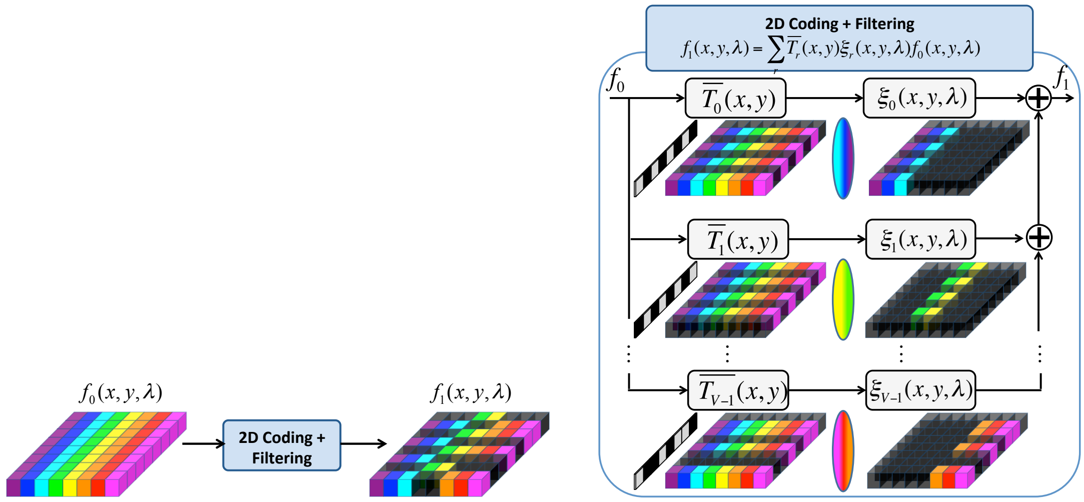
$$\begin{aligned} T(x,y,\lambda) &= \sum_{i,j,k} T^k_{i,j} \text{rect}\left(\frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j\right) \text{rect}\left(\frac{\lambda}{\Delta_d} - k\right) \\ &= \sum_{k} T_k(x,y)\beta_k(\lambda), \end{aligned} \tag{2.2}$$

where $T_k(x,y) = \sum_{i,j} T^k_{i,j}\text{rect}(\frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j)$ can be seen as a block-unblock CA with entries defined by $T^k_{i,j}$, and $\beta_k(\lambda) = \text{rect}(\frac{\lambda}{\Delta_d} - k)$ can be seen as a single-band-pass filter ($\mathcal{B}$) with transmission in the wavelength range $[\lambda_{k-1}, \lambda_k]$. In this way, a CCA can be implemented by synchronizing a set of L block-unblock CA $\{T_k(x,y)\}_{k=0}^{L-1}$, with a set of L band-pass filters $\{\beta_k(\lambda)\}_{k=0}^{L-1}$, during the integration time of the FPA detector.

The model in Eq. (2.2) can be further simplified when the set of single-band-pass filters $\beta_k(\lambda)$ are grouped in a set of user-defined filters. In particular, a low-pass filter ($\mathcal{L}$) with cut-off wavelength λ_c can be represented by setting $\mathcal{L}_c(\lambda) = \sum_{k=0}^{c-1}\beta_k(\lambda)$. A high-pass filter ($\mathcal{H}$) with cut-off wavelength λ_c can by represented as $\mathcal{H}_c(\lambda) = \sum_{k=c}^{L-1}\beta_k(\lambda)$. Similarly, a band-pass filter with cut-off wavelengths λ_{c_1} and λ_{c_2} can be represented as $\mathcal{B}_{c_1,c_2}(\lambda) = \sum_{k=c_1}^{c_2-1}\beta_k(\lambda)$. Let the user-defined set of filters be $\{\xi_r(\lambda)\}_{r=0}^{V-1}$, where $\xi_r(\lambda) \in \{\mathcal{L}_c(\lambda), \mathcal{H}_c(\lambda), \mathcal{B}_{c_1,c_2}(\lambda)\}$, and $V \leq L$. Therefore, Eq. (2.2) can be re-written as,

$$T(x,y,\lambda) = \sum_{r=0}^{V-1} \overline{T}_r(x,y)\xi_r(\lambda), \tag{2.3}$$

where $\overline{T}_r(x,y) = \sum_{i,j} T^r_{i,j}\text{rect}(\frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j)$, such that, $T^r_{i,j} = T^k_{i,j} = \ldots = T^{k'}_{i,j}$, for $k, \ldots, k' \in \xi_r(\lambda)$. Therefore, the set of V block-unblock CA $\{\overline{T}_r(x,y)\}_{r=0}^{V-1}$ can be implemented using a DMD, and the set of corresponding V filters $\{\xi_r(\lambda)\}_{r=0}^{V-1}$ can be changed by means of a filter wheel, during the FPA integration time. The spatial-spectral coding entailed by the CCA resulting in the spectral field $f_1(x,y,\lambda)$, is thus synthesized through



(a) (b)

Figure 2.3: Synthesis of the CCA effect on the $y = q^{th}$ slice of the data cube illustrated in Fig. 1. (a) Input-output relationship of the CCA synthesis. (b) 2D coding and filtering process. The slice is first coded by the q^{th} row of the $\overline{T}_0(x, y)$ CA, and then filtered by the $\xi_0(\lambda)$ optical filter; this process is realized for every r^{th} CA/filter pair, and the resulting spectra are combined.

multiple 2D block-unblock coding, paired with multiple spectral filtering as depicted in Fig. 2.3. A CCA can thus be regarded as a 3D photomask of length V, where each 2D block-unblock CA $\overline{T}_r(x, y)$ is independently applied to a certain set of wavelengths given by $\xi_r(\lambda)$. Note that $V \leq L$, that is, the number of filters used in the CCA are much less than the number of recoverable spectral bands.

Define the ratio between the number of ones and the total number of elements in the CCA as the transmittance (R), which is given by $R = \sum_{i,j,k} T_{ijk}/N^2L$. The transmittance, refers to the proportion of voxels from the scene that will propagate through the optical system. This parameter is of key importance in the performance of CSI systems, since it controls the ill-conditioning of the inverse problem. The 3D coding applied by the CCA to the spatio-spectral density source $f_0(x, y, \lambda)$ results in the coded field $f_1(x, y, \lambda) = T(x, y, \lambda)f_0(x, y, \lambda)$. After the coding-and-filtering stage, the spectral field is sheared by the prism, whose output can be expressed as $f_2(x, y, \lambda) = \iint f_1(x', y', \lambda) \times h(x - x' - S(\lambda), y - y')dx'dy'$, where $h(x - x' - S(\lambda), y - y')$ accounts for the optical impulse response of the system, and $S(\lambda)$ represents the dispersion, which is restricted to occur just in the horizontal axis. Finally, the coded-filtered-and-sheared spectral density being integrated by the detector can be expressed as $g(x,y) = \int f_2(x,y,\lambda)d\lambda$, which may be rewritten as,

$$g(x,y) = \iiint \sum_r \overline{T}_r(x',y')\xi_r(\lambda)f_0(x',y',\lambda)h(x-x'-S(\lambda),y-y')dx'dy'd\lambda. \tag{2.4}$$

Assuming the optical impulse response is an ideal delta function, Eq. (2.4) can be succinctly expressed as,

$$g(x,y) = \int \sum_r \overline{T}_r(x+S(\lambda),y)\xi_r(\lambda)f_0(x+S(\lambda),y,\lambda)d\lambda. \tag{2.5}$$

2.2.1 Discrete Model

The integration of the continuous field $g(x,y)$ given in Eq. (2.5) in a single $(n,m)^{th}$ detector pixel can be expressed as

$$G_{n,m} = \iint g(x,y)\text{rect}\left(\frac{x}{\Delta_d}-m,\frac{y}{\Delta_d}-n\right)dydx + \omega_{n,m} \tag{2.6}$$

where $\omega_{n,m}$ accounts for the noise of the capturing process, $n = 0,\ldots,N-1$ and $m = 0,\ldots,N+L-2$ index the pixels on the FPA, where N is the total number of rows and $N+L-1$ is the total number of columns. Replacing Eq. (2.5) in Eq. (2.6) leads to

$$G_{n,m} = \iiint \sum_r \overline{T}_r(x+S(\lambda),y)\xi_r(\lambda)f_0(x+S(\lambda),y,\lambda)\text{rect}\left(\frac{x}{\Delta_d}-m,\frac{y}{\Delta_d}-n\right)d\lambda dydx + \omega_{n,m}. \tag{2.7}$$

Assume that the response of the dispersive element is linear over the spectral range of the system, such that $S(\lambda_{k+1}) - S(\lambda_k) = \frac{\Delta_d}{\Delta_c}$. Assume also that $\Delta_c = \Delta_d$, and that ideal optical elements are used. Therefore, the source density at the $(n,m-k,k)^{th}$ voxel can be denoted as $F_{n,m-k,k} = \iiint_{\Omega_{n,m-k,k}} f_0(x,y,\lambda)dydxd\lambda$, where $\Omega_{n,m-k,k}$ represents the boundaries of the data cube region being integrated in the $(n,m)^{th}$ FPA pixel. Similarly, the discrete $(n,m)^{th}$ pixel from $\overline{T}_r(x,y)$ can be expressed as $\overline{T}^r_{n,m}$. Therefore, Eq. (2.7) can be rewritten as

$$G_{n,m} = \sum_k \sum_r \overline{T}^r_{n,m-k}(\xi_r)_k F_{n,m-k,k} + \omega_{n,m}, \tag{2.8}$$

where $(\xi_r)_k \in [0,1]$ is a discrete space-invariant value accounting for the effect of the r^{th} filter on the k^{th} spectral band. Let denote the k^{th} spectral band of the hyper-spectral data cube and each r^{th} coded aperture being implemented by the DMD, in vectorial form as,

$$(\mathbf{f}_k)_j = F_{(j-\lfloor \frac{j}{N} \rfloor N), \lfloor \frac{j}{N} \rfloor, k}\ , \tag{2.9}$$

$$\left(\overline{\mathbf{t}}_r\right)_j = \overline{T}^r_{(j-\lfloor \frac{j}{N} \rfloor N), \lfloor \frac{j}{N} \rfloor}\ , \tag{2.10}$$

for $j = 0, \ldots, N^2 - 1$, and $k = 0, \ldots, L-1$. The sensing process of the proposed optical system can be represented in matrix form as,

$$\mathbf{g} = \mathbf{Hf} = \mathbf{PTf}, \tag{2.11}$$

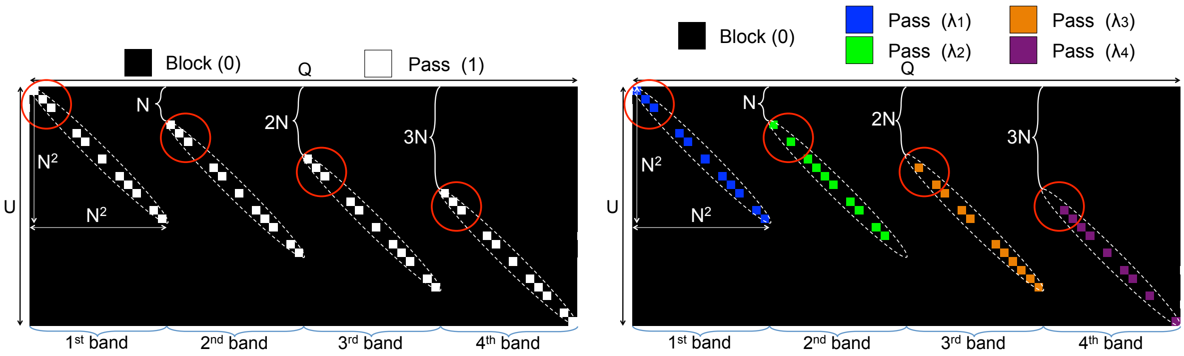
where $\mathbf{f} = [\mathbf{f}_0^T \mathbf{f}_1^T \ldots \mathbf{f}_{L-1}^T]^T$ is the complete data cube in vector form, and $\mathbf{H}$ is called the system transfer function which accounts for the coding ($\mathbf{T}$) and the dispersion ($\mathbf{P}$) processes. Let $Q = N^2 L$ be the total number of data cube voxels, and $U = N(N+L-1)$ be the total number of FPA pixels. Therefore, in Eq. (2.11), $\mathbf{T}$ is a $Q \times Q$ block-diagonal matrix, where each $N^2 \times N^2$ diagonal matrix accounts for the effect of every r^{th} CA/filter pair on the k^{th} spectral band. Similarly, the dispersion matrix $\mathbf{P}$ is a $U \times Q$ rectangular matrix, given by

$$\mathbf{P} = \begin{bmatrix} \text{diag}(\mathbf{1}_{N^2}) & \mathbf{0}_{N\times N^2} & \cdots & \mathbf{0}_{N\times N^2} \\ \mathbf{0}_{N\times N^2} & \text{diag}(\mathbf{1}_{N^2}) & \cdots & \mathbf{0}_{N\times N^2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N\times N^2} & \mathbf{0}_{N\times N^2} & \cdots & \text{diag}(\mathbf{1}_{N^2}) \end{bmatrix}_{U\times Q}\ , \tag{2.12}$$

where $\mathbf{0}_{N\times N^2}$ is a 0-valued $N \times N^2$ matrix, and $\text{diag}(\mathbf{1}_{N^2})$ is an 1-valued $N^2 \times N^2$ diagonal matrix. Notice that, the number of $\mathbf{0}_{N\times N^2}$ matrices in each row or column of Eq. (3.5) is $L-1$, and the number of $\text{diag}(\mathbf{1}_{N^2})$ matrices per row or column is 1.

Figure 2.4 compares the $\mathbf{H}$ matrices corresponding to the traditional CASSI and to the C-CASSI, respectively, using $V = L = 4$ different optical filters, for an $N =$

$L = 4$ spectral data cube. The colored diagonals correspond to each one of the block-unblock coded apertures filtered by the corresponding optical filter, while the black entries correspond to blocking elements. The red circles show that in the traditional CASSI the diagonal elements remain fix along the wavebands, while, they vary in the C-CASSI as the colored coding is wavelength dependent. This in turn, entails a richer coding strategy as the distribution of the voxels being integrated by a single FPA pixel is now manageable.



(a) Traditional CASSI (b) Colored CASSI

Figure 2.4: $\mathbf{H}$ matrices for a $4 \times 4 \times 4$ data cube. (a) Traditional CASSI, and (b) C-CASSI. The diagonal entries correspond to the coding pattern, which remain invariant with the wavelength in CASSI, but are wavelength dependent in C-CASSI.

In the proposed architecture, multiple-snapshot sensing is also attainable, leading to a less ill-posed inverse problem and consequently improved signal recovering [23, 42, 43]. In such a case, the colored coded aperture changes its coding pattern with every snapshot. Consequently, the system transfer function will contain U more rows per additional measurement snapshot, while the number of columns remains unchanged. That is, the matrix $\mathbf{H}$ will be of size $KU \times Q$, where K represents the number of snapshots to be captured. As a constraint, the number of snapshots K allowed to capture is such that $KU \leq Q$, being $KU = Q$ the extreme case, where the inverse problem is no longer under-determined. Particularly, denoting the ℓ^{th} FPA measurement as $\mathbf{g}_\ell = \mathbf{H}_\ell \mathbf{f}$, with $\mathbf{H}_\ell$ representing the effects of the ℓ^{th} colored coded aperture, the set of K FPA measurements is then assembled as $\mathfrak{g} = \left[(\mathbf{g}_0)^T, \ldots, (\mathbf{g}_{K-1})^T\right]^T$, which can be expressed as,

$$\mathfrak{g} = \mathbf{H}\mathbf{f}, \tag{2.13}$$

where $\mathbf{H} = [\mathbf{H}_0^T \mathbf{H}_1^T \dots \mathbf{H}_{K-1}^T]^T$.

2.2.2 Inverse Model

Due to the reduced number of pixels of the FPA compared with the number of voxels from the data cube to be estimated, a direct inversion of the transfer function $\mathbf{H}$ is not feasible. However, CSI theory has been proposed to solve this inverse problem. Particularly, CSI supposes that the hyperspectral signal $F \in R^{N \times N \times L}$, or its vector representation $\mathbf{f} \in R^Q$, is S-sparse when projected over a dictionary or basis $\mathbf{\Psi}$ (supp($\boldsymbol{\theta}$) $= \{j : |\boldsymbol{\theta}_j| > 0\} = S$, such that $\mathbf{f} = \mathbf{\Psi}\boldsymbol{\theta}$). In consequence, $\mathbf{f}$ can be approximated by a linear combination of S columns from $\mathbf{\Psi}$, with $S \ll Q$. Consequently, the measurements in the proposed architecture can be alternatively expressed as $\mathfrak{g} = \mathbf{H}\mathbf{\Psi}\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\theta}$, where $\mathbf{A} = \mathbf{H}\mathbf{\Psi}$ is known as the sensing matrix. Based on this assumption, $\mathbf{f}$ can be recovered from $\mathfrak{g}$ if $\mathbf{A}$ is carefully designed. CSI allows $\mathbf{f}$ to be recovered from compressive projections with high probability. The spectral data cube $\mathbf{f}$ is estimated as,

$$\hat{\mathbf{f}} = \mathbf{\Psi^{-1}} \text{argmin}_{\boldsymbol{\theta}'} \|\mathfrak{g} - \mathbf{A}\boldsymbol{\theta}'\|_2^2 + \tau \|\boldsymbol{\theta}'\|_1, \tag{2.14}$$

where $\tau > 0$ is a regularization parameter, which penalizes the seeking of the sparsest solution. The design of the sensing matrix $\mathbf{A}$ relies directly on the selection of the correct dictionary $\mathbf{\Psi}$ along with the structure of $\mathbf{H}$.

2.3 Simulations

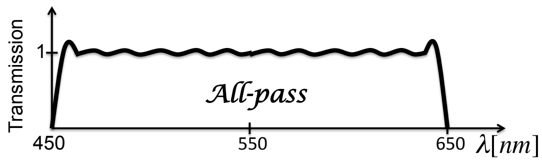
In order to compare the traditional CASSI system against the C-CASSI, a set of compressive measurements $\mathfrak{g}$ are constructed employing two different spectral data cubes depicted in Fig. 2.5. These data cubes were acquired using a wide-band Xenon lamp as the illumination source, modulated by a visible monochromator spanning the spectral range between 450nm and 650nm. The image intensity was captured using a grayscale CCD camera, with pixel size $\Delta_d = 9.9\mu m$, and 8 bits of intensity levels. The resulting test data cubes F have $N \times N = 256 \times 256$ pixels of spatial resolution and $L = 8$ spectral bands.

Figure 2.5: RGB targets along with their corresponding 8-band spectral data cubes. (Left) A scene exhibiting green and blue colors (Database 1). (Right) A scene exhibiting red and yellow colors (Database 2).

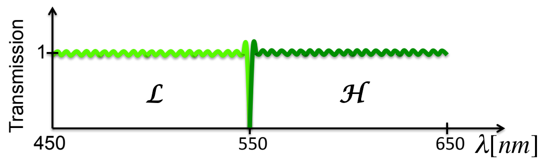
The impact of two important factors in the coding is analyzed. First, the transmittance (R) of the CAs and CCAs (usually fixed to be 50% [23]), and second, the number of different filters (V) employed by the CCAs. To analyze the transmittance effect, the entries of the block-unblock CA used in the traditional CASSI are generated by realizations of Bernoulli random variables with parameter p matching with the transmittance R being varied between 0.1 to 0.8. Transmittances above 0.8 are not analyzed due to they require the solution of an extremely ill-posed inverse problem, and below 0.1 are discarded due to the extremely low light efficiency. Similarly, the CCA entries are random realizations of $\mathcal{L}$, $\mathcal{H}$ or $\mathcal{B}$ optical filters, such that its transmittance is also varied between 0.1 and 0.8.

To analyze the effect of the number of different filters in the CCAs, 3 different configurations are evaluated using either $V = 2$, 4 or 8 optical filters as shown in Fig. 2.6. In particular, we use a 2-color CCA with either $\mathcal{L}$ or $\mathcal{H}$ filter entries with cut-off wavelength half of the spectrum under study ($\lambda_c = 550$ nm), as depicted in Fig. 2.6(b); a 4-color CCA with two $\mathcal{B}$ filters with twice the bandwidth of the spectral bands of the data cube, and a $\mathcal{L}$ and a $\mathcal{H}$ filters for the bands in the extremes, such that each one of them covers two of the eight spectral bands, as seen in Fig. 2.6(c); finally an 8-color CCA with entries being 6 different $\mathcal{B}$ filters, or a $\mathcal{L}$ and a $\mathcal{H}$ filter for the ending bands, with bandwidths matching the bands from the spectral data cube, as depicted in Fig. 2.6(d). The CAs and the CCAs are designed to exhibit 256×256 spatial pixels, matching with the FPA resolution ($\Delta_c = \Delta_d$).

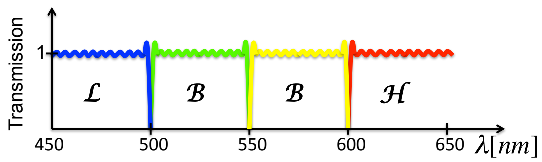
The compressive-sensing-based GPSR reconstruction algorithm [42, 43] is used to recover the underlying data cube $\hat{\mathbf{f}}$ by solving the problem stated in Eq. (4.27). The



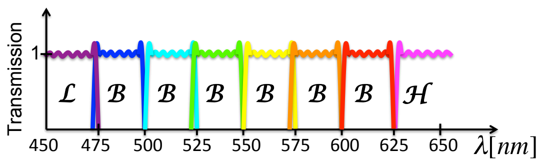
(a) Traditional CASSI



(b) C-CASSI with 2 filters



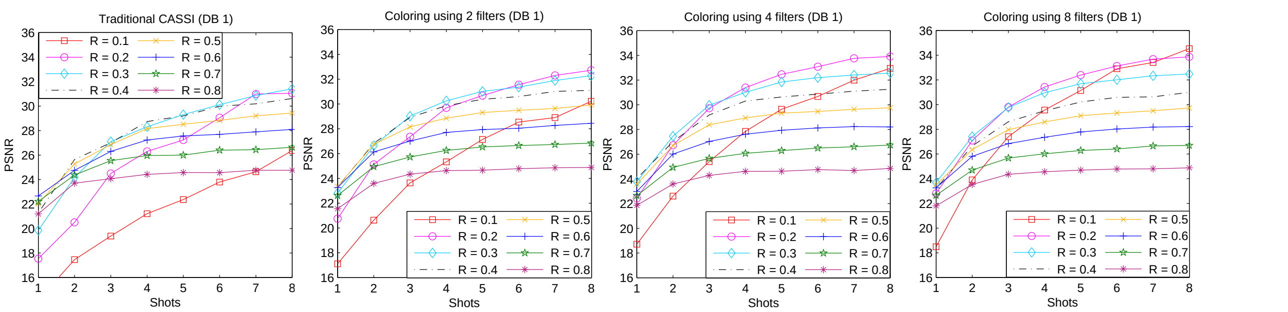
(c) C-CASSI with 4 filters



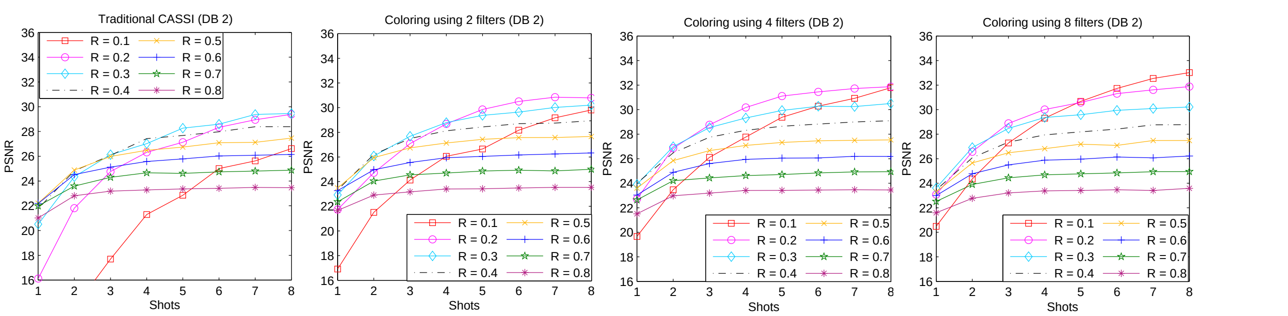
(d) C-CASSI with 8 filters

Figure 2.6: Spectral responses of the different filter configurations. (a) All-pass filter for the traditional CASSI, and (b) $V = 2$, (c) $V = 4$, and (d) $V = 8$, for C-CASSI.

basis representation $\mathbf{\Psi}$ is set to be the Kronecker product of the 2D-Wavelet Symmlet 8 basis and the 1D-DCT basis $\mathbf{\Psi} = \mathbf{\Psi}_{W2D} \bigotimes \mathbf{\Psi}_{DCT}$. Figure 2.7 presents a summary of the



(a) Database 1. (Left) CASSI (Center-left) CCA with 2-colors, (Center-right) CCA with 4 colors, (Right) CCA with 8 colors



(b) Database 2. (Left) CASSI (Center-left) CCA with 2-colors, (Center-right) CCA with 4 colors, (Right) CCA with 8 colors

Figure 2.7: Impact of the transmittance (R) in the reconstruction PSNR as a function of the number of snapshots (K).

reconstruction results obtained from the analysis of the effect of R, at different number of snapshots $K = 1, \dots, 8$, for the databases in Fig. 2.5, using the traditional CASSI and the C-CASSI with $V = 2, 4, 8$ colors. The peak-signal-to-noise-ratio (PSNR) is used as the measure of the reconstruction quality. It can be noticed in Fig. 2.7 that regardless of the database used, the reconstructed PSNR exhibits a similar behaviour for a given R value, as K increases. Further, the achieved PSNR using CCAs improves over that obtained with the traditional CASSI. In addition, note that the transmittances $R = 0.2, 0.3$ achieve the best reconstruction PSNRs overall. From the general behavior of the curves, it can be seen that for a given number of snapshots K, the best results are achieved when the transmittance R is selected to be roughly $R \simeq \frac{1}{K}$. Physically, this relationship entails the sensing of each data cube voxel at least and only once [41].

To analyze the effect of the amount of different filters V in the CCAs, Fig. 2.8 depicts the improvements in PSNR as a function of the number of optical filters, fixing $R = 0.2,\ 0.3$, respectively. It can be noticed that, although the PSNR increases as K increases, the richness acquired by the increasing amount of different filters, converges up to some point, being $V = 4$ colors enough to achieve similar PSNR results as with $V = 8$ colors. Note that, although richer coding is obtained with every new filter added, large transmittances R become increasingly difficult to achieve, due to the fact that the spectral responses of the filters do not overlap, as depicted in Fig. 2.6. Hence, when V filters are used, the spectrum is divided into V evenly distributed sub-spectra; therefore, when the $N \times N$ pixelated filters of the CCA are uniformly distributed, the total CCA transmittance is $1/V$ at most. For instance, if $V = 2$ non-overlapping filters are used, up to $R = 0.5$ would be possible. Then, to achieve $R > 0.5$, some of the $N \times N$ pixelated filters, must be replaced by "all-pass" filters. The resultant CCA coding becomes inefficient, due to, as R continues increasing, this pattern approaches that of the traditional CASSI, where just all-pass and all-block filters are employed

Figure 2.9 compares the best PSNR achieved by the two systems, as a function of the number of snapshots for $R = 0.2,\ 0.3$, respectively. It can be observed that the C-CASSI outperforms the traditional CASSI for any number V of colored filters and any amount of snapshots K. In addition, it can be also noticed the rapid convergence of using higher amounts of colored filters, as the $V = 4$ curve overlaps the $V = 8$ curve.

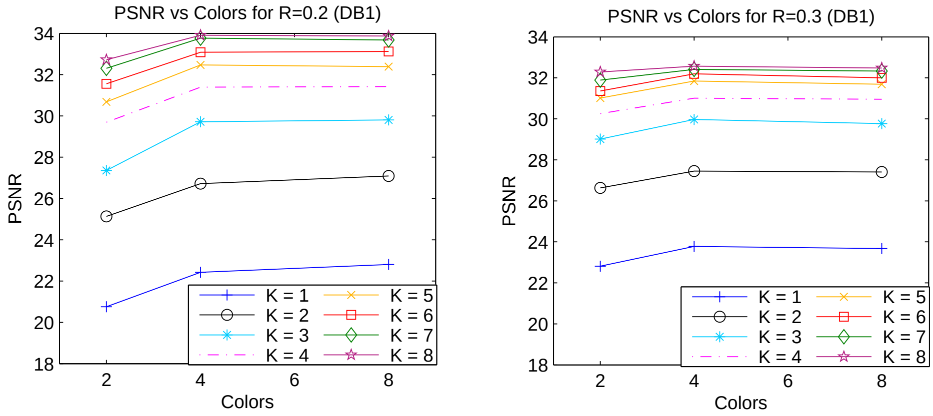


Figure 2.8: Impact of the number of different colors in the reconstruction PSNR for different number of snapshots and (Left) $R = 0.2$, (Right) $R = 0.3$.

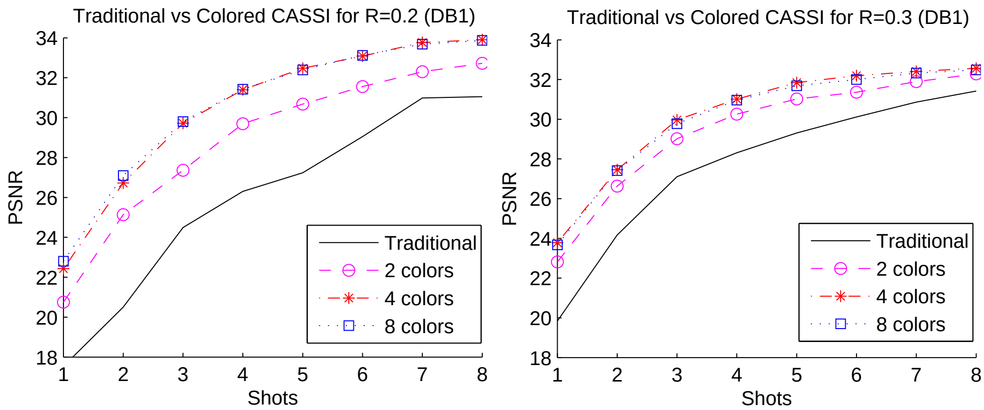


Figure 2.9: Comparison of the best achieved reconstruction PSNR between CASSI and C-CASSI (with $V = 2, 4, 8$) as a function of the snapshots, for fixed transmittances (Left) $R = 0.2$, (Right) $R = 0.3$.

The impact of R on the attained reconstructions is visualized in Fig. 2.10. It depicts the reconstructed data cubes as seen by an RGB camera when $K = 3$ snapshots are captured by both optical architectures, using $R = 0.1, 0.2, 0.3$. Notice that $R = 0.1$ is included just for a broad visual comparison, despite it does not achieve good PSNR. It can be noticed that the C-CASSI results outperform the traditional CASSI, and that the best attainable PSNR is achieved when $R \approx 0.3$. Quantitatively, C-CASSI achieves PSNR improvements up to 8 dBs for $R = 0.1$, 5.3 dBs for $R = 0.2$ and 2.7 dBs for $R = 0.3$. On the other hand, Fig. 2.11 compares the effect of the number of snapshots at a fixed transmittance. It shows the reconstruction results for $K = 1, 4, 8$, with $R = 0.3$.

(a) Database 1. (First row) CASSI, (Second row) C-CASSI with $V = 2$, (Third row) C-CASSI with $V = 8$

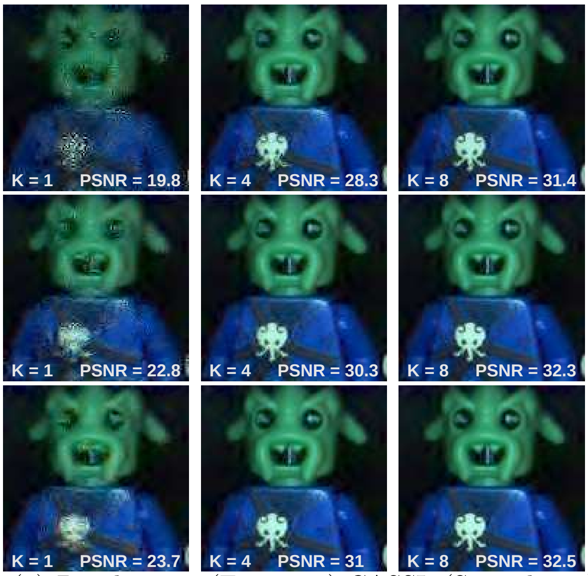
(b) Database 2. (First row) CASSI, (Second row) C-CASSI with $V = 2$, (Third row) C-CASSI with $V = 8$

Figure 2.10: Impact of the transmittance in the reconstructions, using $K = 3$ and varying $R = 0.1, 0.2, 0.3$ from left to right. (a) Database 1, (b) Database 2.

Remark, that the reconstruction quality increases as K increases, due to a different coding pattern is employed with every new snapshot. To validate the spectral fidelity of the reconstructions, 2 spatial points were randomly chosen from the two databases (P1 and P2 in Fig. 2.5), and their spectral signatures plotted in Fig. 2.12. It can be seen how the curves from the C-CASSI reconstructions best fit to the original signatures.

2.4 Experimental Setup

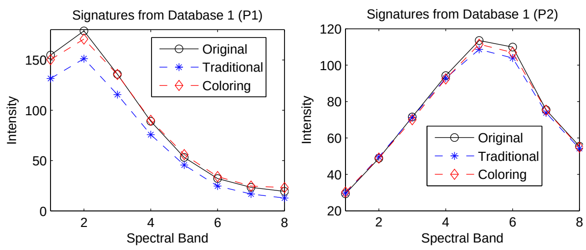
A proof-of-concept optical prototype was designed to experimentally verify the proposed architecture and it is presented in Fig. 2.13. This prototype consists of (i) a Leica COLORPLAN-P2 objective lens, (ii) a D1100 DMD (DLP), (iii) a AC254-100-A-ML doublet (Thorlabs) as the imaging lens, (iv) a custom double Amici prism (Shanghai Optics), (v) a monochrome charge-coupled device FPA detector (Stingray F-033B) and (vi) a set of band-pass filters (Edmund Optics), mounted in a motorized filter wheel (Thorlabs, FW103). A synchronization interface should be designed to control the rapid varying DMD and filter wheel which will modulate the incoming scene in a single FPA integration period.



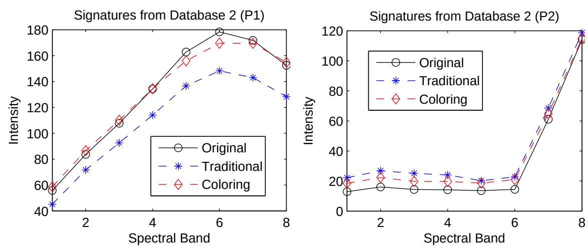
(a) Database 1. (First row) CASSI, (Second row) C-CASSI with $V = 2$, (Third row) C-CASSI with $V = 8$

(b) Database 2. (First row) CASSI, (Second row) C-CASSI with $V = 2$, (Third row) C-CASSI with $V = 8$

Figure 2.11: Impact of the number of snapshots in the reconstructions, fixing $R = 0.3$ and varying $K = 1, 4, 8$, from left to right. (a) Database 1, (b) Database 2.



(a) Database 1. (Left) P1, (Right) P2



(b) Database 2. (Left) P1, (Right) P2

Figure 2.12: Spectral signatures from P1 and P2 in Fig. 2.5, with $R = 0.3$, $K = 3$ and $V = 4$. (a) Database 1. (b) Database 2.

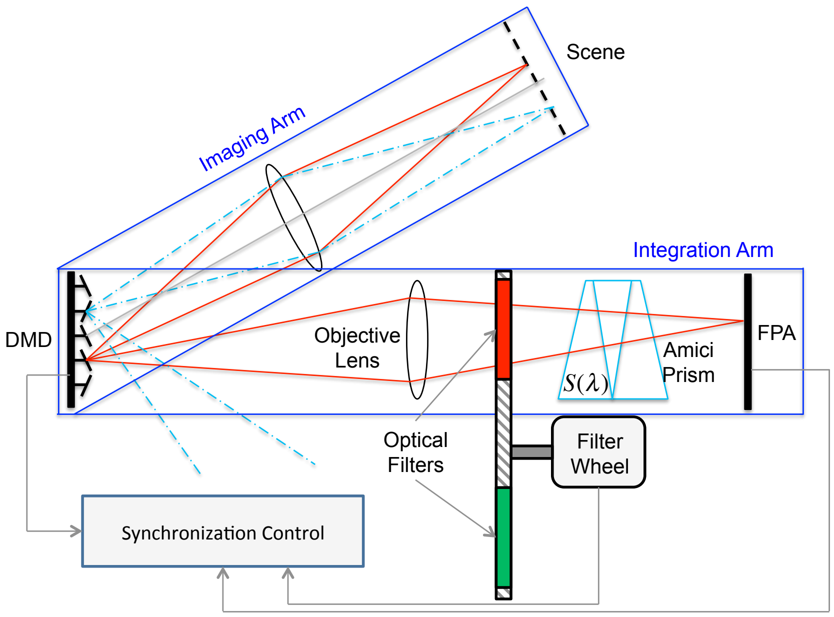


Figure 2.13: Optical scheme of the DMD/Filter-wheel-based C-CASSI. The first lens focuses the scene onto the DMD image plane, which spatially modulates the scene, and a second objective lens focus the coded scene into the image plane of the FPA, which has been first filtered by the rotation wheel, and sheared by the prism. The DMD and filter wheel change rates are faster than the integration time from the FPA.

Based on the simulation results, the CCAs are set to exhibit 128×128 pixels, with each entry being either a $\mathcal{L}$ or a $\mathcal{H}$ filter, with $\lambda_c = 550$ nm (CCAs with $V = 2$). Additionally, the CCAs entries are spatially distributed by means of a random distribution as stated in the simulations. Given that the Amici prism disperses the light up to 8 different FPA columns, the reconstruction algorithm will distinguish up to $L = 8$ spectral bands: {448 – 460}, {461 - 475}, {476 - 494}, {495 - 516}, {517 - 540}, {541 - 572}, {573 - 615} and {616 - 662} nm. Notice that the widths of the intervals are not constant, as a non-linear dispersive element is used. Notice also that, the maximum number of snapshots allowed to capture is $K = 8$ ($K \leq L$), thus, the experiments focused on the evaluation of $K = 1$ and $K = 4$ sensing processes.

The CCA synthesis realized through the use of the DMD and the set of band pass filters, is detailed as follows. (i) Each CCA is mapped to a set of block-unblock CA paired with the corresponding band pass filters given by Eq. (2.3). (ii) The V band-pass filters and block-unblock CA pairs are calibrated to determine the experimental impulse

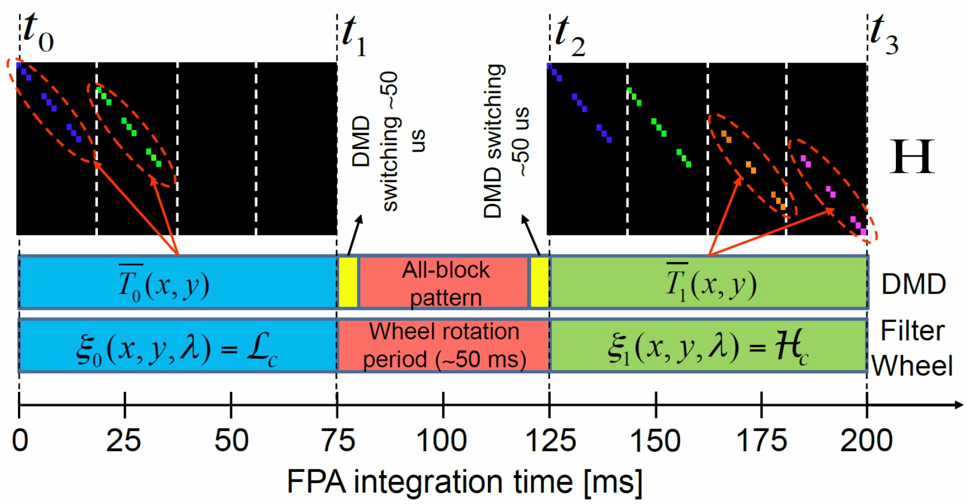
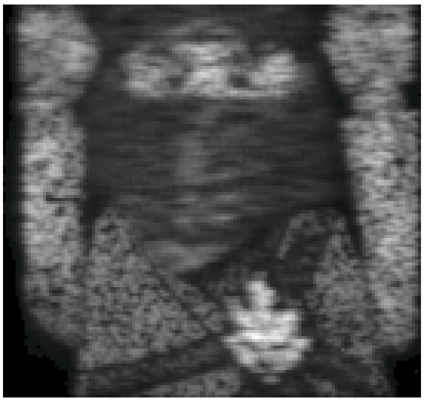


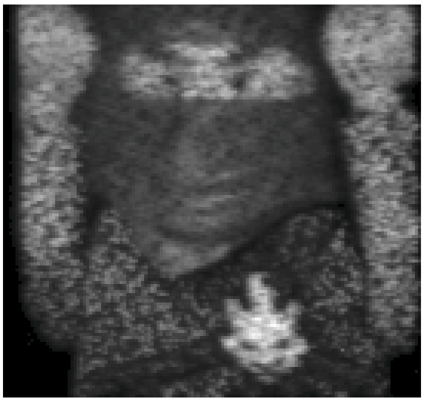
Figure 2.14: Snapshot synchronization control interface tasks using $V = 2$ filters. At t_0, the first DMD pattern/filter pair is set, and it is being measured between t_0 and t_1. At t_1, the control changes the DMD pattern, and rotates the filter wheel to the second filter. The DMD sets an all-black pattern while the filter wheel ends its rotation, to reduce distortion. At t_2 the second DMD pattern/filter pair is set, and it is measured during $t_3 - t_2$ ms. At t_3 the FPA ends its capturing process, and both the DMD and the filter wheel restarts.

response of the system, using monochromatic light at the given Amici prism wavelengths as the input, and a white plate as the target. A calibration data cube is then obtained per each CCA. (iii) The compressive measurements are attained by replacing the white plate by the real target, and using a broadband white light as the illumination source. Notice that for the compressive measurements, the DMD switching time is $\sim 50\mu s$, and the filter wheel rotation time to the adjacent position is $\sim 50ms$, while the FPA integration time is fixed to be $200ms$. In our case, the rotation wheel has $V = 2$ filters synchronized with 2 block-unblock DMD patterns, which rotates/tilts in a 200 ms period. Figure 2.14 details the synchronization control tasks performed in every compressive measurement, for an $N = L = 4$ datacube.

Figure 2.15 presents the $K = 1$ compressive measurements obtained with the traditional CASSI and C-CASSI, respectively. Remark that, both snapshot processes integrate the same amount of intensity restricted by the transmittance of the CA, which was fixed to be $R = 0.3$ (the best transmittance overall obtained from the simulations). There, it can be noticed that the traditional measurement present certain clustered pixels whereas the



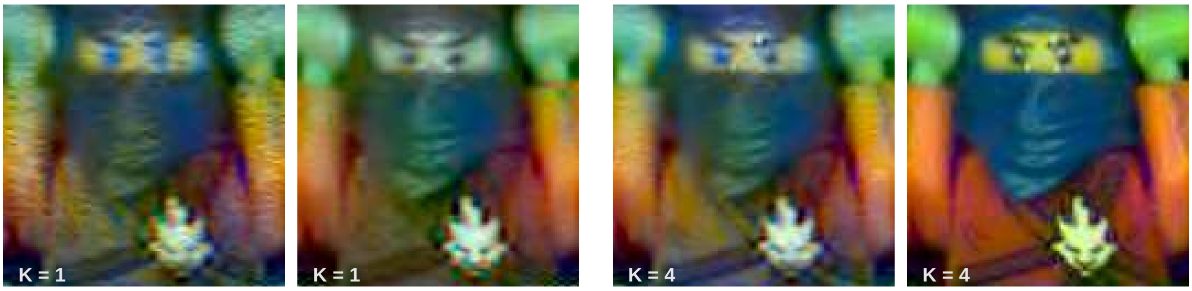
(a) CASSI



(b) C-CASSI

Figure 2.15: Comparison of the compressive measurements using (a) CASSI (b) C-CASSI. In CASSI the intensity appears more clustered in neighboring pixels, rather than in C-CASSI where the intensity is more uniformly distributed.

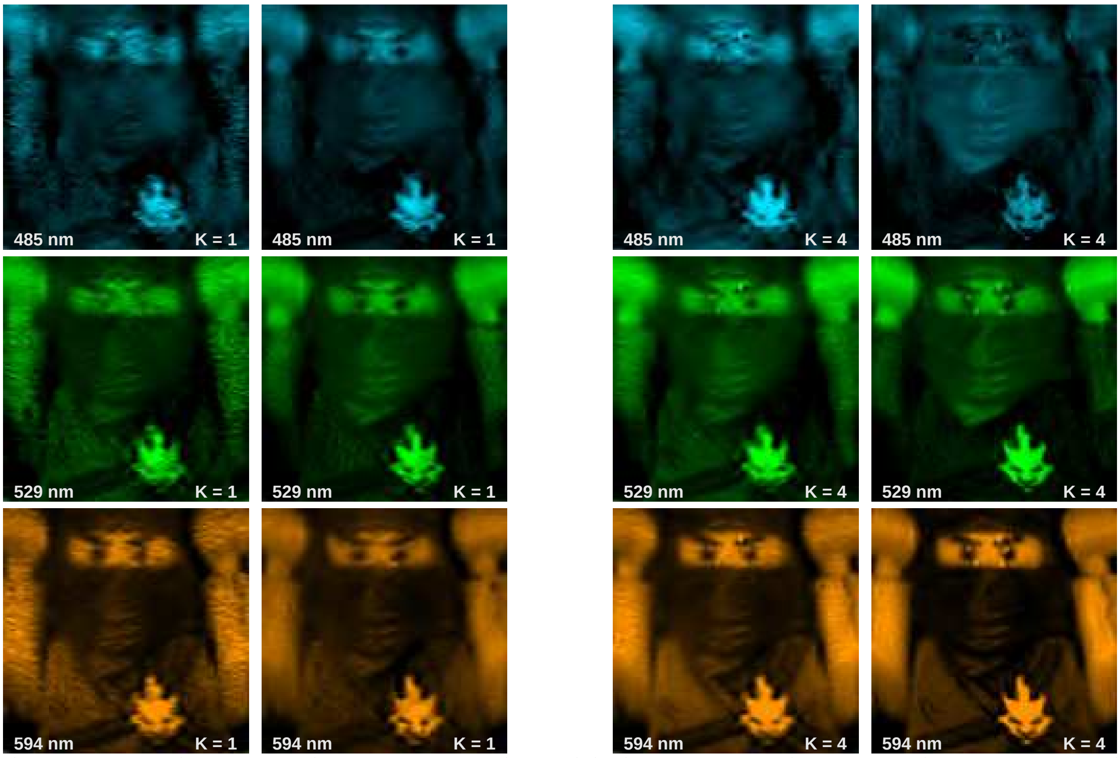
C-CASSI measurement present a more uniform distribution of the intensity, which goes along with the entailed richness of the CCAs. An estimation of the spatio-spectral datacube is recovered by solving Eq. (4.27) using the GPSR algorithm. Figure 2.16 presents the reconstructed data cubes obtained using the traditional CASSI and C-CASSI (with $V = 2$), with $K = 1$ and $K = 4$, respectively, and mapped to an RGB profile. It can be noticed that for a single snapshot process ($K = 1$) the reconstruction quality is not good enough and additional snapshots are desirable. The C-CASSI outperforms the results attained with the traditional CASSI, even for small K. In addition, the gain achieved with C-CASSI can be observed in Fig. 2.17, where three of the eight monochromatic reconstructed bands are depicted. The corresponding band/color discrimination can be contrasted with the RGB reconstructions presented in Fig. 2.16. For instance, the blue helmet is well characterized by the 485 nm wavelength when CCAs are employed with $K = 4$. On the other hand, Fig. 2.18 depicts the spectral signatures of 2 different points from the target scene, compared against a reference measured using a commercially available spectrometer (Ocean Optics USB2000+). The fit of the spectral signatures obtained with C-CASSI overcomes the ones from the traditional CASSI. The improvement in spatial/spectral quality achieved by the proposed procedure is then clearly noticeable.



(a) (Left) CASSI, (right) C-CASSI

(b) (Left) CASSI, (right) C-CASSI

Figure 2.16: Reconstructed data cubes mapped to an RGB profile for: (a) $K = 1$, and (b) $K = 4$.



(a) (First Column) CASSI, (Second Column) C-CASSI

(b) (First Column) CASSI, (Second Column) C-CASSI

Figure 2.17: Experimental reconstructions of the 3rd, 5th and 7th spectral bands. (a) $K = 1$ (b) $K = 4$.

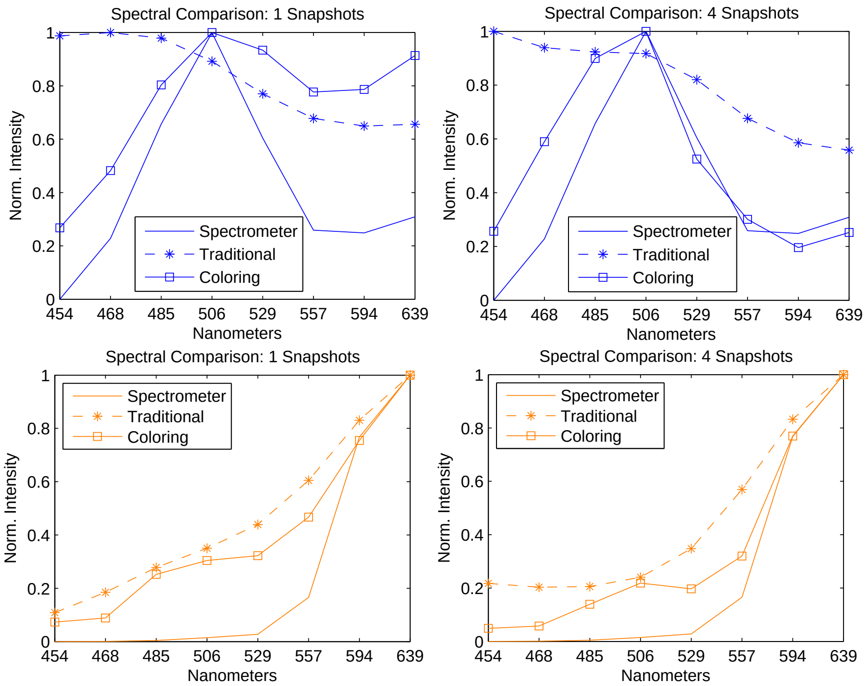


Figure 2.18: Spectral signatures of 2 different points using $K = 1$ and $K = 4$, respectively. (Top row) Helmet blue point, (Bottom row) Arm orange point.

2.5 Conclusions

A proof-of-concept C-CASSI system has been implemented through a DMD and traditional filters. The DMD-based implementation rises as an important way to alleviate the costly implementation of the CCAs. The use of CCAs entails richer coding strategies as not only spatial but also spectral coding is now performed in a single step, thus leading to less ill-posed problems which in turn provide better spectral image reconstructions.

Chapter 3

COMPRESSIVE SPECTRAL TESTBED IMAGING SYSTEM BASED ON THIN-FILM COLOR-PATTERNED FILTER ARRAYS

Compressive spectral imaging systems have been demonstrated to be reliable in capturing multispectral data using far fewer measurements than traditional scanning techniques. In this chapter, a thin-film patterned filter array-based compressive spectral imager is demonstrated, including its optical design and implementation. The use of a patterned filter array entails a single-step 3D spatial-spectral coding on the input data cube, thus providing higher flexibility on the selection of voxels being multiplexed on the sensor. The patterned filter array is designed and fabricated at micrometer pitch size thin-films, referred as pixelated filters, with three different cut-off wavelengths. The performance of the system is evaluated in terms of references measured by a commercially available spectrometer and the visual quality of the reconstructed images. Different distributions of the pixelated filters, including random and optimized structures are explored.

3.1 Introduction

Compressive spectral imaging (CSI) techniques capture multiplexed-and-coded projections of a scene, and then, the 3D underlying datacube is estimated by exploiting the fact that spectral images are highly correlated and admit sparse representations [30]. Different CSI architectures have been proposed to date [35, 23, 44, 45]. All of these utilize block-unblock binary coded apertures and one or two dispersive elements to modulate the optical field from the scene. These coded apertures have been fabricated using materials such as chrome-on-quartz, rendering coded aperture elements which are either opaque or translucent to the whole wavelengths of interest, as illustrated in Fig. 3.1(a). These coded apertures are referred to as photomasks. Recent coating technologies have allowed the design of patterned arrays of pixelated optical filters [37, 46]. Since each pixelated filter attains spatial and spectral coding in a single step, as presented in

Fig. 3.1(b), coding elements based on patterned optical coatings lead to more efficient 3D coding strategies and more compact imaging systems.

There exist various fabrication approaches to attain pixelated optical filters: color-dye gels [47, 48], Fabry-Perot micro-sructures [38, 49], and thin-films [50, 51]. Dye gels have been widely used on CCD cameras due to their low cost, while Fabry-Perot and thin-films technology have emerged with much higher filter precision but at a higher cost, initially, as compared to dye-gels. The principal difference between these technologies, apart from its cost and fabrication procedures, is the extinction ratio. Dye-gels yield broad transition bands, whereas Fabry-Perot and thin films provide a sharper transition, which in turn, enable purer color filtering.

Recently, we have proposed and numerically demonstrated the use of patterned filter arrays, also referred as colored coded apertures, in compressive spectral imaging systems [41, 52]. In this chapter, we report on the development and fabrication of a thin-film-based patterned filter array and its implementation on a CSI testbed architecture. Moreover, an analysis on the impact of the distribution of the pixelated filters on the quality of the reconstructed 3D spectral images is presented, comparing random against optimized distributions. To evaluate the quality of the reconstructed images, the results are presented in terms of monochromatic spectral images, and also mapped to RGB profiles for spatial resolution and visual inspection. To evaluate the spectral fidelity, spectral profiles of selected representative points in the scenes are compared against references measured by a spectrometer. The attained results show improved quality on the images obtained by the optimized distribution of the pixelated filters over the ones with random distributions.

3.2 Compressive optical sensing image model

We focus on the single-disperser coded aperture snapshot spectral imager (SD-CASSI), which typically uses a photomask to encode the input data. A dispersive element is then used to decompose the coded spectrum and a monochrome sensor captures the multiplexed data [30, 23]. In contrast, a 3D patterned filter array is used as the coding element in the proposed optical system, replacing the commonly used photomask as

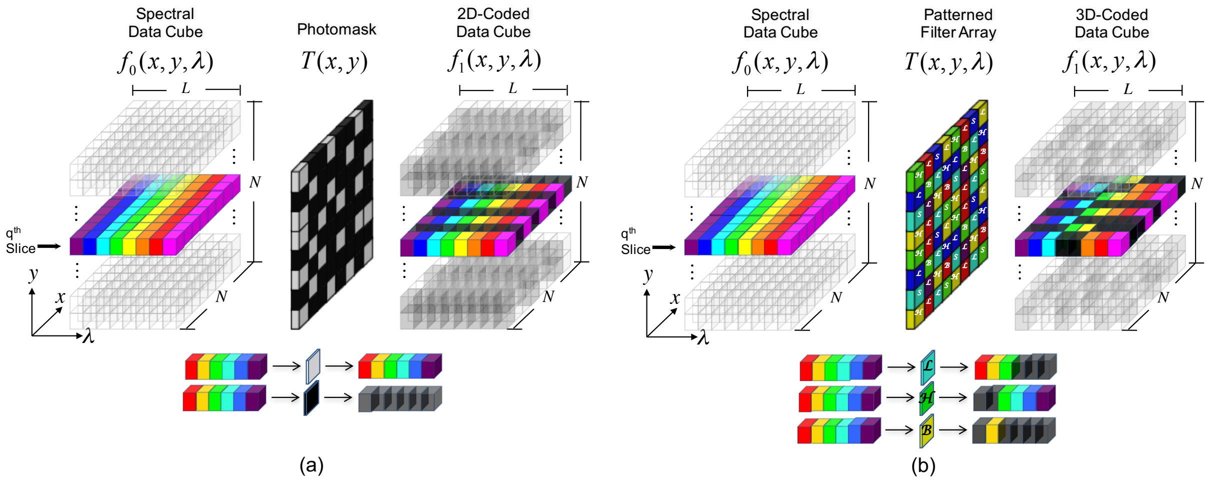


Figure 3.1: Comparison of the coding strategies. (a) Photomask coding. (b) Patterned filter array coding. The filter array enables 3D coding due to the spectral response of the pixelated filters, whereas the photomask performs 2D coding over the data cube.

depicted in Fig. 3.1. The 3D coding is enabled by the 2D array of micro-optical pixelated filters, each of which may be a low-pass, high-pass, band-pass, stop-band or dichroic filter; the dispersive element can be a reflective element, such as a grating, or a transmissive element, such as a prism; the detection element can be a grayscale 2D detector, a line-detector, or even, a single-pixel detector. In particular, the optical system described hereafter uses a custom double Amici prism and a 2D grayscale focal plane array (FPA).

The compressive coded projections of the spectral components of the scene are captured as follows. Denote the continuous spatio-spectral scene as $f_0(x, y, \lambda)$, the patterned filter transfer function as $T(x, y, \lambda)$, and the Amici prism dispersion function as $S(\lambda)$. Coding is realized at the image plane of the field focused by an imaging lens, thus creating the 3D-coded field $f_1(x, y, \lambda) = T(x, y, \lambda)f_0(x, y, \lambda)$. After coding, a relay lens, transmits the coded scene through the Amici prism, dispersing the field along the horizontal axis, forming the image on the FPA detector. The integrated coded and sheared spectral field along the spectral range sensitivity of the detector (Λ) can be expressed as,

$$g(x, y) = \int_{\Lambda} T(x, y, \lambda)f_0(x, y, \lambda)h(x - x' - S(\lambda), y - y')d\lambda, \tag{3.1}$$

where $h(x - x' - S(\lambda), y - y')$ accounts for the optical impulse response of the system and the shifting entailed by the prism along the x-axis. The coding transfer function of the patterned filter array can be modeled as,

$$T(x, y, \lambda) = \sum_{i,j,k} T_{i,j,k} \text{rect}\left(\frac{x}{\Delta_c} - i - \frac{1}{2}, \frac{y}{\Delta_c} - j - \frac{1}{2}, \frac{\lambda}{\Delta_d} - k - \frac{1}{2}\right) \tag{3.2}$$

where $T_{i,j,k} \in [0, 1]$ represents the filtering operation to be performed on the $(i, j, k)^{th}$ data cube voxel; $i, j \in \{0, \ldots, N-1\}$, and $k \in \{0, \ldots, L-1\}$ index the coordinates of an $N \times N \times L$ spectral data cube, and Δ_c, Δ_d account for the pixel sizes of the pixelated filters and the FPA detector, respectively. Notice that the spatial resolution of the resolvable scene is determined by the pitch size of the pixelated filters (Δ_c), while the spectral resolution depends on the pitch size of the FPA pixels (Δ_d) and the dispersion efficiency of the prism $S(\lambda)$ [41].

Let each voxel of the spectral scene be expressed as $F_{i,j,k} = \iiint_{\Omega_{i,j,k}} f_0(x, y, \lambda) dx dy d\lambda = w_{ijk} f_0(x_i, y_j, \lambda_k)$, where $\Omega_{i,j,k}$ represents the $(i, j, k)^{th}$ voxel boundaries, and $w_{i,j,k}$ the voxel mass center weight. Assuming that the response of the dispersive element is linear over the spectral range of the system, and that ideal optical elements are used, the measurement attained on the $(n, m)^{th}$ pixel of the proposed system is given in discrete form by,

$$g_{nm} = \sum_k F_{n,m-k,k} T_{n,m-k,k}, \tag{3.3}$$

where $n \in \{0, \ldots, N-1\}$ and $m \in \{0, \ldots, N+L-1\}$, index the pixels on the rows and columns of the detector, respectively.

3.2.1 Multiple-snapshot matrix model

A single snapshot may not be enough to attain a certain required quality in the reconstructed data cubes. The proposed system is then extended to include multiple snapshots. Multiple-snapshot sensing leads to less ill-posed inverse problems and consequently improved signal recovering [30]. In such a case, the patterned filter array must change its coding pattern with every new snapshot. As will be described later, we use a wafer with numerous patterned filter arrays, contiguously placed to each other, so that, a new coding pattern is attained by mechanically shifting the wafer. Multiple-snapshot sensing can be modeled by $g^{\ell}_{n,m} = \sum_k F_{n,m-k,k} T^{\ell}_{n,m-k,k}$, where $\ell = 0, \ldots, K-1$ accounts for K snapshots, attained by K spatial translations of the patterned filter array wafer.

Denote the k^{th} spectral band of the input spectral data cube as $\mathbf{f}_k \in \mathbb{R}^{N^2}$, and let each ℓ^{th} patterned filter array in vectorial form as $\mathbf{t}^{\ell}_k \in \mathbb{R}^{N^2}$, such that $(\mathbf{f}_k)_z = F_{(z-\lfloor \frac{z}{N} \rfloor N), \lfloor \frac{z}{N} \rfloor, k}$, and $\left(\mathbf{t}^{\ell}_k\right)_z = T^{\ell}_{(z-\lfloor \frac{z}{N} \rfloor N), \lfloor \frac{z}{N} \rfloor, k}$, respectively, for $z = 0, 1, \ldots, N^2$. Also, let $\mathbf{g}^{\ell}$ be a column vector holding all of the recorded pixel values, $g^{\ell}_{n,m}$. Given the lateral dispersion of light, the length of each compressive measurement $\mathbf{g}^{\ell}$ is $N(N+L-1)$ such that all light is accounted for on the sensor. Thus, while the patterned filter array is an $N \times N$ pixel array, the sensor has an extra set of L pixel columns. Assuming that $\mathbf{f}$ stays constant over the K snapshots, we can relate $\mathbf{g}^{\ell}$ to $\mathbf{f}$ in matrix form as

$$\mathbf{g}^{\ell} = \mathbf{H}^{\ell}\mathbf{f} = \mathbf{P}\mathbf{T}^{\ell}\mathbf{f} \tag{3.4}$$

where $\mathbf{f} = [\mathbf{f}_0^T \mathbf{f}_1^T \ldots \mathbf{f}_{L-1}^T]^T$, and $\mathbf{H}^{\ell}$ is the system transfer function of the ℓ^{th} snapshot, which accounts for the coding pattern $\mathbf{T}^{\ell}$ and the dispersion function of the prism $\mathbf{P}$, which remain constant for all the snapshots. Let $Q = N^2L$ be the total number of data cube voxels, and $U = N(N+L-1)$ be the total number of FPA pixels. Therefore, $\mathbf{H}^{\ell} \in \mathbb{R}^{U \times Q}$, $\mathbf{P}$ is an $U \times Q$ rectangular matrix accounting for the shearing of the prism and the integration performed by the FPA. The structure of the matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \begin{bmatrix} \mathrm{diag}(\mathbf{1}_{N^2}) & \mathbf{0}_{N \times N^2} & \cdots & \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{N \times N^2} & \mathrm{diag}(\mathbf{1}_{N^2}) & \cdots & \mathbf{0}_{N \times N^2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N \times N^2} & \mathbf{0}_{N \times N^2} & \cdots & \mathrm{diag}(\mathbf{1}_{N^2}) \end{bmatrix}_{U \times Q}, \tag{3.5}$$

where $\mathbf{0}_{N \times N^2}$ is a 0-valued $N \times N^2$ matrix, and $\mathrm{diag}(\mathbf{1}_{N^2})$ is a 1-valued $N^2 \times N^2$ diagonal matrix. Notice that, the number of $\mathbf{0}_{N \times N^2}$ sub-matrices in each row or column of Eq. (3.5) is $L-1$, and there is just one $\mathrm{diag}(\mathbf{1}_{N^2})$ sub-matrix per row or column. Similarly, $\mathbf{T}^{\ell}$ is a $Q \times Q$ block-diagonal matrix, where each $N^2 \times N^2$ diagonal sub-matrix accounts for the effect of the ℓ^{th} patterned filter array on the k^{th} data cube spectral band, as

$$\mathbf{T}^{\ell} = \begin{bmatrix} \mathrm{diag}(\mathbf{t}_0^{\ell}) & \mathbf{0}_{N^2\times N^2} & \cdots & \mathbf{0}_{N^2\times N^2} \\ \mathbf{0}_{N^2\times N^2} & \mathrm{diag}(\mathbf{t}_1^{\ell}) & \cdots & \mathbf{0}_{N^2\times N^2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times N^2} & \cdots & \mathrm{diag}(\mathbf{t}_{L-1}^{\ell}) \end{bmatrix}_{Q\times Q}, \tag{3.6}$$

where $\mathrm{diag}(\mathbf{t}_k^{\ell})$ is an $N \times N$ diagonal matrix, with diagonal elements $\mathbf{t}_k^{\ell}$.

The complete set of K snapshots in (3.4) can be assembled into a single vector by concatenating each $\mathbf{g}^{\ell}$ vector, end to end, to create the $KU \times 1$ vector $\mathbf{g} = \left[(\mathbf{g}^0)^T, \ldots, (\mathbf{g}^{K-1})^T\right]^T$, such that $\mathbf{g} = \mathbf{Hf}$, where $\mathbf{H} = \left[(\mathbf{H}^0)^T, \ldots, (\mathbf{H}^{K-1})^T\right]^T$ is a $KU \times Q$ matrix. Figure 3.2 shows an sketch of the $\mathbf{H}$ matrix for an $N \times N = 4 \times 4$ spectral data cube with L bands, using $K = 2$ snapshots, and a patterned filter array with L optical filters. The colored diagonals correspond to the filter array applied to each k^{th} waveband, while the black entries correspond to blocking elements. The yellow circles in the figure highlight the variation of the coding elements along the spectral bands, satisfying the wavelength dependency of the filters within the coded pattern. Note that the 2 snapshots are vertically stacked, each one with size U, and the dispersion function is modeled by off-setting the diagonal structure of the patterned filters as the wavelength increases from left to right.

3.2.2 Input data cube reconstruction

Compressive sensing (CS) states that the sensing process can be alternatively expressed as $\mathbf{g} = \mathbf{Hf} = \mathbf{H\Psi}\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\theta}$, such that $\mathbf{A}$ is called the compressive sensing matrix, and $\boldsymbol{\theta} = \mathbf{\Psi}^T\mathbf{f}$ is a vector coordinate of $\mathbf{f}$ in a Q dimensional space $\mathbf{\Psi}$. This transformation assumes $\mathbf{f}$ to be S-sparse on $\mathbf{\Psi}$, so that, only a small subset, $S \ll Q$, of the basis vectors $\mathbf{\Psi}$ can accurately reconstruct $\mathbf{f}$ with little or no distortion. Formally, $\mathbf{f}$ is S-sparse or has sparsity S in a basis $\mathbf{\Psi}$ if $\|\boldsymbol{\theta}\|_0 = S$, where $\|\cdot\|_0$ denotes the ℓ_0-norm, which simply counts the number of nonzero entries in the vector. In this way, an estimation of the

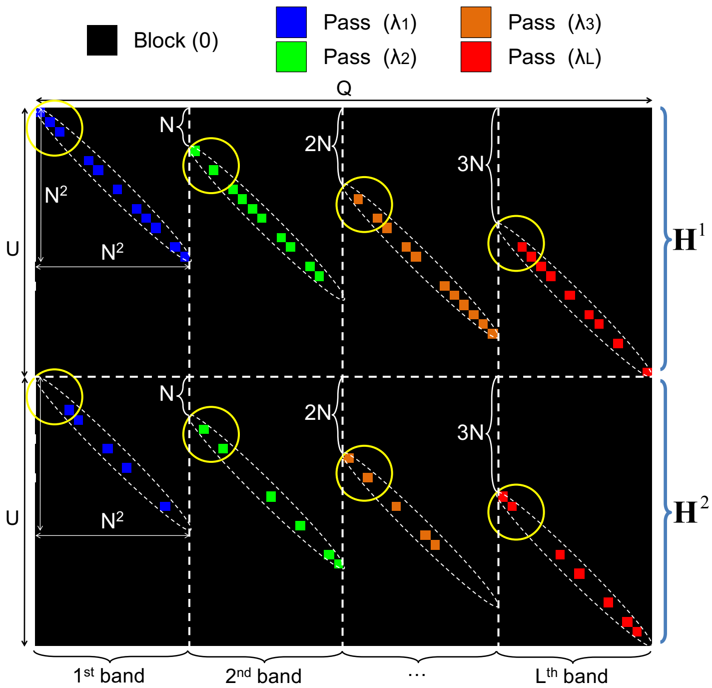


Figure 3.2: Sketch of transfer function matrix $\mathbf{H}$ for a $4\times4\times L$ spectral data cube, using $K = 2$ snapshots, and a patterned filter array with L filters. $Q = N^2L$ is the total number of data cube voxels, and $U = N(N+L-1)$ is the total number of FPA pixels. Note that the coding pattern is wavelength-dependent, and a different distribution of its entries is used in each snapshot.

spatio-spectral input data cube can be attained by solving the regularization problem,

$$\hat{\mathbf{f}} = \mathbf{\Psi}\left(\underset{\boldsymbol{\theta}'}{\operatorname{argmin}}\|\mathbf{g}-\mathbf{A}\boldsymbol{\theta}'\|_2^2+\tau\|\boldsymbol{\theta}'\|_1\right), \tag{3.7}$$

where $\boldsymbol{\theta}'$ is the estimation of the vector coordinate of $\hat{\mathbf{f}}$ in $\mathbf{\Psi}$, $\|\cdot\|_2^2$ represents the ℓ_2 norm, $\|\cdot\|_1$ the ℓ_1 norm, and τ is a regularization parameter which penalizes the fact of $\boldsymbol{\theta}'$ being sparse, while reducing the error between the estimated $\mathbf{A}\boldsymbol{\theta}'$, and the captured set of compressive measurements $\mathbf{g}$. To solve the optimization problem in (7.16), different algorithms have been proposed, including the gradient projection for sparse reconstruction (GPSR) [29], the GPSR-based but faster block-processing approach proposed in [43], the two-step iterative shrinkage/thresholding (TwIST) using the total-variation (TV) as the regularization function [53], gaussian mixture models (GMM) [54] or hyperspectral denosing algorithms based on approximate message passing (AMP) [33]. This chapter uses the method in [29] although any of the other methods could be used as well.

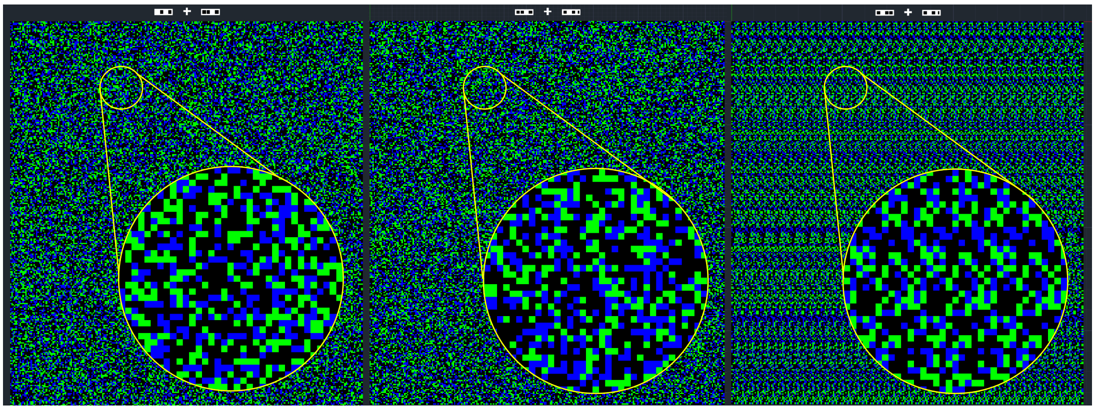
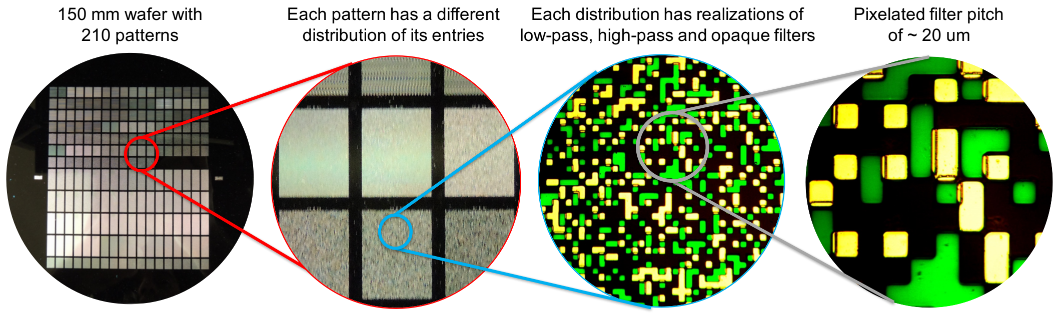


Figure 3.3: 3-color patterned filter array design using 3 different spatial distributions. The blue pixels are low-pass filters, the green are high-pass filters, and the black are all-stop filters.

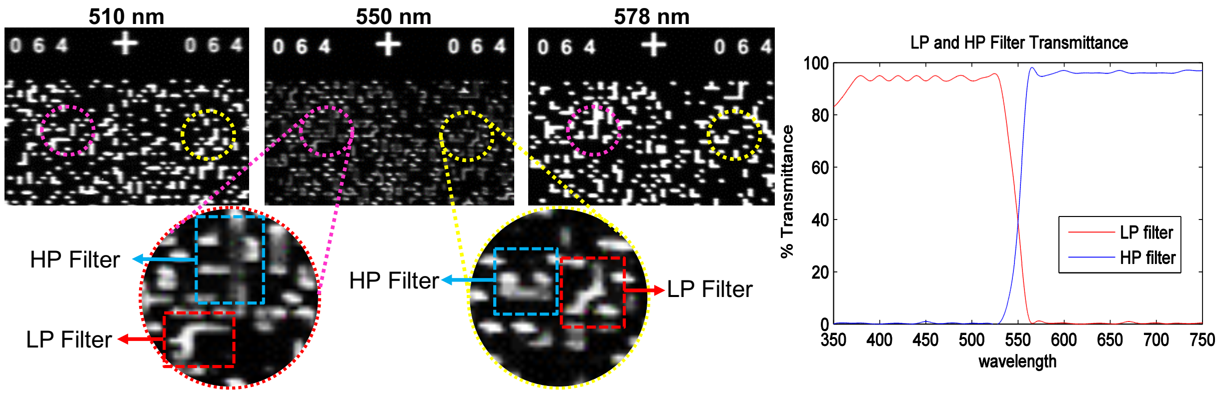
3.3 Design and fabrication of a patterned filter array wafer

3.3.1 Patterned filter array design

The number of different filters to be used, the filters spectral profile, and the way these filters are spatially distributed have to be carefully designed. First, if the filter array has $N \times N$ pixels of spatial resolution, then, it is possible to have at most a different optical filter per pixel, that is, up to N^2 pixelated filters with different cut-off wavelength. The higher the number of different filters, however, the higher the fabrication cost. Furthermore, we have observed in practice that the gains gradually decrease as the number of filters increase [52]. In fact, only a few different type of optical filters are typically needed to obtain significant improvement over imaging that uses conventional photomask technology. As an example [52], only 3 filters are sufficient to improve image reconstruction quality by 4 dBs, of an 8 spectral band data cube, in terms of peak-signal-to-noise-ratio (PSNR). It is expected, however, that the number of filters used, for improved performance, should be larger with increases of the spectral bands. Based on these results, and taking into account the fabrication costs, we used 3 filters: a low-pass (denoted by $\mathcal{L}(\lambda)$) and a high-pass (denoted by $\mathcal{H}(\lambda)$) complementary filters, with cut-off wavelength $\lambda_c = 550nm$, and a blocking (all-stop, denoted by $\mathcal{B}(\lambda)$) filter, represented in Fig. 3.3 with colors blue, green and black, respectively.



(a) Zoomed versions of the wafer are shown from left to right.



(b) The HP and LP filters appear in the transition band of 550 nm.

Figure 3.4: Fabricated wafer of thin-film patterned filter arrays. (a) A microscope was used to capture zoomed versions of the wafer shown from left to right. (b) Analysis of the transition band of the filters. Due to the cut-off wavelength of the filters (λ_c = 550 nm), both HP and LP filters show-up attenuated when the patterned filter is illuminated at that wavelength.

Second, the pixelated filters, or coded aperture entries in previous photomasks implementations, have been usually placed in an $N \times N$ grid using simply random spatial distributions. However, we have shown in previous publications that the spatial distribution of the entries in a coded aperture plays an important role in the quality of the reconstructions [30, 55, 36]. Therefore, in addition to the random distribution of the filters, in this chapter, we also test and analyze the performance of another two distributions called Boolean, and optimized [41] distributions, as explained below.

3.3.1.1 Random distributed filters

Random distributions of block-unblock filters have been widely used in CASSI. These are justified if high incoherence is attained with the signal representation basis $\boldsymbol{\Psi}$

such as Fourier, Wavelet or Cosine [19]. The entries of a coded aperture with random distributions can be modeled as,

$$T_{i,j}^{\ell} = \begin{cases} \mathcal{B}(\lambda), & p \leq \frac{1}{3} \\ \mathcal{L}(\lambda), & \frac{1}{3} < p \leq \frac{2}{3} \\ \mathcal{H}(\lambda), & p > \frac{2}{3} \end{cases}, \tag{3.8}$$

where p is a random variable with distribution function $U[0,1]$. Notice that $T_{i,j}^{\ell_1}$ and $T_{i,j}^{\ell_2}$ ($\ell_1 \neq \ell_2$) are independently generated from each other.

3.3.1.2 Boolean distributed filters

Boolean distributions keep the properties of random distributions for a single snapshot. However, they have the characteristic that in multiple-shot sensing, each realization is mutually-complementary to the previous realizations, thus guaranteeing that each voxel of the datacube is captured only once by the FPA. Note that, such characteristic is not guaranteed by random distributions, where each realization is generated independently of the others; therefore, a single voxel could be sensed more than once, or even worse, a single voxel may be always filtered-out. Boolean distributions can be modeled as in (3.8), but satisfying the mutually-complementary constraint $\sum_{\ell=0}^{K-1} T_{i,j}^{\ell} = A(\lambda)$, where $A(\lambda)$ is an all-pass filter, representing the full spectrum of interest. For instance, in a $K = 3$ snapshot sensing, if the $(i,j)^{th}$ filter in the first snapshot is a low-pass ($T_{i,j}^{0} = \mathcal{L}(\lambda)$), then to satisfy the constraint, the second snapshot must use a high pass filter and the third snapshot a blocking filter, or vice versa ($T_{i,j}^{1} = \mathcal{H}(\lambda)$ and $T_{i,j}^{2} = \mathcal{B}(\lambda)$; or $T_{i,j}^{1} = \mathcal{B}(\lambda)$ and $T_{i,j}^{2} = \mathcal{H}(\lambda)$).

3.3.1.3 Optimized spatial distribution of the filters

We recently proposed and demonstrated an optimization of the distribution of the entries of a patterned filter array [41]. This optimization extracts the benefits of Boolean distributions, but in addition, it exploits the correlations created by the structure of the CS matrix $\mathbf{A}$, such that the Restricted Isometry Property (RIP) is better satisfied. In

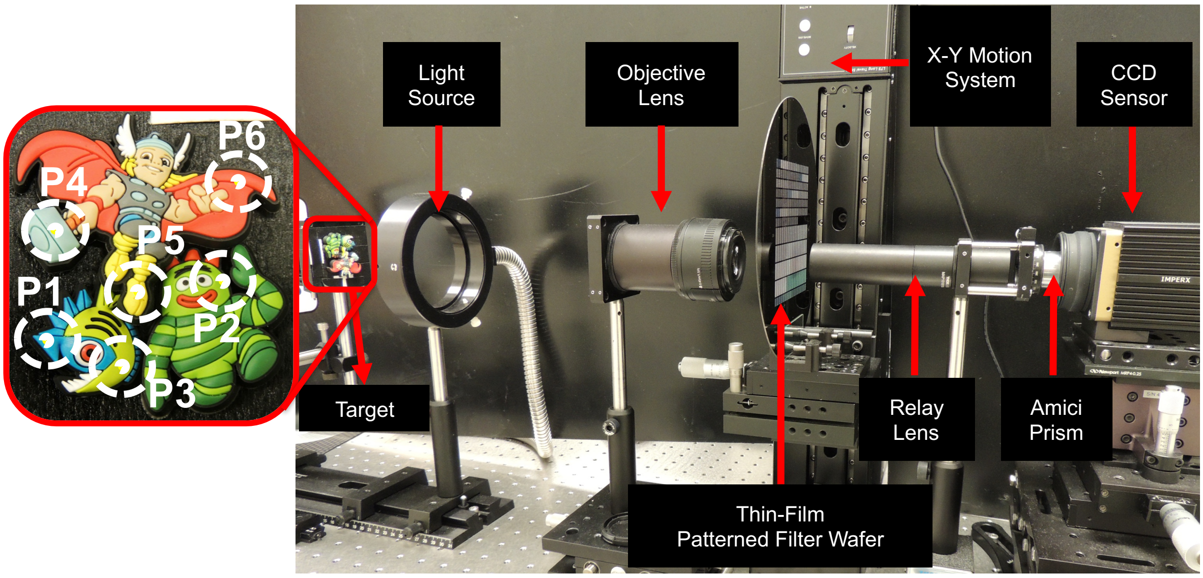


Figure 3.5: Thin-film-based compressive spectral imaging testbed. The objective lens images the captured scene onto the thin-film patterned filter wafer; then the relay lens transmit the coded light through the Amici prism before it is integrated by the CCD sensor. The X-Y motion system moves the wafer precisely to access different coding patterns.

particular, the optimization of the distribution of the entries can be expressed as,

$$\begin{aligned} \underset{T^{\ell}_{i,0},T^{\ell}_{i,1}\cdots,T^{\ell}_{i,L-1}}{\operatorname{argmin}} \quad & \sum_{k=1}^{L} c_k \gamma_k + \tau_1 U \\ \text{subject to} \quad & \sum_{\ell=0}^{K-1} T^{\ell}_{i,j} = A(\lambda), \end{aligned} \tag{3.9}$$

for $i = 1, \ldots, N$, where, c_k are the entries of a penalty cost vector, τ_1 is a regularization constant, γ_k forces to reduce the correlations along the spectral domain, U attempts to reduce the correlations along the spatial domain, and the constraint guarantees the mutual-complementariness along the K multiple-snapshots. Intuitively, the optimization problem seeks to spread as far as possible the passing bands of the pixelated filters along the spectral and the spatial domains, exploiting the correlations entailed by the sensing structure of the optical system. For further detailed information, the reader is referred to [41].

Figure 3.3 shows a single realization of the 3 spatial distributions as seen in the final design of the wafer, developed using AutoCAD. Notice, that the random and the Boolean realizations present a pure random distribution of the entries, whereas the optimized realization exhibit a pseudo-random structure since just a single L-long section is designed per row and then horizontally replicated.

3.3.2 Thin-film-based patterned filter array fabrication

The fabrication process was realized by Pixelteq Inc., and it mainly consisted on deposition and lithographic patterning cycles of the 3 filters, in a random, Boolean or optimized fashion [37]. The low-pass filter has an average transmittance of 93% between 350-540 nm, and of 0.25% between 560-750 nm; and the high-pass filter has an average transmittance of 0.04% between 350-540 nm, and of 96% between 560-750 nm, as shown in Fig. 3.4(b). The blocking pixel has a 0.04% transmittance in the entire domain between 350-750 nm. The array was fabricated on a 150 mm diameter wafer of Eagle XG substrate, with 1.1 mm thickness . Each patterned filter array in the wafer was designed to exhibit $N \times N = 256 \times 256$ pixels of spatial resolution, where each pixelated filter has a pixel-pitch size of $\Delta_c = 20\mu m$. Due to the micrometer size of the pixelated filters, they are created by multilayer stacks of high- and low-index materials via plasma-assisted deposition or magnetron sputtering [37].

The final product is shown in Fig. 3.4(a). It contains 210 patterned filter arrays, and zoomed versions show the fine details within the full wafer. The 210 patterned arrays are different realizations of the spatial distributions, for different number of snapshots. Note that the pixelated filters exhibit sloped edges, since the lift off fabrication process requires the lithography to be removed and the coating in the pixel to remain. These sloped edges cause areas of non-uniformity in the transmission properties of the coatings. In consequence, fine striped chrome areas are used around the pixelated filters, allowing to aperture the pixel, so that, only the uniform portion is projected. If these opaque areas are not used, the effective pixel performance will suffer. In order to analyze the transition-band of the pixelated filters, Fig. 3.4(b) shows three images of a portion from a single patterned filter array illuminated with three different wavelengths (510 nm, 550 nm, and 578 nm). It can be seen that when the pattern is illuminated with monochromatic light at 510 nm, just the low-pass pixelated filters let the light pass; similarly, when it is illuminated with monochromatic light at 578 nm, just the high-pass pixelated filters let the light to propagate. Note that when the patterned array is illuminated with monochrome light at the cut-off wavelength of the filters ($\lambda_c = 550$ nm), both high- and low-pass filters let light go through the system, but its intensity is attenuated around 40%, agreeing with the given transition characteristics.

3.4 Optical testbed implementation and experiments

To test the performance of the patterned filter array wafer, the CSI testbed setup presented in Fig. 7.5 was assembled in our laboratory. A colorful target object was attached to a black surface and illuminated by a broadband Xenon lamp playing the role of ambient light. The wafer of patterned filter arrays was placed at the image plane of the objective lens Canon EF 50mm f/1.8 II. Two Thorlabs LTS300 linear translation stages with integrated stepper motor controller were used to precisely move the wafer along the x-y axis, in order to capture different patterned filter arrays. Each independent linear stage exhibits a velocity of 50 mm/s, on-axis accuracy of 5 μm, and a bidirectional repeatability of 2 μm. A matched achromatic doublet pair lens with 100 mm focal length was used to relay the 3D coded field through the custom designed double Amici prism, which horizontally spreads a single voxel of the data cube into 42 different columns of the FPA. The FPA sensor is a Bobcat B2021 CCD GigE Vision (with PoE), exhibiting a pixel pitch of $\Delta_d = 7.4\mu m$ and a resolution of 2048×2048. A 3×3 macro-pixel windowing was realized on the FPA, in order to have an one-to-one correspondence with each pixelated filter, that is, $\Delta_d \approx 3\Delta_c$. Therefore, the attainable resolution in the reconstructions is $N \times N = 256 \times 256$ pixels in space, and $L = 14$ spectral bands.

To obtain the estimations of the input data cubes, the gradient projection for sparse reconstruction (GPSR) algorithm [29] was used to solve the optimization problem in (7.16). It is important to notice, however, that there exist different strategies to do this estimation, as mentioned in Section 2.2. GPSR has been broadly used with traditional basis functions $\boldsymbol{\Psi}$ such as the wavelet ($\boldsymbol{\Psi}_{W2D}$) and discrete cosine transforms ($\boldsymbol{\Psi}_{DCT}$) as well as pre-trained dictionaries [30, 43, 54]. However, it is also possible to simultaneously exploit the sparsity properties of the input data cube along each of its dimensions to provide a new representation for their structure. Doing so, a single sparsity transform for the entire data cube is obtained as the Kronecker product of the bases used for each of its dimensions [45, 52, 56]. Particularly, we used the basis expressed as the Kronecker product between $\mathbf{\Psi}_{W2D}$ and $\mathbf{\Psi}_{DCT}$ where $\mathbf{\Psi}_{DCT}$ is used to sparsify the spectral axis while $\mathbf{\Psi}_{W2D}$ sparsifies the spatial coordinates.

A single reconstruction with GPSR takes around 90 seconds to run 300 iterations for $K = 1$, and up to 210 seconds for $K = 8$ snapshots, when the data is processed in Matlab R2015b on an Intel Xeon CPU E5-1660 v3 at 3.00GHz and 80GB of RAM. Figures 3.6–3.11 summarize the reconstruction results. In these figures, reconstruction results for $K = 2$ (compression ratio of 1/7) and $K = 8$ (compression ratio of 4/7) snapshots are shown, for the random, Boolean and optimized spatial filter distributions.

First, in Figs. 3.6 and 3.7 the quality of the spatial reconstructions is evaluated by mapping the set of 14 reconstructed spectral bands to an RGB profile. It can be noticed that the Boolean distribution of the filters attain better quality than the random generated filters, and the optimized reconstructions exhibit the best quality between the three, for both $K = 2$, and $K = 8$ snapshots.

Second, to evaluate the reconstruction results at each wavelength level, Figs. 3.8 and 3.9 present the spectral footprint of 6 different points of the target scene, denoted as P1, P2, P3, P4, P5 and P6 in Fig. 7.5, which are compared against a reference spectrum measured with an Ocean Optics USB2000+ spectrometer, assumed to be the ground truth. All spectra are averaged in a 5×5 window corresponding to the same color and normalized to the maximum value in their respective curves. The root-mean-square errors (RMSE) between the reconstructed and the reference spectra are included in each subplot for ease of interpretation. It can be noticed that the curves attained with the optimal (Opt) distribution presents a better fit to the reference spectrum than the Boolean (Bool) and random (Rand) reconstructions; even more, when the number of captured snapshots increases.

Lastly, Figs. 3.10 and 3.11 show 7 out of the 14 reconstructed spectral bands, false-colored with the respective wavelength, for $K = 2$ and $K = 8$, respectively. Notice in these figures that with only two snapshots the reconstructions lack of details, and when 8 snapshots are captured the quality of the reconstruction is quite improved, even

Figure 3.6: Full data cube reconstructions when random, Boolean and optimized patterned filter arrays are used, and $K = 2$ snapshots are captured. The 14-bands data cubes are mapped to false-colored RGB profile to visually compare the color fidelity with the original target scene.

Figure 3.7: Full data cube reconstructions when random, Boolean and optimized patterned filter arrays are used, and $K = 8$ snapshots are captured.The 14-bands data cubes are mapped to false-colored RGB profile to visually compare the color fidelity with the original target scene.

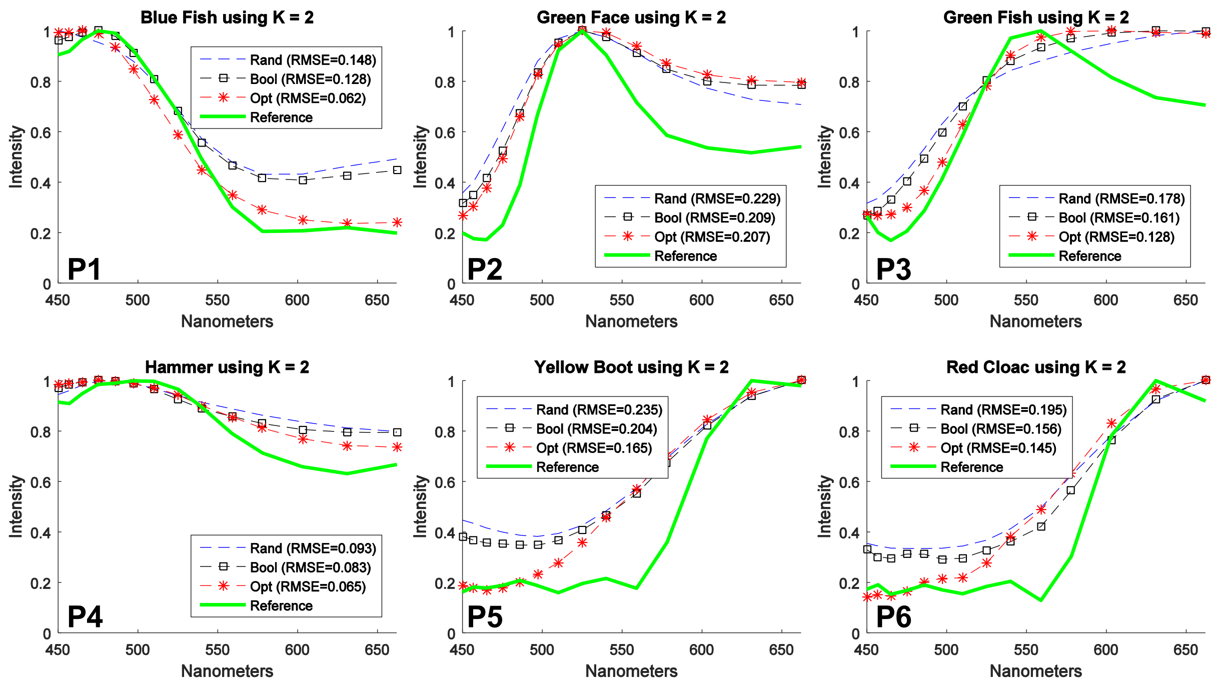


Figure 3.8: Comparison of the spectrum reconstructions when $K = 2$ snapshots are captured. Six different points (P1, P2, P3, P4, P5 and P6) from the target scene (Fig. 7.5) were measured by a spectrometer and compared against the reconstructed data cubes.

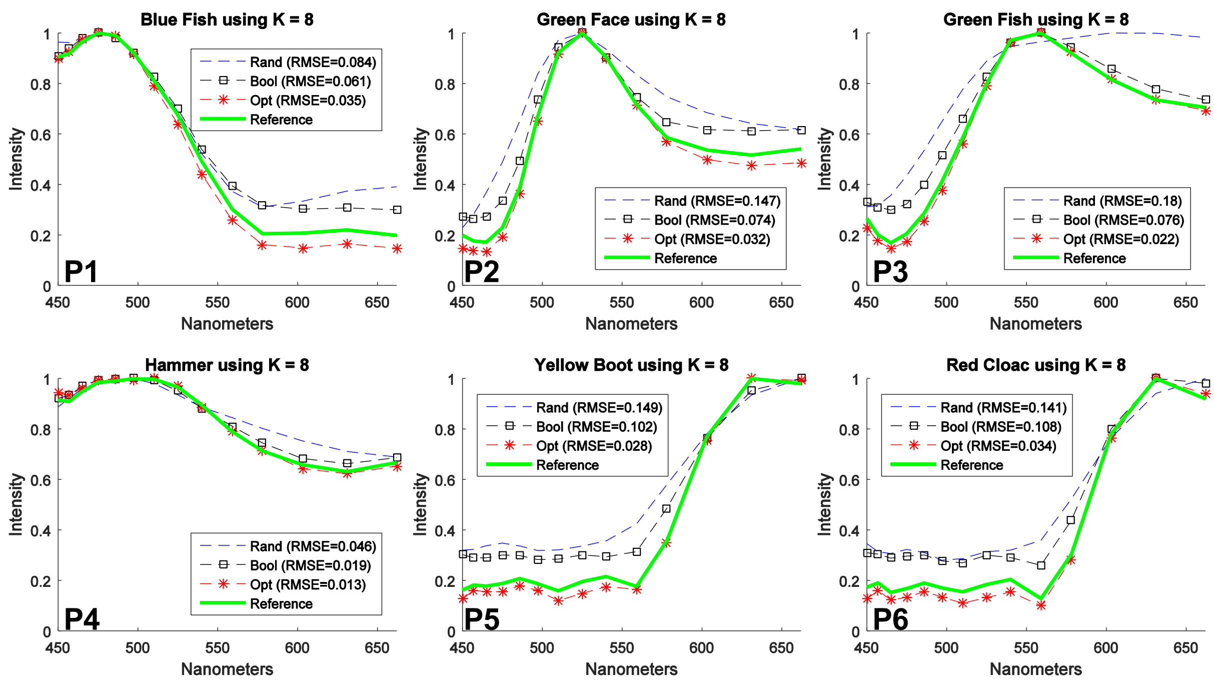


Figure 3.9: Comparison of the spectrum reconstructions when $K = 8$ snapshots are captured. Six different points (P1, P2, P3, P4, P5 and P6) from the target scene (Fig. 7.5) were measured by a spectrometer and compared against the reconstructed data cubes.

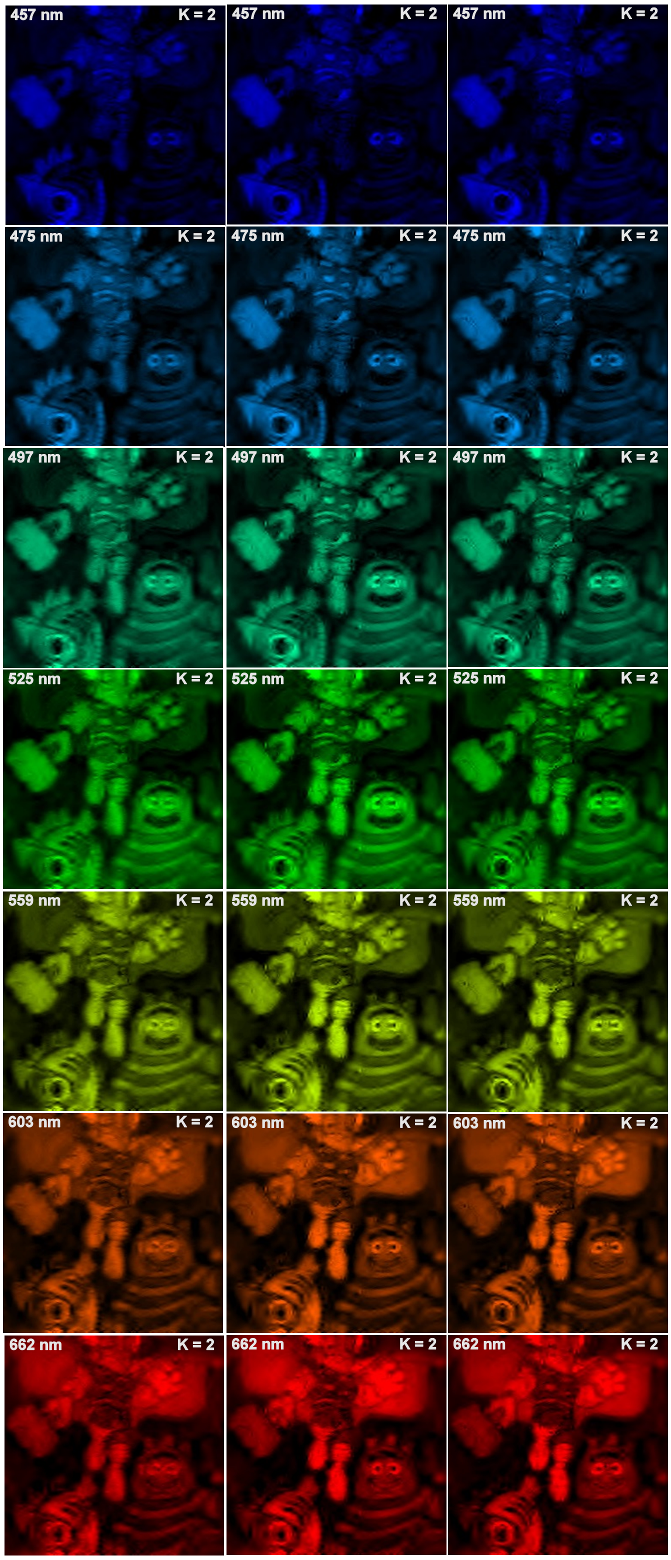


Figure 3.10: Comparison of the reconstructions at monochrome level. The 2^{th}, 4^{th}, 6^{th}, 8^{th}, 10^{th}, 12^{th}, and 14^{th} spectral bands for the 3 different distributions are shown for $K = 2$ snapshots. (First column) Random, (Second column) Boolean, (Third column) Optimized distribution.

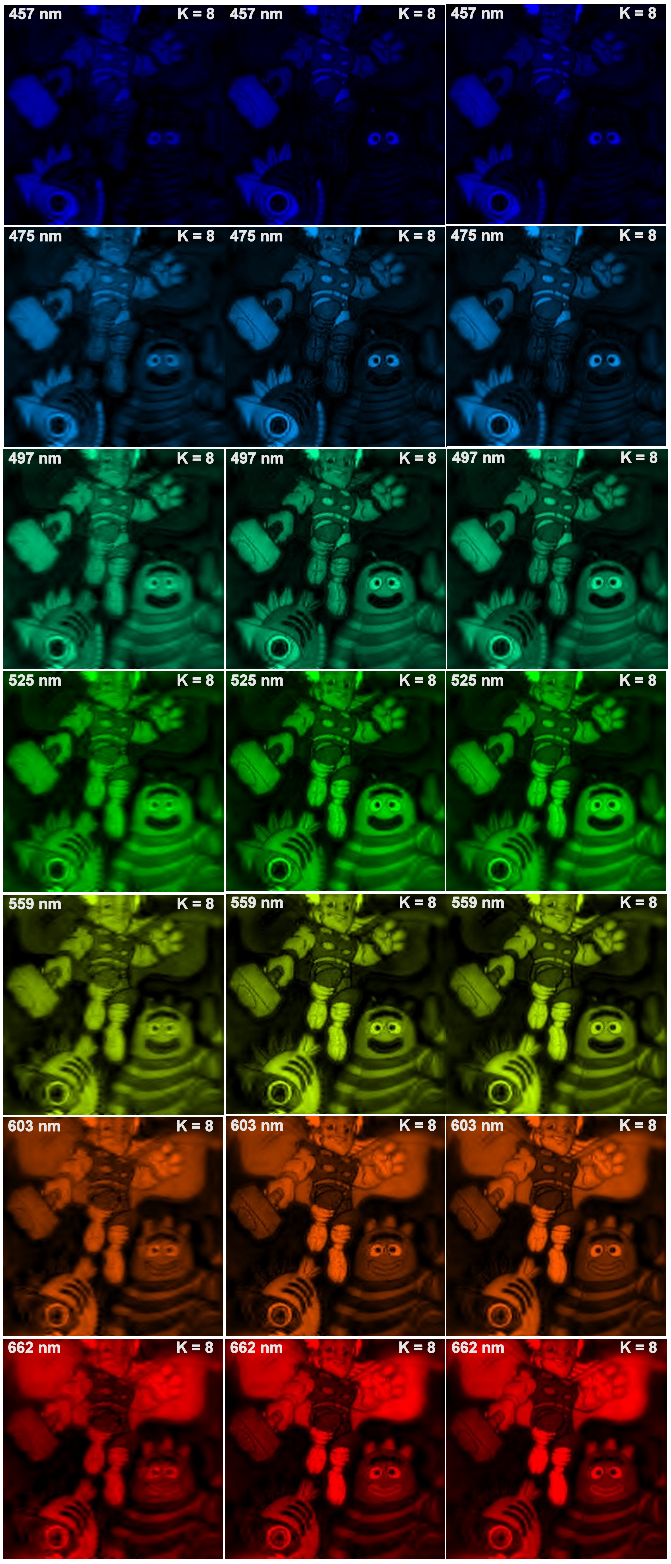


Figure 3.11: Comparison of the reconstructions at monochrome level. The 2^{th}, 4^{th}, 6^{th}, 8^{th}, 10^{th}, 12^{th}, and 14^{th} spectral bands for the 3 different distributions are shown for $K = 8$ snapshots. (First column) Random, (Second column) Boolean, (Third column) Optimized distribution.

more when either the Boolean or the optimized distribution of the pixelated filters is employed. The corresponding band/color discrimination of the monochrome wavelengths can be contrasted with the RGB target scene on Fig. 7.5, where the red cloak of the top toy is well characterized by the 662 nm wavelength, the blue fish fins are well resolved at 486 nm, and the green body of the right toy is well reconstructed at 525 nm.

3.5 Discussion

This chapter focuses on the implementation and analysis of patterned filter arrays into a compressive spectral imaging architecture. We did not include comparisons against the use of photomasks in our setup since in our previous results we already demonstrated that the use of patterned filter arrays entail better reconstruction results even just from simply randomly generated pixelated filter patterns [41, 52]. In addition, we preferred to compare different distributions of the pixelated filters to evaluate their impact on the reconstruction quality. It was shown that either Boolean or the optimized distributions attained better reconstruction results, than simply using random distributions. Note that further optimization on the distribution of the entries can be still performed, as well as a higher-order precision model of the continouos sensing model [42] can be used to achieve even better quality results.

It is shown that thin-film technology opens a door to higher-quality and higher compression sensing methodologies. In this path, our group has recently demonstrated the use of patterned filter arrays attached to FPAs [57]. Although this approach entails a more-compact architecture, in terms of size-weight-and-power, it only permits single snapshot sensing since the filter array is glued to the FPA surface. However, it can be assumed as a first step towards the use of conventional dyed-gels-based technology which has demonstrated to be low-cost, and widely used [58].

As future work, it would be useful to perform a comparison on the use of dyed-gels filters, which are less costly but exhibit a wider extinction ratio, and the impact of their use in coded apertures in compressive spectral imaging. We expect that dyed-gels-based filter arrays will cause a rougher filtering due to the wide response of the gels, and in consequence the spectral results will not be as good (sharp) as the ones shown in this manuscript. That is, the wider the response of the filter, the closer the behavior to black and white (photomasks) coded apertures. However, the use of dyed-gels could promote the expansion of the use of this technology due to cheaper costs. An alternative solution is to use spatial light modulators (SLMs) such as the liquid-crystal-based, which can be exploited to perform spatio-spectral coding by tuning each mirror with a different voltage, thus permitting to attain infinite patterns of higher number of filters just by tuning different voltages [44]. This will entail a compromise between no moving parts and the power required to manipulate the SLM.

3.6 Conclusions

This chapter experimentally demonstrated and analyzed a thin-film-based patterned filter array compressive spectral imaging system. The use of a patterned filter array entailed a higher flexibility in the coding step of the imaging system, allowing to independently encode the space and the spectrum of the input data cube. This flexibility entailed higher degrees of freedom in the design of the sensing operator $\mathbf{H}$, which in turn permitted to better satisfy the conditions of the compressive sensing theory, such as the RIP, therefore attaining higher quality in the reconstructions.

The spatial distribution of the thin-film filters impact the quality of the attained reconstructions, since the optimized distribution outperformed the simply random and Boolean distributions. The experimental results support the promising imaging capabilities introduced in our previous work [41, 52], where the imaging conditions, such as the variability of the cut-off wavelengths of the color filters, and the measurement noise, have been also discussed.

Chapter 4

OPTIMIZATION OF COLOR CODED APERTURES FOR COMPRESSIVE SPECTRAL IMAGING VIA 3D SPREADING

A spectral image can be regarded as a three-dimensional cube where each pixel is not a single value but a vector of intensities representing a spectral reflectance signature. Knowledge of the spectral reflectance of an object permits to know its components, which in turn can be used to detect or classify the elements in a certain image. Compressive spectral imaging (CSI) is a sensing and reconstruction framework, based on the fundamentals of the compressive sensing (CS) theory, which focuses on capturing spectral images efficiently. CSI exploits the highly correlated information contained in the spectral images, by encoding its spectral characteristics with a coded aperture. A coded aperture is a 2D matrix, with each pixel being either a black or translucent element in the conventional setting, or a tiny optical filter in recent developments. The selection of the spectral characteristics of each pixel, as well as the spatial distribution of them within the 2D matrix, control the quality of the captured compressive measurements, which in turn determine the quality of the spectral image to be estimated. State of the art methods have used random distributions of the entries, as suggested by CS, and some optimization procedures have focused on the design of the coded apertures per rows. However, these approaches do not fully exploit the spatial and spectral correlations entitled by CSI imagers. To that end, in this chapter, it is proposed a high-dimensional optimization procedure to design coded apertures for CSI systems, which exploits not only the spectral correlations but the spatial correlations as well. The high-dimensional nature of the optimization refers to the design and selection of each element in the coded aperture, including the cut-off frequencies of the optical filter-based coded apertures, as well as the minimization of the correlation between rows, columns and depth spectral planes of the scene. Simulations analyzing the conditioning of the sensing matrices, as well as the reconstruction quality of the attained spectral images show the improvement entailed by the proposed method.

4.1 Introduction

4.1.1 Motivation

Spectral imaging techniques capture sequences of two-dimensional images along the electromagnetic spectrum, where each pixel is a vector of intensities representing a spectral signature. Information of the spectral properties of an object permits to know its components, which in turn can be used to detect or classify the elements in a certain image. These techniques have been massively used during the last century, mainly in areas such as remote sensing [7] and medical imaging [59]. As a result of the appearance of compressive sensing (CS) [19], new methodologies were proposed to effectively capture spectral images, such as compressive spectral imaging (CSI) [30, 60]. CSI is an imaging framework that focuses on capturing spectral images efficiently, by encoding their spectral characteristics along their spatial extent. CSI systems are modeled based on the principles of a traditional spectrophotometer which records all wavelengths simultaneously. In particular, a ray of light from the scene propagates through a series of optical elements, including apertures, gratings or prisms, which spectrally decompose the ray over before it impinges onto a sensor array. Given the nature of the aperture, it can be designed to allow multiple rays of light to simultaneously impact onto the sensor, meaning that a single pixel will collect light from multiple spatial locations of the underlying object. In other words, CSI systems capture multiplexed measurements, which are then processed by a CS-based estimation algorithm to decouple the spectral profiles within each sensor pixel. An arrangement of 2D spatially located apertures, is widely known as a coded aperture, where each pixel is a black, or translucent element in the conventional setting [23], or a tiny optical filter in recent developments [41]. Both the selection of the spectral characteristics of each pixel, as well as the spatial distribution of them within the 2D arrangement determine the quality of the spectral image to be estimated. Regarding the spectral characteristics, the use of optical filter-based coded apertures, so-called color coded apertures [41, 61, 52], entails a single-step three-dimensional coding on the input data cube, a richer coding strategy compared to black-and-white coded apertures, which only perform 2D spatial coding. Thus, the color coded aperture provides higher flexibility on the selection of which voxels from the scene are going to be integrated on the sensor. On the other hand, regarding the spatial distribution of the entries, coded apertures have been conventionally generated randomly, since they suffice to satisfy the CS requirements [19]. However, better encoding strategies can be obtained by optimizing their spectral characteristics and spatial distribution.

4.1.2 Related work

Optimization of black-and-white coded apertures has been demonstrated in the past. In particular in [55] the authors use a filter-bank strategy to design the coded apertures for agile CSI. In a more recent work [36] they propose a rank minimization strategy to perform selective spectral sensing. Application-free strategies have been also reported in [62], where the optimization of coded apertures is carried-out by exploiting blue-noise halftone ideas. Other works have focused on the minimization of the restricted isometry constant [63, 64], and the minimization of the mutual coherence constant using dictionary learning strategies [65, 66]. Regarding color coded apertures, and given its recent proposal, just a single work has reported their optimization [41]. In that work, a row per row optimization (row-wise) approach based on concentration of measure was proposed. This approach, however, does not fully exploit the spatial correlations of the spectral image as it individualizes the optimization from the spatial extent.

4.1.3 Contributions

In this chapter, we propose an optimization procedure to design color coded apertures for CSI systems, which exploits not only the spectral but the spatial correlations within an spectral image. The optimization refers to the design and selection of each element in the coded aperture, including the cut-off frequencies of the optical filters, as well as the exploitation of the correlation between rows, columns and spectral planes of the scene. The proposed optimization method seeks for a better conditioned sensing matrix, so that it satisfies tightly the restricted isometry property. To do this, a higher-order discretization model of the continuous phenomena within color coded aperture CSI systems is first proposed. This model entails a better approximation of the real phenomena, which in turn leads to a more precise, still structured, sparse sensing matrix. An iterative random-walk optimization algorithm is then developed to carefully design the color coded apertures, which minimizes the shifted cross-correlation between the columns of the, more precise, sparse sensing matrix. Exploiting the fact that a color coded aperture can be modeled as a 3D black-and-white coded aperture, the optimization algorithm thus seeks to spread the translucent elements (band-pass filters of the color coded aperture) along the 3D cube extent. The latter is achieved by 3D filtering the coded aperture with a 3D Euclidean filter, in each iteration of the random-walk.

The performance of the proposed 3D optimized color coded apertures are then compared against the random, and the row-wise optimized, showing the benefits in terms of the conditioning of the sensing matrix, and in terms of the peak-signal-to-noise-ratio (PSNR) and structural similarity index (SSIM) metrics in 6 different spectral scenes.

4.2 Higher-order Compressive Spectral Imaging Model with Colored Coded Apertures

Different CSI architectures have been proposed to date, most of them using black-and-white coded apertures or spatial light modulators (SLMs) [23, 66], and just recent works use color coded apertures to perform the coding [41, 57, 61]. Prisms and gratings are conventionally used to disperse the optical field from the scene, and monochrome sensors are preferred over RGB sensors to integrate the modulated field, although the latter can provide better spectral estimations [58]. For the scope of this chapter, we focus on the single-disperser coded aperture snapshot spectral imager, which uses a prism, a monochrome camera, and has been demonstrated using black-and-white (CASSI) [1, 23] and color coded apertures (C-CASSI) [41, 52, 61]. Figure 4.1 shows an sketch of both architectures. There, the (colored-) coded aperture is placed in the image plane of the objective lens. Then, a set of relay lenses transmits the encoded scene through the dispersive element and onto the sensor, where multiplexing occurs. Note that, color coded apertures have demonstrated improved results compared with black-and-white apertures due to the richer coding entailed by the pixelated optical filters [41]. Therefore we focus on the optimization of the C-CASSI, although an ease extension can be done for CASSI as well, given the high similarities between both architectures.

In this section, we will first show the continuous model of the C-CASSI; then the higher order discretization is detailed, as well as the matrix representation.

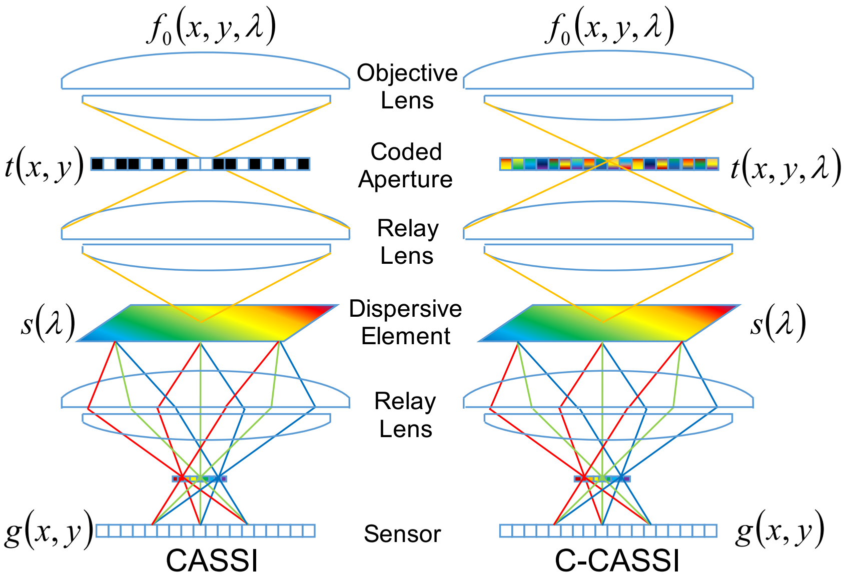


Figure 4.1: CSI systems of interest. (Left) Coded aperture snapshot spectral imager (CASSI). (Right) CASSI with color-coded apertures (C-CASSI)

4.2.1 C-CASSI Continuous Model

Let denote the input spectral scene as the 3D function $f_0(x, y, \lambda)$, and the color coded aperture as the 3D function $t(x, y, \lambda)$, where x, y represent the spatial coordinates and λ the spectral domain. The coded field just after the color coded aperture is given by

$$f_1(x, y, \lambda) = t(x, y, \lambda) f_0(x, y, \lambda). \tag{4.1}$$

The relay lens projects the coded field through the prism exhibiting a dispersion function $s(\lambda)$; therefore, the dispersed field just after the prism is given by,

$$\begin{aligned} f_2(x, y, \lambda) &= \iint f_1(x', y', \lambda) h(x - x' - s(\lambda), y - y', \lambda) dx' dy' \\ &= \iint t(x', y', \lambda) f_0(x', y', \lambda) h(x - x' - s(\lambda), y - y', \lambda) dx' dy'. \end{aligned} \tag{4.2}$$

Finally, the coded and dispersed field is integrated on the sensor, along its sensitive spectrum (Λ) as,

$$\begin{aligned} g(x,y) &= \int_{\Lambda} f_2(x,y,\lambda)d\lambda \\ &= \iiint t(x',y',\lambda)f_0(x',y',\lambda)h(x-x'-s(\lambda),y-y',\lambda)dx'dy'd\lambda. \end{aligned} \quad (4.3)$$

Equation (4.3) can be simplified by assuming an ideal space-invariant optical impulse response, so that $h(x,y,\lambda) = \delta(x,y,\lambda)$. Note that $\delta(x'-x-s(\lambda), y'-y, \lambda) = 1 \Leftrightarrow x' = x + s(\lambda)$ and $y' = y$. Thus, Eq. (4.3) becomes,

$$g(x,y) = \int_{\Lambda} t(x+s(\lambda),y,\lambda)f_0(x+s(\lambda),y,\lambda)d\lambda. \quad (4.4)$$

4.2.2 C-CASSI Higher-Order Discrete Model

The continuous measurement $g(x,y)$ can be discretized by means of the pixelated detector array described by

$$p(m,n;x,y) = \text{rect}\left(\frac{x}{\Delta_d} - m, \frac{y}{\Delta_d} - n\right), \quad (4.5)$$

where rect represents the rectangular function, and Δ_d is the detector pixel size. Note that we assume the rectangular function is not centered to facilitate the notation, that is, $\text{rect}\left(\frac{x}{\Delta}\right) = \text{rect}\left(\frac{x'}{\Delta} - \frac{1}{2}\right)$. Then, the discrete measurement at the $(m,n)^{th}$ sensor pixel is given by,

$$\begin{aligned} g_{n,m} &= \iint g(x,y)p(m,n;x,y)dxdy \\ &= \iiint t(x+s(\lambda),y,\lambda)f_0(x+s(\lambda),y,\lambda)\text{rect}\left(\frac{x}{\Delta_d} - m, \frac{y}{\Delta_d} - n\right) d\lambda dxdy \end{aligned} \quad (4.6)$$

Similarly, the patterned filter array is discretized as,

$$t(x,y,\lambda) = \sum_{n',m',k'} t_{n',m',k'} \text{rect}\left(\frac{x}{\Delta_c} - m', \frac{y}{\Delta_c} - n', \frac{\lambda}{\Delta_L} - k'\right), \quad (4.7)$$

where $t_{n',m',k'} \in [0,1]$ represents the filtering on the $(n',m',k')^{th}$ voxel of the scene, Δ_c is the coded aperture pixel size, and Δ_L represents the bandwidth of the optical filters. Therefore, Eq. (4.6) can be rewritten as,

$$\begin{aligned} g_{n,m} &= \sum_{n',m',k'} t_{n',m',k'} \iiint f_0(x+s(\lambda),y,\lambda)\text{rect}\left(\frac{x+s(\lambda)}{\Delta_c}-m', \frac{y}{\Delta_c}-n', \frac{\lambda}{\Delta_L}-k'\right) \\ &\times\text{rect}\left(\frac{x}{\Delta_d}-m, \frac{y}{\Delta_d}-n\right) dxdyd\lambda. \end{aligned} \tag{4.8}$$

The spatial resolution of the system is given by the pixel size of the coded aperture (Δ_c), and the spectral resolution depends on the ratio between the dispersion function ($s(\lambda)$) and the pixel size of the sensor (Δ_d). In particular,

$$f_0(x,y,\lambda) = \sum_{n',m',k} f_{n',m',k}\text{rect}\left(\frac{x}{\Delta_c}-m', \frac{y}{\Delta_c}-n', \frac{\lambda}{\Delta_d}-k\right), \tag{4.9}$$

where $f_{n',m',k} \in [0,1]$ represents the normalized scene intensity at the $(n',m',k')^{th}$ voxel. Therefore, Eq. (4.8) becomes,

$$\begin{aligned} g_{n,m} &= \sum_{n',m',k',k} t_{n',m',k'} f_{n',m',k} \\ &\iiint \text{rect}\left(\frac{x+s(\lambda)}{\Delta_c}-m', \frac{y}{\Delta_c}-n', \frac{\lambda}{\Delta_d}-k\right) \\ &\times\text{rect}\left(\frac{x+s(\lambda)}{\Delta_c}-m', \frac{y}{\Delta_c}-n', \frac{\lambda}{\Delta_L}-k'\right)\text{rect}\left(\frac{x}{\Delta_d}-m, \frac{y}{\Delta_d}-n\right) dxdyd\lambda. \end{aligned} \tag{4.10}$$

For ease of calculations, let assume that $\Delta = \Delta_c = \Delta_d$, and that the number of optical filters in the color coded aperture, indexed by k', can be discretized to match with the wavelengths of the datacube indexed by k (that is, $\Delta_d = \Delta_L$). Also, given that the dispersion occurs only along the x-axis, then $n = n'$, that is, the number of rows in the datacube and in the sensor remains the same, and thus they can be similarly indexed. Therefore Eq. (4.10) becomes,

$$\begin{aligned} g_{n,m} &= \sum_{m',k} t_{n,m',k} f_{n,m',k} \qquad (4.11) \\ &\iiint \text{rect}\left(\frac{x+s(\lambda)}{\Delta}-m', \frac{y}{\Delta}-n, \frac{\lambda}{\Delta}-k\right)\text{rect}\left(\frac{x}{\Delta}-m, \frac{y}{\Delta}-n\right) dxdyd\lambda. \end{aligned}$$

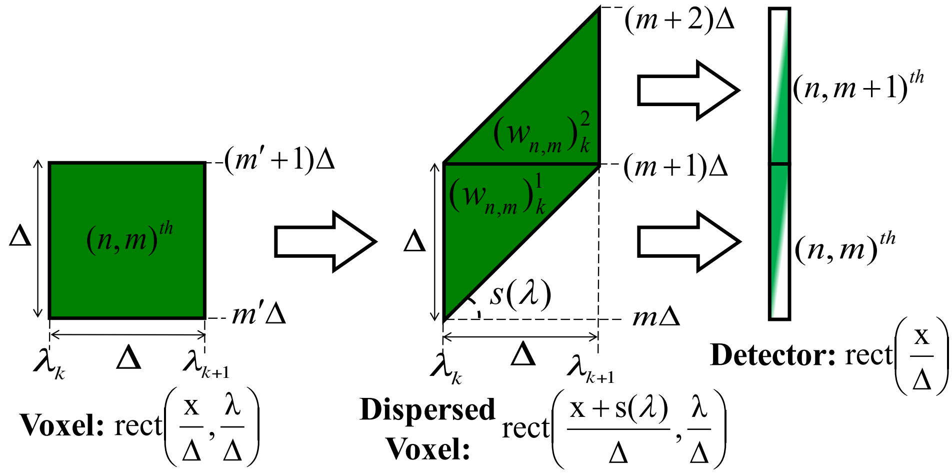


Figure 4.2: Propagation of a single voxel through the CSI system

To evaluate the triple-integral in Eq. (4.11), let us define the dispersion function as, $s(\lambda_k) - s(\lambda_{k+1}) = \Delta$; that is, the bandwidth of each spectral band of the datacube is given by how much of the dispersion spread impinges onto a single pixel of the sensor Δ. Figure 4.2 shows the propagation of a single voxel of the datacube through the dispersive element, seen from atop. Note that the single voxel (n, m, k) spreads onto more than a single sensor pixel (n, m) and $(n, m + 1)$, due to the continuous nature of the dispersion function [42]. Mathematically, the portion of the $(n, m', k)^{th}$ voxel that impinges onto the $(n, m)^{th}$ sensor pixel can be written as,

$$(w_{n,m})_k^{m-k-m'} = \iiint\limits_{n\Delta;m\Delta;\lambda_k}^{(n+1)\Delta;(m+1)\Delta;\lambda_{k+1}} \mathrm{rect}\left(\frac{x+s(\lambda)}{\Delta} - m', \frac{y}{\Delta} - n, \frac{\lambda}{\Delta} - k\right) \mathrm{rect}\left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n\right) dxdyd\lambda, \tag{4.12}$$

therefore, Eq. (4.11) can be rewritten as,

$$g_{n,m} = \sum_{m',k} t_{n,m',k} f_{n,m',k} \left(w_{n,m}\right)_k^{m-k-m'}. \tag{4.13}$$

Letting $r = m - k - m'$, index the number of sensor pixels a single voxel impinges onto, $r = 0, \ldots, R - 1$ with R being the maximum number of pixels affected, the succinctly version of the discrete measurement is given by

$$g_{n,m} = \sum_{r,k} t_{n,m-k-r,k} f_{n,m-k-r,k} \left(w_{n,m}\right)_k^r . \tag{4.14}$$

4.2.3 C-CASSI Matrix Representation

Let assume the underlying datacube is of size $N \times N \times L$, with $N \times N$ spatial pixels and L spectral bands. Then, denote $Q = N^2 L$ as the number of voxels of the datacube, $V = N(N+L+R-2)$ as the number of sensor pixels, and K as the number of snapshots to be captured. The discrete values of the color coded aperture $t_{n',m',k'}$, the datacube $f_{n',m',k'}$, and the dispersion weights $(w_{n,m})_k^r$ can be transformed to vector form by,

$$(\mathbf{f})_q = f_{\left(q-\lfloor \frac{q}{N} \rfloor N\right),\left(\lfloor \frac{q}{N} \rfloor - \lfloor \frac{q}{N^2} \rfloor N\right),\lfloor \frac{q}{N^2} \rfloor}, \tag{4.15}$$

$$\left(\mathbf{t}_k^\ell\right)_p = t^\ell_{\left(p-\lfloor \frac{p}{N} \rfloor N\right),\lfloor \frac{p}{N} \rfloor,k}, \tag{4.16}$$

$$(\mathbf{w}_k^r)_p = (w_{\left(p-\lfloor \frac{p}{N} \rfloor N\right),\lfloor \frac{p}{N} \rfloor})_k^r \tag{4.17}$$

for $p = 0, 1, \ldots, N^2 - 1$, $q = 0, 1, \ldots, Q - 1$ and $\ell = 0, 1, \ldots, K - 1$. Then, it is possible to denote the compressive measurements given in Eq. (4.14) as the linear system,

$$\mathbf{g}^\ell = \mathbf{H}^\ell \mathbf{f} = \mathbf{P}\mathbf{T}^\ell \mathbf{f}, \tag{4.18}$$

where the sensing matrix $\mathbf{H}^\ell$ can be decomposed in the effect of the color coded aperture (matrix $\mathbf{T}^\ell \in \mathbb{R}^{Q\times Q}$), and that of the dispersive element (matrix $\mathbf{P} \in \mathbb{R}^{V\times Q}$), such that,

$$\mathbf{T}^\ell = \begin{bmatrix} \text{diag}(\mathbf{t}_0^\ell) & \mathbf{0}_{N^2\times N^2} & \cdots & \mathbf{0}_{N^2\times N^2} \\ \mathbf{0}_{N^2\times N^2} & \text{diag}(\mathbf{t}_1^\ell) & \cdots & \mathbf{0}_{N^2\times N^2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N^2\times N^2} & \mathbf{0}_{N^2\times N^2} & \cdots & \text{diag}(\mathbf{t}_{L-1}^\ell) \end{bmatrix} \tag{4.19}$$

and $\mathbf{P} = \sum_{r=0}^{R-1} \widetilde{\mathbf{P}}_r$, with,

$$\widetilde{\mathbf{P}}_r = \begin{bmatrix} & \mathbf{0}_{rN\times Q} & & \\ \text{diag}(\mathbf{w}_0^r) & \mathbf{0}_{N\times N^2} & \cdots & \mathbf{0}_{N\times N^2} \\ \mathbf{0}_{N\times N^2} & \text{diag}(\mathbf{w}_1^r) & \cdots & \mathbf{0}_{N\times N^2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N\times N^2} & \mathbf{0}_{N\times N^2} & \cdots & \text{diag}(\mathbf{w}_{L-1}^r) \\ & \mathbf{0}_{(R-1-r)N\times Q} & & \end{bmatrix}, \tag{4.20}$$

where $\mathbf{0}_{a\times a}$ represents an $a \times a$ all-zero matrix, and diag($\mathbf{v}$) an $N^2 \times N^2$ matrix, with the vector $\mathbf{v}$ in the diagonal. Figure 4.3 presents an sketch of the higher-order sensing matrix $\mathbf{H}$ in Eq. (4.18), for a $16 \times 16 \times 4$ datacube, with $R = 2$ and $K = 1$.

4.3 Sensing Matrix Analysis

The sensing process in Eq. (4.18) models the sensing operator as a "black-box" process with $\mathbf{f}$ as the input, and $\mathbf{g}^\ell$ as the output. In order to recover an estimation of the datacube $\mathbf{f}$ from the measurements $\mathbf{g}^\ell$, the inverse of the sensing operator $\mathbf{H}^\ell$ must be found. However, note that this linear system is under-determined, and exhibits a compression ratio of $R = V/Q$, thus direct inversion is not possible. Therefore, we need to force the sensing operator to exhibit a close to well-conditioning behavior in order to obtain a good inversion. Particularly, let first focus on the entries of the sensing matrix $\mathbf{H}$, and their dependency on the color coded aperture. The goal is to express the sensing matrix in terms of the entries of the coded aperture, such that a correct design of the later entails a well conditioning of the former.

4.3.1 Transfer Function (Matrix H) Entries

Matrices in Eqs. (4.19) and (4.20) can be denoted also as $\mathbf{T}^\ell = [\tilde{\mathbf{t}}_0^\ell, \tilde{\mathbf{t}}_1^\ell, \ldots, \tilde{\mathbf{t}}_{Q-1}^\ell]$ and $\mathbf{P} = [\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{Q-1}]$, respectively, where $\tilde{\mathbf{t}}_q \in \mathrm{IR}^Q$ and $\mathbf{p}_q \in \mathrm{IR}^V$. Therefore, the entries of the matrix $\mathbf{H}^\ell = \left[\mathbf{h}_0^\ell, \mathbf{h}_1^\ell, \ldots, \mathbf{h}_{Q-1}^\ell\right]$ are given by,

$$(\mathbf{h}_q^\ell)_v = \sum_{q'=0}^{Q-1} (\mathbf{p}_{q'})_v (\tilde{\mathbf{t}}_q^\ell)_{q'}, \tag{4.21}$$

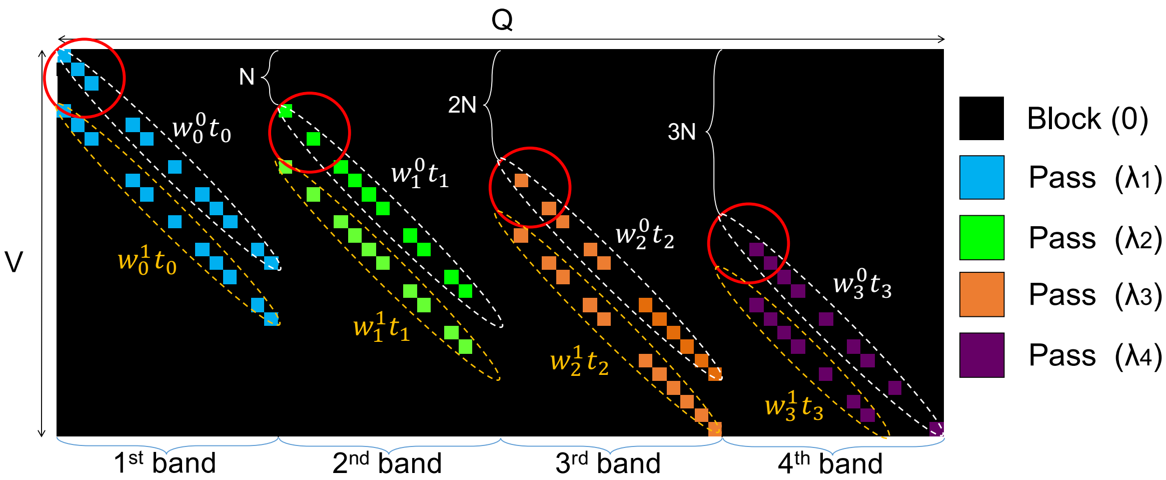


Figure 4.3: Sketch of the higher-order sensing matrix $\mathbf{H}^\ell$ for a datacube of size $16 \times 16 \times 4$. The diagonals refer to the wavelength-dependent coding pattern $(\mathbf{t}_k^\ell)$ weighted by the corresponding dispersion distribution $(\mathbf{w}_k^r)$. The N off-set between the diagonals correspond to the column-wise representation of the dispersion.

where,

$$(\tilde{\mathbf{t}}_q^\ell)_{q'} = \begin{cases} \left(\mathbf{t}^\ell_{\lfloor \frac{q}{N^2} \rfloor}\right)_{q' - \lfloor \frac{q'}{N^2} \rfloor N^2}, & \text{if } q = q' \\ 0, & \text{otherwise} \end{cases} \tag{4.22}$$

$$(\mathbf{p}_{q'})_v = \begin{cases} \left(\mathbf{w}^r_{\lfloor \frac{q'}{N^2} \rfloor}\right)_{q' - \lfloor \frac{q'}{N^2} \rfloor N^2}, & \text{if } v = q' - q''_r \\ 0, & \text{otherwise} \end{cases} \tag{4.23}$$

with $q''_r = \left\lfloor \frac{q'}{N^2} \right\rfloor N^2 + rN$, for $v = 0, 1, \ldots, V-1$. Expanding Eq. (4.21) in terms of Eqs. (4.22) and (4.23), and noting that Eq. (4.22) is a block-diagonal matrix, it is obtained,

$$(\mathbf{h}_q^\ell)_v = \begin{cases} \left(\mathbf{w}^r_{\lfloor \frac{q}{N^2} \rfloor}\right)_{q - \lfloor \frac{q}{N^2} \rfloor N^2} \left(\mathbf{t}^\ell_{\lfloor \frac{q}{N^2} \rfloor}\right)_{q - \lfloor \frac{q}{N^2} \rfloor N^2}, & \text{if } v = g_r \\ 0, & \text{o.w.} \end{cases} \tag{4.24}$$

where $g_r = q - \left\lfloor \frac{q}{N^2} \right\rfloor N^2 + rN$. In order to obtain a reconstruction from a under-determined linear system of equations, certain conditions must be imposed such that the recovery becomes feasible. The theory of compressive sensing (CS) has emerged to solve this kind of problems in a reliable way. One of the general conditions CS imposes is over the input data. It says that, the input data should follow a compact (sparse) representation when projected over an orthonormal basis domain. This can be mathematically modeled as,

$$\mathbf{g} = \mathbf{H}\mathbf{f} = \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\theta} \tag{4.25}$$

where $\mathbf{H} = \left[(\mathbf{H}^0)^T, (\mathbf{H}^1)^T, \dots, (\mathbf{H}^{K-1})^T\right]^T$ and $\boldsymbol{\theta} \in \mathrm{IR}^Q$ is a compact (or sparse) version of the signal $\mathbf{f}$, in the basis $\boldsymbol{\Psi} \in \mathrm{IR}^{Q\times Q}$. The vector $\boldsymbol{\theta}$ is said to be a compact representation of $\mathbf{f}$ because all the information of the latter is concentrated in just few coefficients of the former. That is, $\|\mathrm{supp}(\boldsymbol{\theta})\| = \|\{j : |\boldsymbol{\theta}_j| > 0\}\| = S$, with $S << Q$. The entries of the multiple-shot matrix $\mathbf{H} \in \mathrm{IR}^{KV\times Q}$ in Eq. (4.25) are given by,

$$(\mathbf{h}_q)_{i'} = \begin{cases} \left(\mathbf{w}^r_{\lfloor \frac{q}{N^2} \rfloor}\right)_{q-\lfloor \frac{q}{N^2}\rfloor N^2} \left(\mathbf{t}^{\lfloor \frac{i'}{V} \rfloor}_{\lfloor \frac{q}{N^2} \rfloor}\right)_{q-\lfloor \frac{q}{N^2}\rfloor N^2}, & \text{if } i' = g'_r \\ 0, & \text{o.w.} \end{cases} \tag{4.26}$$

where $g'_r = g_r - \lfloor \frac{i'}{V} \rfloor V$, for $i' = 0, 1, \dots, KV - 1$. Note that the entries of $\mathbf{H}$ depend indeed on the entries of the color coded aperture and the dispersion weight distribution.

4.3.2 Analysis of the Compressive Sensing Matrix A

By multiplying the transfer function $\mathbf{H}$ with the representation basis $\boldsymbol{\Psi}$, the compressive sensing matrix $\mathbf{A} \in \mathrm{IR}^{KV\times Q}$ is obtained, as shown in Eq. (4.25). Based on the sparse assumption, CS permits to recover an estimation of the data cube $\hat{\mathbf{f}}$ from the compressive measurements $\mathbf{g}$ by solving

$$\hat{\mathbf{f}} = \boldsymbol{\Psi}^{-1}\left(\mathrm{argmin}_{\boldsymbol{\theta}'} \|\mathbf{g} - \mathbf{A}\boldsymbol{\theta}'\|_2^2 + \tau\|\boldsymbol{\theta}'\|_1\right), \tag{4.27}$$

with $\tau > 0$ being a regularization parameter, which penalizes the seeking of the sparsest solution. Apart from the sparsity assumption over the input data, CS also requires that the matrix $\mathbf{A}$ satisfy certain condition, such as the restricted isometry property (RIP), or similarly, that the matrices $\mathbf{H}$ and $\boldsymbol{\Psi}$ be mutually incoherent [19]. Satisfying these conditions guarantee, with high probability, the correct estimation of the input signal. In particular, the RIP for the matrix $\mathbf{A}$ of order S is defined as the smallest constant δ_S such that,

$$(1-\delta_S)\,\|\boldsymbol{\theta}\|_2^2 \leq \|\mathbf{A}\boldsymbol{\theta}\|_2^2 \leq (1+\delta_S)\,\|\boldsymbol{\theta}\|_2^2, \tag{4.28}$$

where,

$$\delta_S = \max_{\rho\subset[Q],|\rho|\leq S} \|\mathbf{A}_{|\rho|}^T\mathbf{A}_{|\rho|} - \mathbb{I}\|_2^2, \tag{4.29}$$

with $\mathbf{A}_{|\rho|}$ being a $KV \times |\rho|$ matrix whose columns are equal to $|\rho|$ columns of $\mathbf{A}$ indexed by the set ρ, and $\mathbb{I}$ is an identity matrix [41]. Note that, to satisfy the RIP condition, the matrix $\mathbf{A}$ must be carefully designed, such that a small δ_S is obtained. The design of the sensing matrix $\mathbf{A}$ relies directly on the selection of a suitable representation matrix $\boldsymbol{\Psi}$ along with the structure of the sensing matrix $\mathbf{H}$. In our case, $\mathbf{H}$ directly depends on the entries of the color coded apertures $\mathbf{t}_k^{\ell}$ and the dispersion function weights $\mathbf{w}_k^r$. In particular, the entries of $\mathbf{A} = \mathbf{H}\boldsymbol{\Psi} = [\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_{Q-1}]$, where $\boldsymbol{\Psi} = \left[\boldsymbol{\psi}_0\;\boldsymbol{\psi}_1\ldots\boldsymbol{\psi}_{Q-1}\right]$, are given by,

$$\begin{aligned}
(\mathbf{a}_{j'})_{i'} =& \sum_{q=0}^{Q-1} (\mathbf{h}_q)_{i'}\left(\boldsymbol{\psi}_{j'}\right)_q \\
=& \sum_{\{q|i'=g_r'\}} \left(\mathbf{w}^r_{\lfloor\frac{q}{N^2}\rfloor}\right)_{q-\lfloor\frac{q}{N^2}\rfloor N^2} \left(\mathbf{t}^{\lfloor\frac{i'}{V}\rfloor}_{\lfloor\frac{q}{N^2}\rfloor}\right)_{q-\lfloor\frac{q}{N^2}\rfloor N^2} \left(\boldsymbol{\psi}_{j'}\right)_q \\
=& \sum_{r=0}^{R-1}\sum_{k=0}^{L-1} (\mathbf{w}^r_k)_{i'-(k+r)N-\lfloor\frac{i'}{V}\rfloor V} \left(\mathbf{t}_k^{\lfloor\frac{i'}{V}\rfloor}\right)_{i'-(k+r)N-\lfloor\frac{i'}{V}\rfloor V} \left(\boldsymbol{\psi}_{j'}\right)_{i'-(k+r)N-\lfloor\frac{i'}{V}\rfloor V+kN^2}
\end{aligned} \tag{4.30}$$

where the last equation is obtained after some math explained in Appendix 4.7. In consequence, the matrix $\mathbf{A}_{|\rho|}$ is given by,

$$\begin{aligned}
\left(\mathbf{A}_{|\rho|}\right)_{i',l} =& \left(\mathbf{a}_{\Omega_l}\right)_{i'}, \\
=& \sum_{r=0}^{R-1}\sum_{k=0}^{L-1} (\mathbf{w}^r_k)_{i'-(k+r)N-\lfloor\frac{i'}{V}\rfloor V} \left(\mathbf{t}_k^{\lfloor\frac{i'}{V}\rfloor}\right)_{i'-(k+r)N-\lfloor\frac{i'}{V}\rfloor V} \left(\boldsymbol{\psi}_{\Omega_l}\right)_{i'-(k+r)N-\lfloor\frac{i'}{V}\rfloor V+kN^2}
\end{aligned} \tag{4.31}$$

for $l = 0, \ldots, |\rho|-1$ and $\Omega_l \in \{0, \ldots, Q-1\}$.

4.3.3 Grammian Matrix Analysis

Following Eq. (4.29), let define the Gram matrix of $\left(\mathbf{A}_{|\rho|}\right)$ as $\mathbf{G}_{|\rho||\rho|} = \mathbf{A}_{|\rho|}^T\mathbf{A}_{|\rho|}$. Therefore, the constant δ_S can be rewritten as, $\delta_S = \max_{\rho\subset[Q],|\rho|\leq S}\|\mathbf{G}_{|\rho||\rho|} - \mathbb{I}\|_2^2$, where the entries of $\mathbf{G}_{|\rho||\rho|}$ are given by, $\left(\mathbf{G}_{|\rho||\rho|}\right)_{l_1,l_2} =$

$$
\begin{aligned}
&= \sum_{i'=0}^{KV-1} \left(\mathbf{A}^T_{|\rho|}\right)_{l_1,i'} \left(\mathbf{A}_{|\rho|}\right)_{i',l_2} \\
&= \sum_{i'=0}^{KV-1} \left[\sum_{r=0}^{R-1}\sum_{k=0}^{L-1} (\mathbf{w}_k^r)_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V} \left(\mathbf{t}_k^{\lfloor \frac{i'}{V}\rfloor}\right)_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V} \left(\boldsymbol{\psi}_{l_1}\right)_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V+kN^2}\right] \\
&\times \left[\sum_{\tilde{r}=0}^{R-1}\sum_{\tilde{k}=0}^{L-1} (\mathbf{w}_{\tilde{k}}^{\tilde{r}})_{i'-(\tilde{k}+\tilde{r})N-\lfloor \frac{i'}{V}\rfloor V} \left(\mathbf{t}_{\tilde{k}}^{\lfloor \frac{i'}{V}\rfloor}\right)_{i'-(\tilde{k}+\tilde{r})N-\lfloor \frac{i'}{V}\rfloor V} \left(\boldsymbol{\psi}_{l_2}\right)_{i'-(\tilde{k}+\tilde{r})N-\lfloor \frac{i'}{V}\rfloor V+\tilde{k}N^2}\right] \qquad (4.32)
\end{aligned}
$$

where $l_1, l_2 = 0, \ldots, |\rho|-1$. Equation (4.32) can be rewritten succinctly as, $\left(\mathbf{G}_{|\rho||\rho|}\right)_{l_1,l_2} =$

$$
\begin{aligned}
&\sum_{\ell=0}^{K-1}\sum_{i=0}^{V-1}\sum_{r=0}^{R-1}\sum_{k=0}^{L-1}\sum_{\tilde{r}=0}^{R-1}\sum_{\tilde{k}=0}^{L-1} (\mathbf{w}_k^r)_{i-(k+r)N} (\mathbf{w}_{\tilde{k}}^{\tilde{r}})_{i-(\tilde{k}+\tilde{r})N} \\
&\quad \times \left(\mathbf{t}_k^\ell\right)_{i-(k+r)N} \left(\mathbf{t}_{\tilde{k}}^\ell\right)_{i-(\tilde{k}+\tilde{r})N} \left(\boldsymbol{\psi}_{l_1}\right)_{i-(k+r)N+kN^2} \left(\boldsymbol{\psi}_{l_2}\right)_{i-(\tilde{k}+\tilde{r})N+\tilde{k}N^2}. \qquad (4.33)
\end{aligned}
$$

Note that minimizing δ_S requires minimizing the maximum off-diagonal element of the matrix $\mathbf{G}_{|\rho||\rho|} - \mathbb{I}$. Therefore, minimizing δ_S requires minimizing the cross-products in Eq. (4.33), or in other words, the shifted-cross correlation of the Grammian matrix.

4.4 Higher order color coded aperture optimization

Given that the weights $\mathbf{w}_k^r$ and the basis $\boldsymbol{\Psi}$ are fixed and bounded, the first depends on the dispersive element, and the second on a-priori selection, the entries of $\mathbf{G}_{|\rho||\rho|}$ in Eq. (4.33) can be bounded by the variable,

$$
\beta = \sum_{\ell=0}^{K-1}\sum_{i=0}^{V-1}\sum_{r=0}^{R-1}\sum_{k=0}^{L-1}\sum_{\tilde{r}=0}^{R-1}\sum_{\tilde{k}=0}^{L-1} \left(\mathbf{t}_k^\ell\right)_{i-(k+r)N} \left(\mathbf{t}_{\tilde{k}}^\ell\right)_{i-(\tilde{k}+\tilde{r})N} \qquad (4.34)
$$

whose minimization yields to an upper bound of the parameter δ_s, and so, to better satisfy the RIP. Remark that keeping the basis vectors ($\boldsymbol{\psi}$) yield to a joint design of the sensing matrix and the representation basis, which was deeply discussed in [65] for a general sensing matrix. Note however, that designing both matrices implies a high computational burden, as well as it leads to an ad-hoc basis design.

Note that β can be minimized by minimizing the products of the entries of the color coded aperture $(\mathbf{t}_k^\ell)_{i-(k+r)N} \times (\mathbf{t}_{\tilde{k}}^\ell)_{i-(\tilde{k}+\tilde{r})N}$. For ease of interpretation let express Eq. (4.34) in terms of the actual entries of the color coded aperture by using the 1D-to-3D transformation introduced in Eq. (4.16). Thus,

$$\beta = \sum_{\ell=0}^{K-1} \sum_{i=0}^{N-1} \sum_{j=0}^{\frac{V}{N}-1} \sum_{r=0}^{R-1} \sum_{k=0}^{L-1} \sum_{\tilde{r}=0}^{R-1} \sum_{\tilde{k}=0}^{L-1} t^\ell_{i,j-(k+r),k} t^\ell_{i,j-(\tilde{k}+\tilde{r}),\tilde{k}}. \tag{4.35}$$

Analyzing the range of variability of the subindices $i, j, r, k, \tilde{r}$ and $\tilde{k}$, minimizing β requires separating the translucent elements (band-pass filters) of the coded aperture, $R+L$ pixels apart horizontally, along all the L spectral bands. That is, this problem can be cast as the minimization of the products of the entries of the coded aperture within a window of size $2(R+L) \times L$ along the $(x-\lambda)$ axis, centered at each pixel (i, j, k), for every ℓ^{th} snapshot being captured. In addition, even though we do not consider the basis vectors $\boldsymbol{\psi}$ in the minimization of β, its a-priori selection gives us clues about its structure. Common selections of this basis for CSI include L-long Kronecker products between Wavelets and Cosines; thus, high correlations in the basis will occur every L steps. Therefore, we aim to spread the band-pass elements of each color coded aperture in every of the sub-windows of size $2L \times 2(R+L) \times L$ along all its rows, columns and slices.

In order to avoid reaching the trivial solution (all-zero entries), the color coded aperture entries have to satisfy the minimum light throughput constraint $\sum_{\ell=0}^{K-1} \mathbf{t}_k^\ell = 1$, for $k = 0, \ldots, L-1$, and $K \geq 2$. That is, each voxel of the data cube must be sensed at least and only once. Consequently, for a single snapshot sensing ($K = 1$), the light throughput will be 50%. Formally, this can be written as the optimization problem

$$\begin{aligned} &\underset{\{\mathbf{t}_0^\ell, \ldots, \mathbf{t}_{L-1}^\ell\}_{\ell=0}^{K-1}}{\operatorname{argmin}} && \beta \\ &\text{subject to} && \sum_{\ell=0}^{K-1} \mathbf{t}_k^\ell = 1, \text{for } k = 0, \ldots, L-1. \end{aligned} \tag{4.36}$$

4.4.1 Optimization Algorithm

To solve the optimization problem proposed in Eq. (4.36), we developed an iterative algorithm which randomly walks along all the pixels of a set of randomly generated color coded apertures satisfying the complementarity constraint. In every step, the algorithm filters the subwindow of size $2L \times 2(R+L) \times L$ centered at the current pixel with an specially designed 3D-Euclidean filter, and evaluates the objective function of the problem for every ℓ^{th} snapshot. The snapshot pixel that gives the lowest objective value is set to 1 (band-pass pixel), and the rest of the snapshots at the same position are set to 0 (blocking pixel). This way, the algorithm satisfies the complementary constraint at the same time that the objective function is minimized. The algorithm iterates on the updated coded apertures until the objective function stops decreasing, or until β lies within a certain threshold.

Algorithm 1 details the steps followed by the optimization algorithm. It requires the size of the datacube $N \times N \times L$, the number of snapshots K, the higher-order dispersion pixels R and the 3D filter $\boldsymbol{\Xi}$. Lines 2-7 perform the initialization of the complementary coded apertures by using Algorithm 2 (see Appendix 4.8). Lines 8-19 detail the main loop of the algorithm. Line 9 returns the triplets of the random walk to be performed along the 3D coded aperture, detailed in Algorithm 3 (see Appendix 4.8). For each one of the triplets, a subwindow of size $2L \times 2(R+L) \times L$ centered at the triplet pixel is selected for every snapshot in Line 13, as performed in Algorithm 4 (see Appendix 4.8), and filtered in Line 14, where β is also evaluated. Line 16 returns the snapshot that gives the smaller β, and Line 17 assigns the band-pass element to that snapshot. Lines 8-19 are repeated until the cost function stop decreasing, or the maximum number of iterations is achieved.

4.4.2 Euclidean Filter Design

The filtering step is critical for the minimization of β, and even more critical the filter being used. We follow the intuition on the goal of spreading the translucent elements along a 3D cube of size $2L \times 2(R+L) \times L$, and so we decided to design, what we have called, a 3D spreading Euclidean filter. This filter is of size $2L-1 \times 2(R+L)-1 \times L-1$ and the weights are calculated inversely proportional to the Euclidean distance between the points in a 3D mesh, with respect to the central pixel. That is, the entries of the filter are given by

$$\Xi_{a,b,c} = \left(\sqrt{(a-a_0)^2 + (b-b_0)^2 + (c-c_0)^2}\right)^{-1}, \tag{4.37}$$

Algorithm 1 CCA Optimization Algorithm

procedure CCA-OPT(N, L, K, R, Ξ)
  for $k \leftarrow 0, L-1$ **do** ▷ Initialization of CCAs
    $\mathbf{T}^k \leftarrow \text{BooleanCCA}(N, K)$
    for $\ell \leftarrow 0, K-1$ **do**
      $\mathbf{t}_k^\ell \leftarrow \text{vec}\left(\mathbf{T}^k(:,:,\ell)\right)$
    end for
  end for
  repeat ▷ Main iterative loop
    $\mathbf{M} \leftarrow \text{3DMeshRandWalk}(N, L)$
    for each triplet $(i, j, k) \in \mathbf{M}$ **do**
      for $\ell \leftarrow 0, K-1$ **do**
        $t_{i,j,k}^\ell \leftarrow 0$
        $\hat{\mathbf{T}}^\ell \leftarrow \text{Window}(\mathbf{T}^\ell, (i,j,k), L, R)$
        $\beta^\ell \leftarrow \text{sum}(\text{vec}(\hat{\mathbf{T}}^\ell \circ \Xi))$
      end for
      $\hat{\ell} \leftarrow \text{MinIdx}(\beta^0, \ldots, \beta^{L-1})$
      $t_{i,j,k}^{\hat{\ell}} \leftarrow 1$
    end for
  until Stopping criteria is met (Iter, Tol, No change)
  return $\left\{\mathbf{t}_0^\ell, \ldots, \mathbf{t}_{L-1}^\ell\right\}_{\ell=0}^{K-1}$ ▷ K optimized CCAs
end procedure

for each triplet (a, b, c) within the filter. Note that $a_0 = b_0 = c_0 = 0$, when the origin of the filter is assumed to be the central pixel. In addition, the entries of the filter along the central column, central row, and central slice are set to the maximum value of the filter. This is done after realizing that these central coordinates contribute the most correlations to β. Figure 4.4 shows an sketch of the spreading filter.

4.5 Simulation Results

To evaluate the outcome of the proposed method, we perform a comparison against random, boolean, and the coded apertures resulting of the row-wise optimization proposed in [41]. We first evaluate the conditioning of the sensing matrix using each of these coded apertures, and then we compare their performance with actual sensing and reconstruction of 6 different spectral datacubes in terms of PSNR and SSIM.

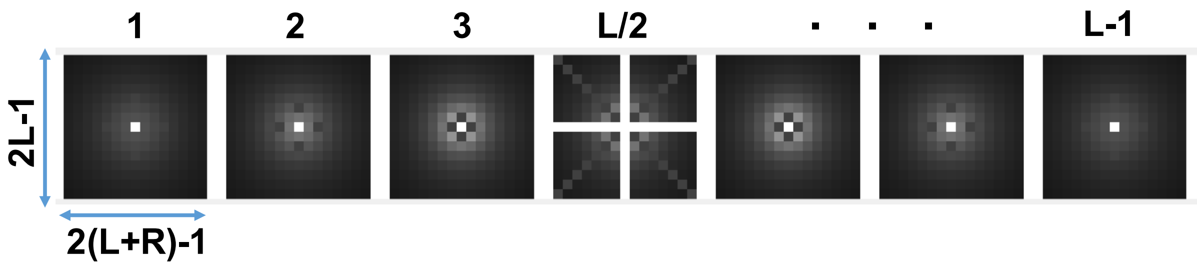


Figure 4.4: Sketch of the 3D spreading Euclidean filter Ξ.

4.5.1 Coded Aperture Optimization Comparison

A widely used method to compare different measurement strategies is the singular value spread analysis of the sensing matrices $\mathbf{A}$ from each method [67]. This analysis does not depend on prior information about the scene, but directly on the sensing matrix. In particular, it calculates the condition number (κ) of the matrix, defined as the ratio between the greatest singular value and the smallest nonzero singular value, to measure how ill-conditioned is the problem. The closer the condition number to 1, the better conditioned the matrix, being 1 the ideal scenario. Figure 4.5 shows the SVD curves of the 4 methods being compared for a color coded aperture with $N = 256$, $L = 8$, using $K = 2$ (Fig. 5(a)) and $K = 4$ (Fig. 5(b)) snapshots. It can be seen that the proposed method attain the best condition number for the first 1000 components, $\kappa = 1.0168$ (for $K = 2$), and $\kappa = 1.034$ (for $K = 4$), and a slower-decaying behavior of the SVD coefficients, which mean, that the sensing matrix is prone to capture more orthogonal components of the scene, and thus to attain a better recovery [67].

Subsections of the previously compared color-coded apertures, along with zoomed versions, are depicted in Fig. 4.6. The boolean realization is not shown because of its similarity to the random. Recall that a color-coded aperture can be regarded as a 3D cube of black-and-white elements; thus, a front view (along $x - y$ axis) and a top view (along $x - \lambda$ axis) are shown. There, it can be seen that the row-wise optimization exhibits a structured pseudo-random behavior due to its row-wise optimization nature, and the spread is done just spatially. In contrast, the random and our proposed method spread the non-zero values along the 3 dimensions. Note however, that the spread attained by

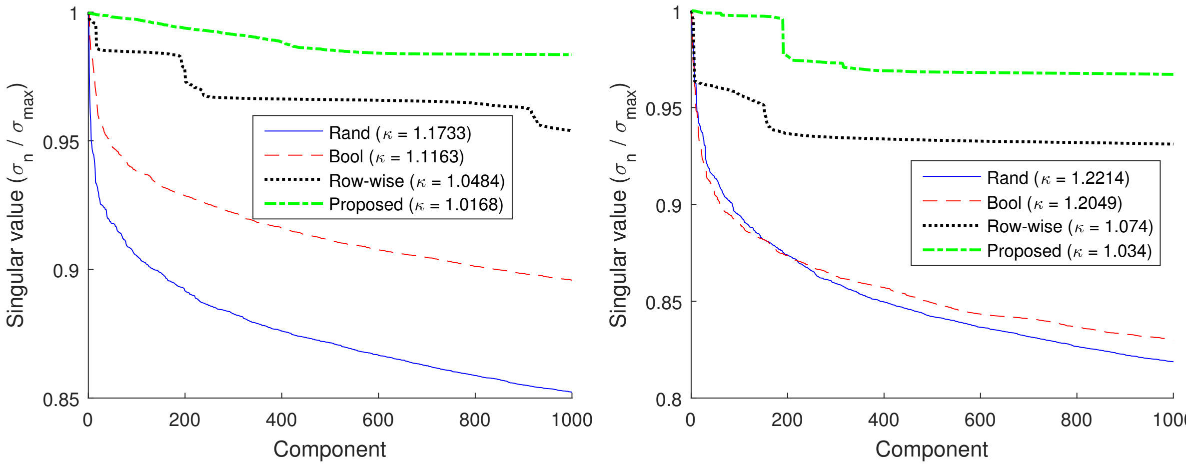


Figure 4.5: Analysis of the singular value spread of the sensing matrices $\mathbf{A}$ for the 3 different color coded apertures, with $N^2 = 256^2$ spatial pixels, $L = 8$ spectral bands, and (a)$K = 2$ snapshots, (b) $K = 4$ snapshots.

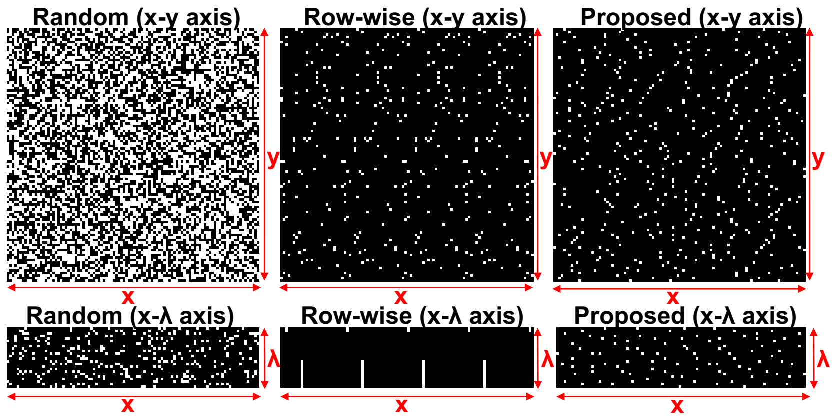
the proposed method is more uniform spatially and spectrally than the random, where clusters of translucent elements are present.

Table 4.1: PSNR and SSIM results for the 6 spectral images with $L = 8$

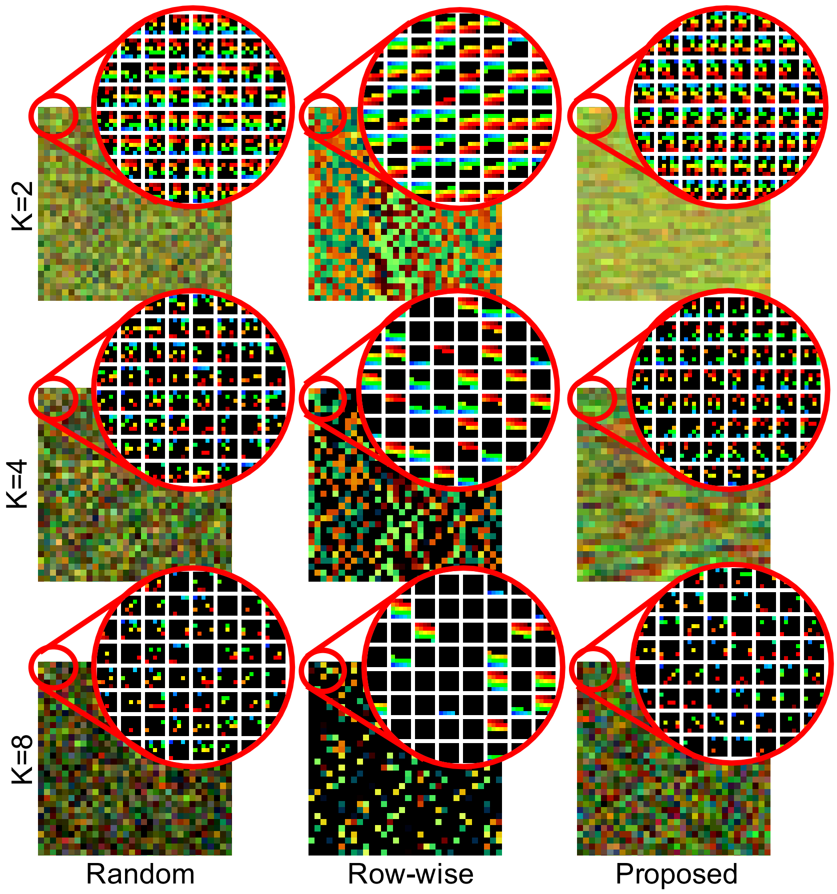
	Database	**Lego**		**Beads**		**Feathers**		**Flowers**		**Pompons**		**Thread-spools**	
Shots	**Method**	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
K=2	Rand	29.64	0.64	22.23	0.51	28.45	0.49	28.68	0.41	29.84	0.65	30.66	0.54
	Bool	30.08	0.65	22.38	0.52	28.75	0.50	28.79	0.42	30.50	0.66	31.07	0.55
	Row-wise	30.21	0.65	22.68	0.52	28.90	0.50	29.05	0.43	30.52	0.66	31.20	0.56
	Proposed	**30.37**	**0.65**	**23.25**	**0.53**	**29.25**	**0.51**	**29.49**	**0.43**	**30.60**	**0.66**	**31.55**	**0.57**
K=4	Rand	35.36	0.73	27.49	0.71	33.09	0.59	33.08	0.49	34.72	0.75	36.15	0.63
	Bool	37.71	0.79	28.10	0.74	34.60	0.68	34.96	0.58	37.71	0.82	38.21	0.70
	Row-wise	37.99	0.84	28.98	0.77	35.58	0.74	35.59	0.63	38.67	0.84	39.42	0.78
	Proposed	**40.24**	**0.85**	**29.81**	**0.79**	**36.41**	**0.76**	**37.09**	**0.64**	**39.73**	**0.86**	**40.52**	**0.80**
K=6	Rand	37.28	0.76	29.08	0.76	33.46	0.65	35.53	0.57	35.57	0.80	37.47	0.73
	Bool	41.10	0.84	31.19	0.82	37.87	0.76	38.54	0.65	40.85	0.88	42.31	0.81
	Row-wise	46.69	0.90	36.11	0.92	42.49	0.87	42.68	0.75	46.04	0.94	47.99	0.88
	Proposed	**48.37**	**0.91**	**36.03**	**0.92**	**42.61**	**0.88**	**43.70**	**0.75**	**46.21**	**0.94**	**48.31**	**0.89**
K=8	Rand	38.31	0.76	29.84	0.79	35.93	0.70	36.56	0.60	37.07	0.84	38.79	0.74
	Bool	44.35	0.88	33.28	0.87	40.10	0.83	40.91	0.71	42.73	0.91	44.57	0.85
	Row-wise	55.06	0.97	44.23	0.97	50.29	0.96	51.14	0.88	53.01	0.98	55.84	0.95
	Proposed	**57.60**	**0.97**	**45.65**	**0.98**	**52.21**	**0.98**	**53.79**	**0.91**	**55.76**	**0.99**	**58.79**	**0.96**

4.5.2 Testing the CCAs in the CSI system

To test the impact of the proposed coded aperture optimization algorithm in the quality of the reconstruction of spectral images, we evaluate the performance in 6 different spectral scenes with $N \times N = 256 \times 256$ pixels of spatial resolution, and $L = 8$ and $L = 24$ spectral bands equally spaced along the visible spectrum between 400 and 700 nm. We tested several scenarios using different amounts of snapshots $K = 2, 4, \ldots, L$, and four



(a) (First row) Front view $(x - y)$ axis. (Second row) Top view $(x - \lambda)$ axis.



(b) Each pixel is the average color of the filters in the zoomed version.

Figure 4.6: Comparison of color coded apertures with $L = 24$ and $N = 128$. (a) Front and top-view of the 3D black-and-white representation to easily identify the 3D spreading. (b) Single realization for $K = 2, 4$ and 8 snapshots.

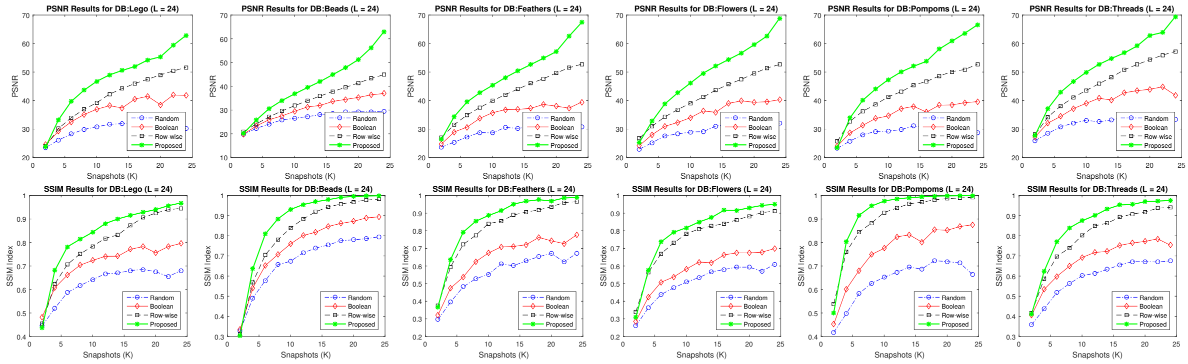


Figure 4.7: Comparison of the reconstructions in terms of PSNR (first row) and SSIM (second row) for different snapshots using the four kinds of coded apertures for the 6 databases with $L = 24$ bands.

types of coded apertures: Random, Boolean, Row-wise optimized and the Proposed. The Boolean coded apertures are similar to the Random ones, but they satisfy the complementary constraint in Eq. (4.36). To solve the CS inverse problem in Eq. (4.27), we employed the gradient projection for sparse reconstruction (GPSR) algorithm [29], using the Kronecker product between the 2D Wavelet Symlet 8 transform and the 1D discrete cosine transform, as the representation basis function $\mathbf{\Psi}$. This configuration and this algorithm have shown to be reliable and relatively fast to attain good reconstructions, although a large variety of solvers exist including, the constrained-split augmented lagrangian shrinkage algorithm (C-SALSA) [68], the two-step iterative shrinkage/thresholding (TwIST) [53], Gaussian mixture models (GMM) [54], and denosing algorithms based on approximate message passing (AMP) [33]. The penalization parameter τ in Eq. (4.27) was found by try and error within the range $\tau \in [2e-6, 3e-4]$, and the one that attained the maximum PSNR and SSIM was selected for each case. Ten realizations of this experiment were executed and the final results, described next, were averaged.

Table 4.1 and Fig. 4.7 summarize the reconstruction results in terms of the peak signal to noise ratio (PSNR) and the structural similarity index (SSIM), for the databases with $L = 8$ and $L = 24$, respectively. Remark that the larger the PSNR the better the reconstruction, and similarly, the closest the SSIM index to 1, the better the reconstruction. It can be noticed in this table and figure, that the proposed optimization algorithm generates better color coded apertures for the CSI system, as its PSNR and SSIM overcomes the ones attained with the row-wise optimization and the random/boolean coded apertures. It worth pointing out that the proposed algorithm perform better for larger number of snapshots since the number of non-zero values, per snapshot, decreases to satisfy the complementary constraint in Eq. (4.36), thus having more room to freely spread the non-zero values.

Figure 4.8 shows a comparison of the estimated spectral images when $K = 4$ snapshots are captured by the CSI system. There, it can be noticed the better quality attained by the proposed coded apertures over the other methods, for the 6 different databases. Two zoomed versions of each reconstruction are shown to facilitate the comparison. Finally, Figs. 4.9 - 4.11 shows 6 out of the 24 spectral bands of the zoomed versions for the Lego, Beads and Feathers databases, to evaluate the improvement at wavelength level. Note that the images in the down-most row of each subfigure preserves better the edges, and they are more artifact-free.

4.6 Conclusions

A higher order discrete sensing matrix of the optical phenomena in color-coded aperture based CSI systems has been reported. Also, an optimization algorithm exploiting this higher order discrete matrix was developed for the correct design of colored coded apertures by means of a 3D Eucledian spreading filter. The proposed method entail better conditioned sensing matrices, reflected in the condition numbers $\kappa = 1.0168$ (for $K = 2$), and $\kappa = 1.034$ (for $K = 4$), for the first 1000 components of the singular value spread. Furthermore, the proposed optimal color coded apertures overcomes the reconstruction quality attained with up-to-date proposed coded aperture optimization procedures. In particular, ours entail an improvement of up to 13 dBs in terms of PSNR against random/boolean distributions and up to 3 dBs against the row-wise optimization proposed in [41] for spectral scenes with $L = 8$ bands. When the spectral scenes exhibited more spectral bands ($L = 24$), the proposed coded apertures entailed an improvement of up to 20 dBs against random/boolean, and up to 7 dBs against the row-wise optimized.

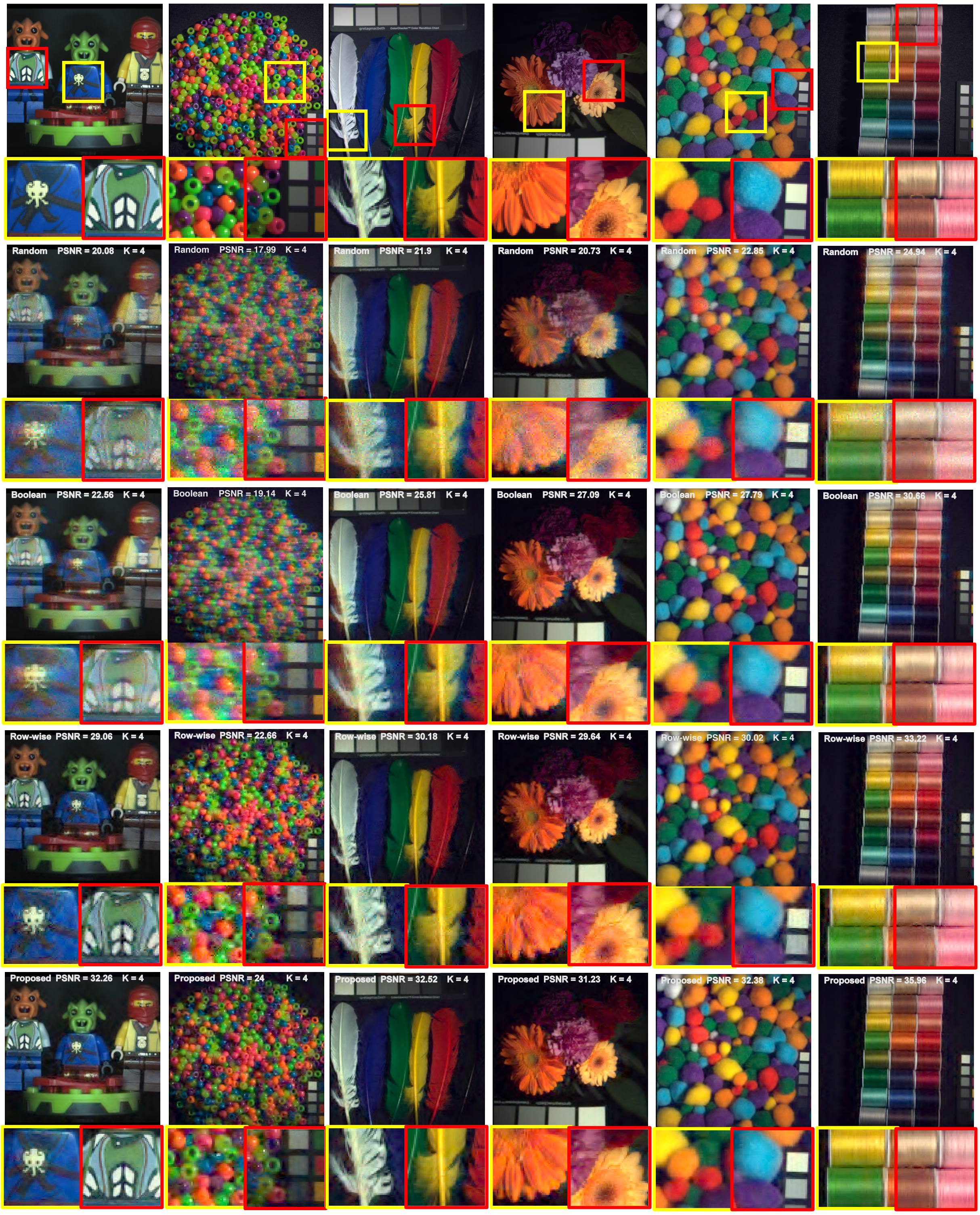


Figure 4.8: Reconstructed datacubes mapped to RGB profile when $K = 4$ snapshots ($\approx$ 16% of the data) are captured. From left to right there is shown the 6 databases: First, the Lego database, which was captured in our lab, and then, the other 5 (beads, feathers, flowers, pompoms and thread-spools), respectively, downloaded from [2]. The insets show 2 zoomed regions of each scene for easy comparison. From top to bottom we have the ground-truth image, and then the results attained with Random, Boolean, Row-wise and the Proposed color coded apertures.

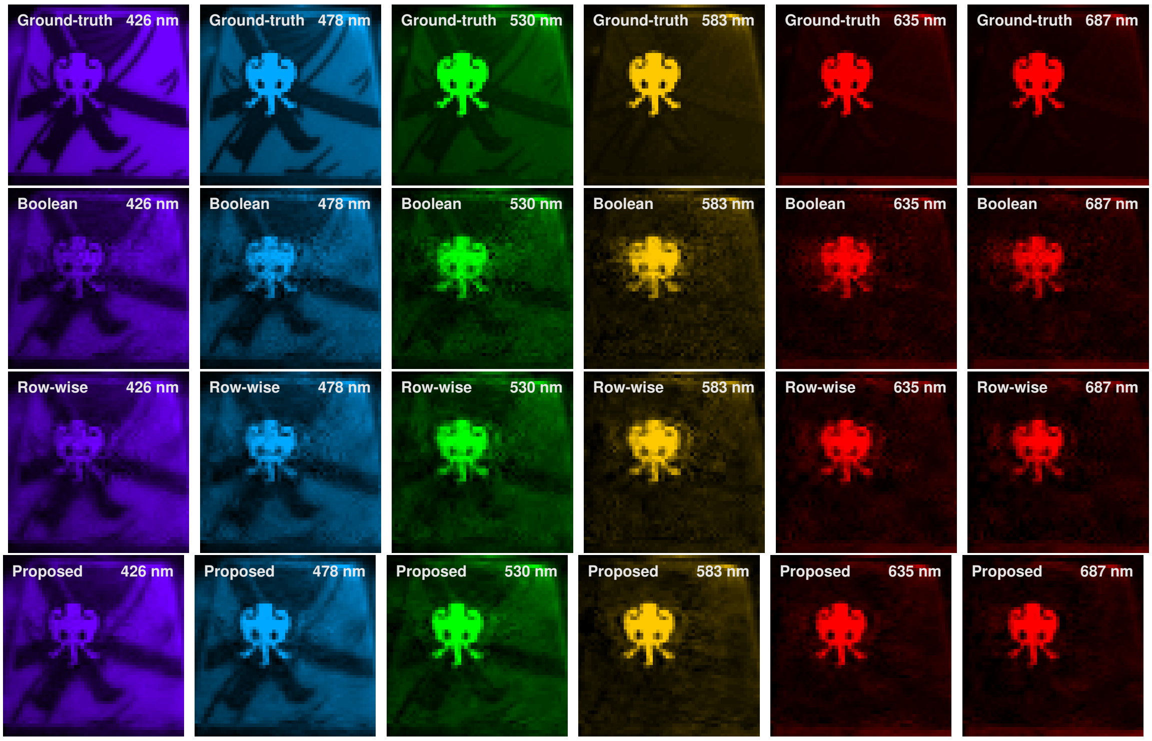


Figure 4.9: Zoomed version of the 3^{rd}, 7^{th}, 11^{th}, 15^{th}, 19^{th} and 23^{th} reconstructed spectral bands of the $L = 24$ Lego database, using $K = 4$ snapshots ($\approx 16\%$ of the data). (First row) Ground truth, (Second row) Boolean, (Third row) Row-wise optimization, and (Fourth row) Proposed optimization.

4.7 How to obtain Eq. (4.30)

The entries of the matrix $\mathbf{A} = \mathbf{H}\boldsymbol{\Psi}$ are given by,

$$
\begin{aligned}
(\mathbf{a}_{j'})_{i'} =& \sum_{q=0}^{Q-1} (\mathbf{h}_q)_{i'} \left(\boldsymbol{\psi}_{j'}\right)_q \\
=& \sum_{\{q|i'=g_r'\}} \left(\mathbf{w}^r_{\lfloor \frac{q}{N^2} \rfloor}\right)_{q-\lfloor \frac{q}{N^2} \rfloor N^2} \left(\mathbf{t}^{\lfloor \frac{i'}{V} \rfloor}_{\lfloor \frac{q}{N^2} \rfloor}\right)_{q-\lfloor \frac{q}{N^2} \rfloor N^2} \left(\boldsymbol{\psi}_{j'}\right)_q
\end{aligned}
$$

Recall Fig. 4.3. Note that each matrix $\mathbf{H}^\ell$ is an L-long horizontal concatenation of R vertically N-times shifted $N^2 \times N^2$ diagonal matrices with the coding pattern in each diagonal. Therefore we can perform an analysis of the entries of the $\mathbf{A}$ matrix per rows as follows.

1) *For* $0 \leq i' \leq N - 1$:

$$(\mathbf{a}_{j'})_{i'} = (\mathbf{h}_{i'})_{i'} (\boldsymbol{\psi}_{j'})_{i'}$$

2) *For* $N \leq i' \leq 2N - 1$:

$$(\mathbf{a}_{j'})_{i'} = (\mathbf{h}_{i'-N})_{i'} (\boldsymbol{\psi}_{j'})_{i'-N} + (\mathbf{h}_{i'})_{i'} (\boldsymbol{\psi}_{j'})_{i'} + (\mathbf{h}_{N^2+i'-N})_{i'} (\boldsymbol{\psi}_{j'})_{N^2+i'-N} \tag{4.38}$$

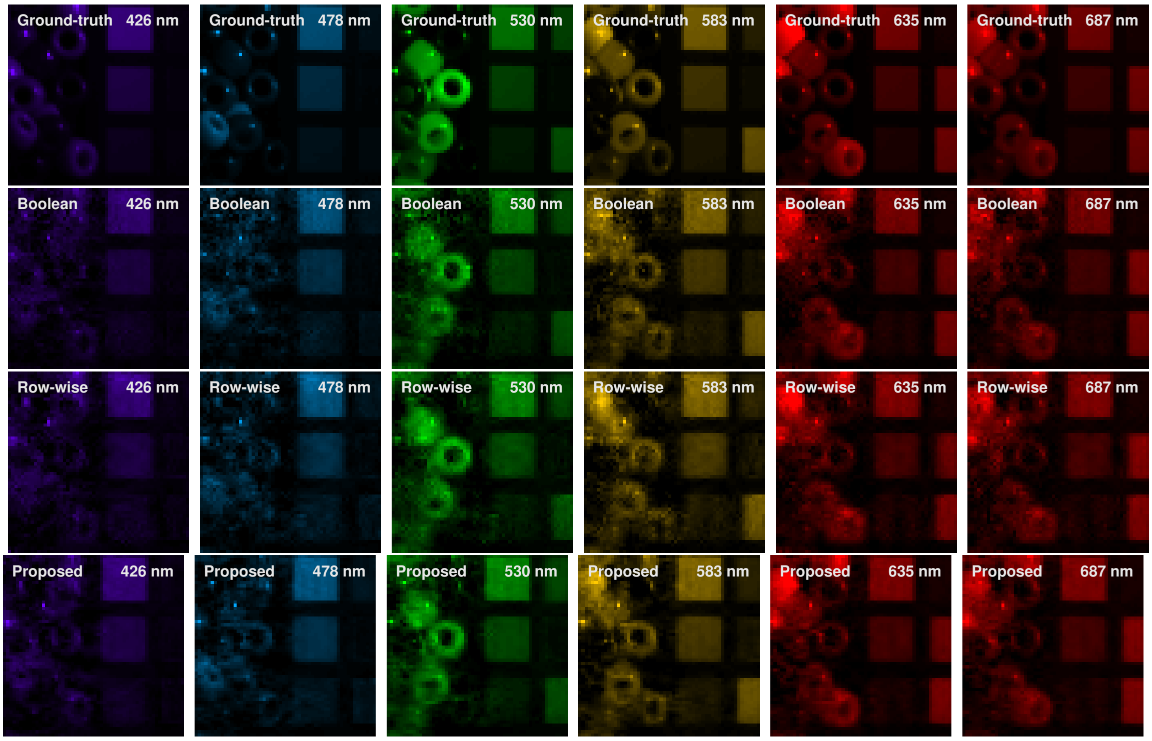


Figure 4.10: Zoomed version of the $3^{rd}, 7^{th}, 11^{th}, 15^{th}, 19^{th}$ and 23^{th} reconstructed spectral bands of the $L = 24$ Beads database, using $K = 4$ snapshots ($\approx 16\%$ of the data). (First row) Ground truth, (Second row) Boolean, (Third row) Row-wise optimization, and (Fourth row) Proposed optimization.

3) *For* $2N \leq i' \leq 3N - 1$:

$$\begin{aligned}(\mathbf{a}_{j'})_{i'} = \; & (\mathbf{h}_{i'-2N})_{i'}(\boldsymbol{\psi}_{j'})_{i'-2N} + (\mathbf{h}_{i'-N})_{i'}(\boldsymbol{\psi}_{j'})_{i'-N} \\ & +(\mathbf{h}_{i'})_{i'}(\boldsymbol{\psi}_{j'})_{i'} + (\mathbf{h}_{N^2+i'-2N})_{i'}(\boldsymbol{\psi}_{j'})_{N^2+i'-2N} \\ & +(\mathbf{h}_{N^2+i'-N})_{i'}(\boldsymbol{\psi}_{j'})_{N^2+i'-N}\end{aligned}$$

In general, from these 3 cases we can generalize $(\mathbf{a}_{j'})_{i'} =$

$$\sum_{r=0}^{R-1}\sum_{k=0}^{L-1}\left(\mathbf{h}_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V+kN^2}\right)_{i'}\left(\boldsymbol{\psi}_{j'}\right)_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V+kN^2}.$$

Replacing the entries of $\mathbf{h}$ given in Eq. (4.26), we get

$$(\mathbf{a}_{j'})_{i'} = \sum_{r=0}^{R-1}\sum_{k=0}^{L-1}(\mathbf{w}_k^r)_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V}\left(\mathbf{t}_k^{\lfloor \frac{i'}{V}\rfloor}\right)_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V}\left(\boldsymbol{\psi}_{j'}\right)_{i'-(k+r)N-\lfloor \frac{i'}{V}\rfloor V+kN^2}.$$

4.8 Algorithm subroutines

Algorithm 2 Subroutine BooleanCCA

```
procedure BooleanCCA(N, K)
    T ← 0 ∈ ℝ^{N×N×K}                                   ▷ All-zero initialization
    for each spatial position (i, j) do
        ℓ̂ ← rand(1, K)
        t_{i,j,ℓ̂} ← 1
    end for
    return T
end procedure
```

Algorithm 3 Subroutine 3DMeshRandWalk

```
procedure 3DMeshRandWalk(N, L)
    X ← randperm(N)
    Y ← randperm(N)
    Z ← randperm(L)
    for i ← 1, N do
        for j ← 1, N do
            for k ← 1, L do
                [M[i], M[j], M[k]] ← [X[i], Y[j], Z[k]]
            end for
        end for
    end for
    return M
end procedure
```

Algorithm 4 Subroutine Window

```
procedure Window(T^ℓ, (i, j, k), L, R)
    for i ← 1, 2L − 1 do
        for j ← 1, 2(L + R) − 1 do
            for k ← 1, L − 1 do
                t̂_{i,j,k} = t^ℓ_{(i−L),(j−L−R),(k−L/2)}
            end for
        end for
    end for
    return T̂
end procedure
```

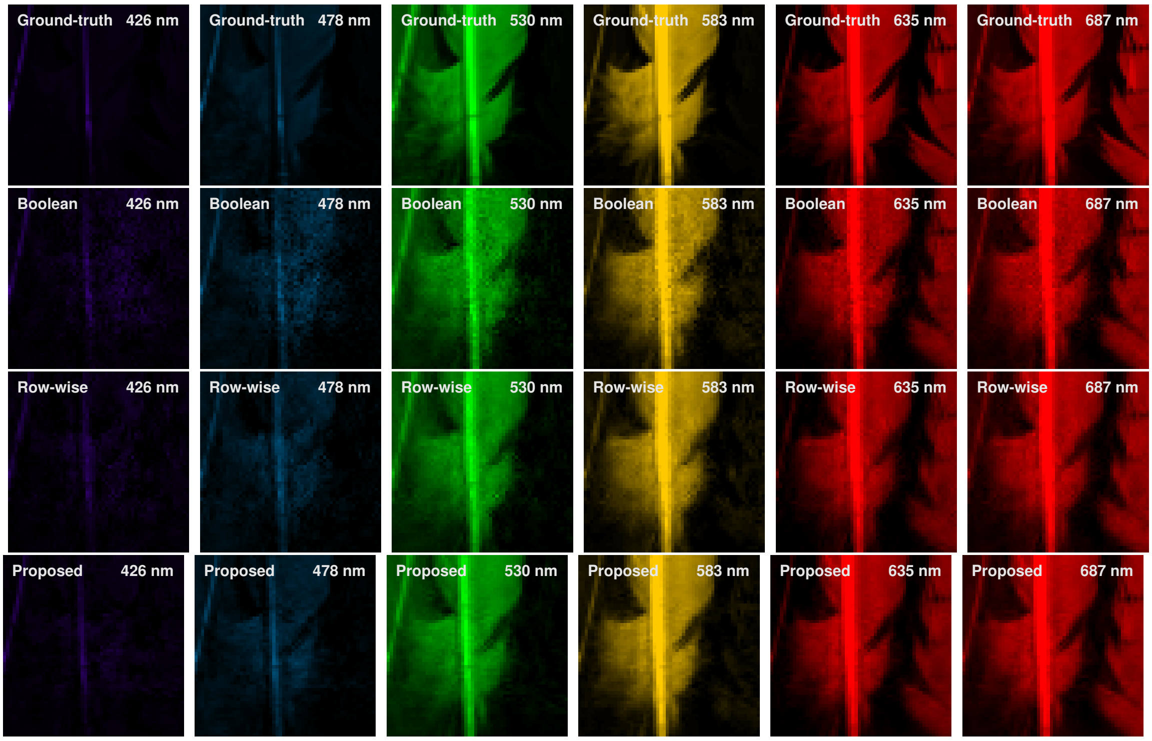


Figure 4.11: Zoomed version of the 3^{rd}, 7^{th}, 11^{th}, 15^{th}, 19^{th} and 23^{th} reconstructed spectral bands of the $L = 24$ Feathers database, using $K = 4$ snapshots ($\approx 16\%$ of the data). (First row) Ground truth, (Second row) Boolean, (Third row) Row-wise optimization, and (Fourth row) Proposed optimization.

Chapter 5

MULTI-SPECTRAL COMPRESSIVE SNAPSHOT IMAGING USING RGB IMAGE SENSORS

Compressive sensing is a powerful sensing and reconstruction framework for recovering high dimensional signals with only a handful of observations and for spectral imaging, compressive sensing offers a novel method of multispectral imaging. Specifically, the coded aperture snapshot spectral imager (CASSI) system has been demonstrated to produce multi-spectral data cubes color images from a single snapshot taken by a monochrome image sensor. In this chapter, we expand the theoretical framework of CASSI to include the spectral sensitivity of the image sensor pixels to account for color and then investigate the impact on image quality using either a traditional color image sensor that spatially multiplexes red, green, and blue light filters or a novel Foveon image sensor which stacks red, green, and blue pixels on top of one another.

5.1 Introduction

Multispectral imaging in the visible light spectrum refers to color cameras that record more than three primaries [69]. Composed of a monochrome image sensor taking a series of pictures through an array of narrowband interference filters, these cameras are especially common in the fine arts since full spectral recordings avoid issues associated with metamerism – the apparent change in color of an object caused by changes in ambient light. A more recent approach to multispectral imaging, the coded aperture snapshot spectral imager (CASSI) system [30], is modeled after a traditional spectrophotometer and records all wavelengths simultaneously. Here a ray of light emanating from the scene is passed through a series of optical elements culminating in a prism that spreads that single ray of light over a lateral sequence of sensor pixels, in a wavelength dependent manner, just like a spectrophotometer would spread light across a linear CCD array. Allowing for multiple rays of light simultaneously incident upon the sensor means that the camera can resolve a two dimensional image; however, because the spreading of light rays across lateral sequences of pixels means a single pixel will collect light from multiple sources, compressive sensing techniques must be employed to decouple the spectral profiles of neighboring pixels [23, 42].

Perhaps simplifying the problem somewhat, color sensors may be employed such that each pixel will only record those wavelengths of light falling within the range of its corresponding color filter, either red, green, or blue, but such sensors lose spatial resolution since only one quarter of the pixels are red, one quarter of the pixels blue, and the rest green [70]. Noting that photons of light penetrate the silicon in a wavelength dependent manner before being absorbed by the crystal lattice, Foveon manufactures a color image sensor that effectively layers red, green, and blue pixels on top of one another such that a 1 megapixel sensor records 1 million unique red, 1 million unique blue, and 1 million unique green values [71, 72]. Such a stacked color image sensor, like Foveon's, could provide a significant performance improvement for spectral recordings using CASSI, and in this chapter, we investigate that improvement in both theoretical and experimental venues. Specifically, we will introduce a new sensing model that incorporates the spectral response of the image sensor, we will present and analyze its forward sensing operator, and we will show simulated and experimental data to validate the findings. In particular, the Foveon-based system attains around a 4 dB improvement over the traditional monochrome system, and the color CCDs that employs a Bayer filter array attains up to 3 dB improvement, regardless of the number of snapshots.

5.2 Camera model

In the CASSI system, the front end lens system effectively creates a light field in front of the image sensor where all light rays travel in parallel to the Z-axis as depicted in Fig. 5.1(a) such that $f(x, y, \lambda)$ represents the light incident upon the sensor at 2-D coordinate (x, y) and having wavelength λ. Accumulating all of the light incident upon a discrete pixel of the sensor with row and column coordinate $[m, n]$ results in the discrete-space, continuous-frequency signal $f_{m,n}(\lambda)$. If we now define $\chi_{m,n}(\lambda)$ as the spectral sensitivity of the sensor for that same discrete pixel, then we can define the gray or

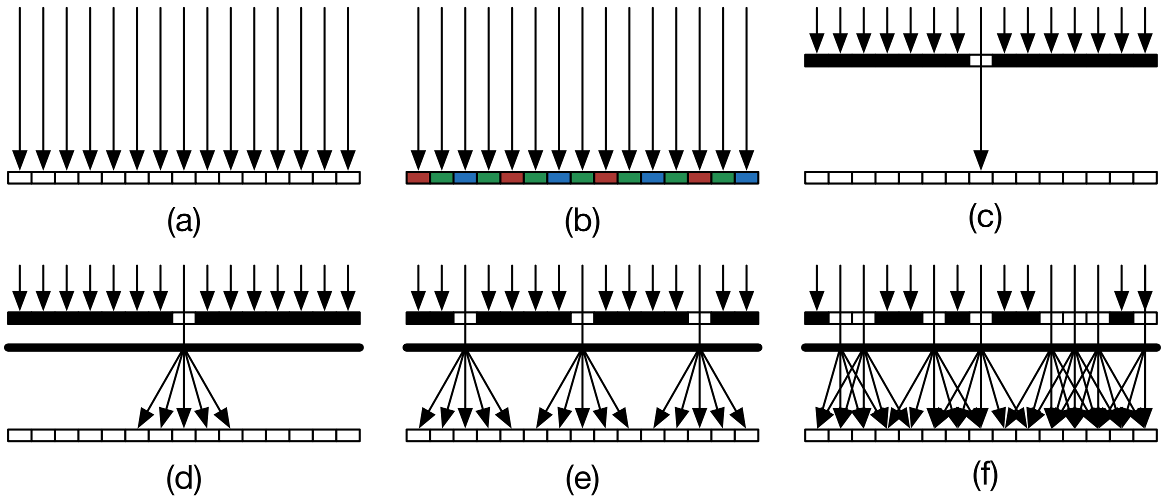


Figure 5.1: Illustration of the multispectral imaging principles used by CASSI.

monochrome pixel, $g[m, n]$, of the resulting digital image according to

$$g[m,n] = \int_{\lambda} \chi_{m,n}(\lambda) f_{m,n}(\lambda) d\lambda. \tag{5.1}$$

One would expect $\chi_{m,n}(\lambda)$ to be either constant for all pixels and absorb light in the visible light range from 450 to 670 nm or selectively absorb red, green, or blue light in a traditional Bayer pattern as illustrated in Fig. 5.1(b).

As light approaches the sensor, CASSI employs a coded aperture or diaphragm to block all but a single ray of light from reaching the sensor as indicated in Fig. 5.1(c). We can account for this aperture in our model by means of the binary modulation function $T_{m,n} \in [0, 1]$. In order to separate the individual wavelengths of the incoming signal, the CASSI system places a dispersive element in front of the image sensor such that $f_{m,n}(\lambda)$ is divided into discrete bands spread across K consecutive pixels along the lateral or n dimension in a wavelength dependent manner, indicated by $S(\lambda)$. We depict this discrete-space behavior in Fig. 5.1(d) such that,

$$g[m,n] = \sum_{k} \chi_{m,n,k} T_{m,n+k} f_{m,n+k,k}, \tag{5.2}$$

where $k = \frac{-K}{2}, \ldots, 0, \ldots, \frac{+K}{2}$ indexes the pixels affected by the spreading of $f_{m,n}(\lambda)$ as a result of $S(\lambda)$. At this point, we note, as indicated in Fig. 5.1(e), that we could use the coded aperture to allow light rays sufficiently far apart so that there is no overlap of the corresponding light spectrums; however, doing so means only a fraction of the incoming light field can be recorded. We could, therefore, take a series of snapshots where the open pixels in the coded aperture sweep across the focal plane such that, in this case, we would need to take exactly K snapshots in order to capture the K available spectral bands. Compressive sensing (CS) [73, 17, 27, 74] tells us that, perhaps, we may take a much smaller set of L snapshots and yet produce a spectral image very close to that produced by the K snapshots by randomly enabling/disabling pixels through the coded aperture as depicted in Fig. 5.1(f). Moreover, L can be further reduced when the distribution of the enabled pixels is optimized [36].

5.3 Compressive sensing model

In order to establish a matrix representation of CASSI consistent with CS, $\mathbf{f}$ is defined as a column vector containing $f_{m,n,k}$ for all spectral components k and all pixels m and n ordered first by row, then column, and finally by wavelength. Assuming we have a coded aperture with $M \times N$ pixels, $\mathbf{f}$ will be MNK in length. We now define $\mathbf{g}^\ell$ as a column vector holding all of the recorded pixel values, $g[m,n]$, from the monochrome image sensor for the ℓ^{th} image in our sequence of L images. Noting the laterally dispersion of light, the length of $\mathbf{g}^\ell$ is $M(N+K-1)$ such that all light is accounted for on the sensor. So while the coded aperture has an $M \times N$ pixel array, the sensor has an extra set of K pixel columns. Assuming that $\mathbf{f}$ stays constant over the L snapshots, we can relate $\mathbf{g}^\ell$ to $\mathbf{f}$ according to,

$$\mathbf{g}^\ell = \mathbf{X}\mathbf{P}\mathbf{T}^\ell\mathbf{f} = \mathbf{H}^\ell\mathbf{f}, \tag{5.3}$$

where $\mathbf{H}^\ell$ represents the combined effects of the sensor's pixel-wise spectral sensitivity in matrix form ($\mathbf{X}$), the dispersion effect in matrix form ($\mathbf{P}$), and the ℓ^{th} coded aperture ($\mathbf{T}^\ell$).

As an illustration of the system transfer function, Fig. 5.2 depicts the non-zero entries for single $\mathbf{H}^\ell$ for an $M = N = 4$ camera sensor used to record a $K = 6$ multi-spectral image where the matrix width $Q = MNK$ is the total number of scalar measurements

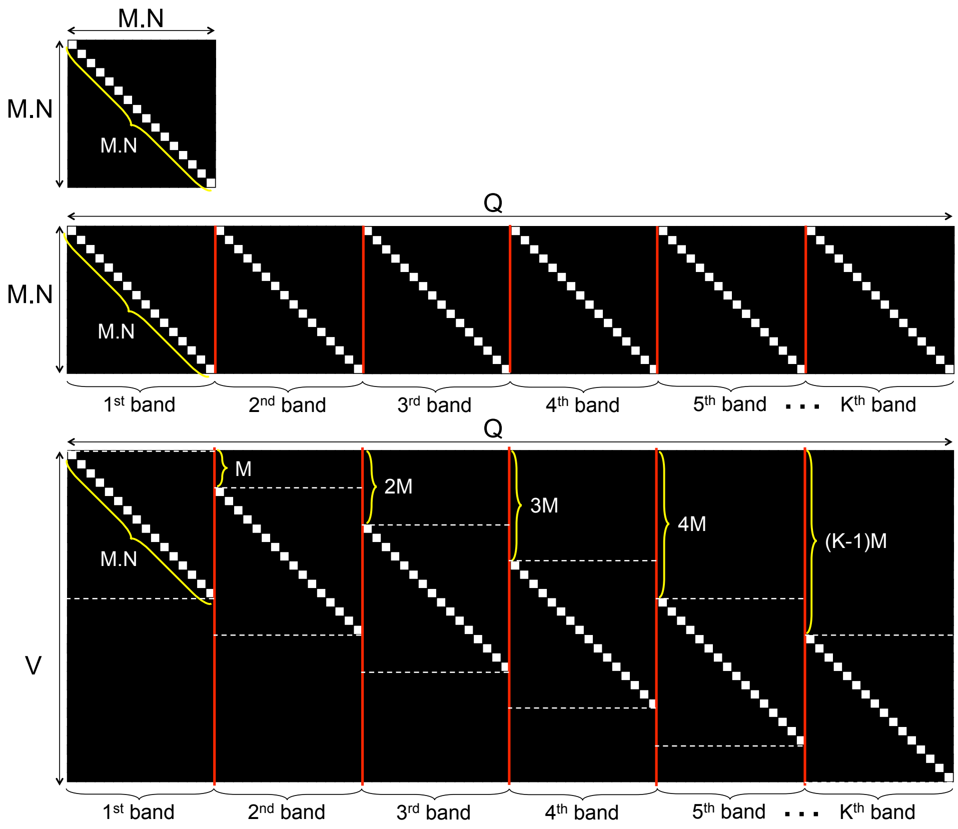


Figure 5.2: Structure of the system transfer functions $\mathbf{H}$ for a $4 \times 4 \times 6$ data cube, using a monochromatic sensor, where (top) shows the transfer function for a typical monochrome camera, (center) shows the same camera but with the incoming light divided into spectral bins, and (bottom) models the effects of the dispersive element by offsetting the diagonal structure from band to band. Note that the non-zero diagonals correspond to the entries of the $M \times N$ coded aperture set to be all ones.

(K values per pixel) while $V = M(N + K - 1)$ is the total number of observed pixels. In the case of Fig. 5.2 (top), the system transfer function is for a typical camera where none of the incoming pixels are blocked by the coded aperture nor is there any dispersion. Hence, there is no need to decompose the incoming light into spectral bins. For Fig. 5.2 (center), the system transfer function divides the incoming light into K spectral bins, but without any dispersion, the resulting observation $\mathbf{g}^\ell$ is the same as Fig. 5.2 (top). In Fig. 5.2 (bottom), the depicted transfer function models dispersion by off-setting the diagonal structure of the previous matrices vertically.

Because we have arranged vector $\mathbf{f}$ in row-major order, a lateral shift of light one pixel to the right on the sensor corresponds to a vertical shift of M pixels. As we progress through all K pixels on the sensor, we end up with a transform matrix that has $M(K-1)$ additional rows compared to Fig. 5.2 (center). Now having modeled the collection of incoming light rays on a per frame basis, we can now assemble the complete set of all L snapshots into a single transform by concatenating each $\mathbf{g}^\ell$ vector end to end to create the $LM(N+K-1)\times 1$ vector $\mathbf{g} = \left[(\mathbf{g}^0)^T, \ldots, (\mathbf{g}^{L-1})^T\right]^T$, which can be modeled in Eq. (5.3) by similarly concatenating the matrices $\mathbf{H}^\ell$ on top of one another to form the single matrix $\mathbf{H} = \left[(\mathbf{H}^0)^T, \ldots, (\mathbf{H}^{L-1})^T\right]^T$ and the transformation $\mathbf{g} = \mathbf{Hf}$.

Having now modeled the spectral imaging system as a single vector to single vector transform, CS requires us to find a linear transform of $\mathbf{f}$ from its current spatial-spectral (m, n, k) space into some other MNK-dimensional space with basis $\boldsymbol{\Psi}$ such that $\boldsymbol{\theta} = \boldsymbol{\Psi}^T\mathbf{f}$ is the vector coordinate of $\mathbf{f}$ in the new space. For this process to be worth-while, $\mathbf{f}$ needs to be S-sparse where only a small subset, $S \ll (M \cdot N \cdot K)$, of the basis vectors $\boldsymbol{\Psi}$ can largely reconstruct $\mathbf{f}$ with little or no distortion [24]. Formally, $\mathbf{f}$ is S-sparse or has sparsity S in a basis $\boldsymbol{\Psi}$ if $\|\boldsymbol{\theta}\|_0 = S$, where $\|\boldsymbol{\theta}\|_0$ denotes the ℓ_0-norm, which simply counts the number of nonzero entries in the vector.

Traditional basis functions encompass the wavelet ($\boldsymbol{\Psi}_{W2D}$) and discrete cosine transforms ($\boldsymbol{\Psi}_{DCT}$) as well as pre-trained dictionaries [36, 43, 75], but it is also possible to simultaneously exploit the sparsity properties of a multidimensional signal along each of its dimensions to provide a new representation for their structure. Doing so, we obtain a single sparsity basis for the entire multidimensional signal as the Kronecker product of the bases used for each of its dimensions [56, 45, 52]. Particularly for spectral images, we will use the basis expressed as the Kronecker product between $\boldsymbol{\Psi}_{W2D}$ and $\boldsymbol{\Psi}_{DCT}$ where $\boldsymbol{\Psi}_{DCT}$ is used to make sparse the k-axis while $\boldsymbol{\Psi}_{W2D}$ makes sparse the (m, n) coordinates.

Having our $\boldsymbol{\Psi}$, CS allows $\mathbf{f}$ to be recovered from a single $\mathbf{g}^\ell$ vector since the length of $\mathbf{g}^\ell$, V, is greater than $S\log(M \cdot N \cdot K)$ which is much smaller than $M \cdot N \cdot K$ [36]. Having multiple $\mathbf{g}^\ell$ will, therefore, only improve the result since it corresponds to increasing the dimensionality, S, of the sub-space, in which we are projecting $\mathbf{f}$. To finally estimate $\mathbf{f}$, we need to solve the minimization problem,

$$\tilde{\mathbf{f}} = \boldsymbol{\Psi}^T \left(\underset{\boldsymbol{\theta}'}{\text{argmin}} ||\mathbf{g} - \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}'||_2 + \tau ||\boldsymbol{\theta}'||_1 \right), \tag{5.4}$$

where $\boldsymbol{\theta}'$ is an S-sparse representation of $\mathbf{f}$ on the basis $\boldsymbol{\Psi}$, and τ is a regularization constant. To solve Eq. (5.4), different methodologies and frameworks have been proposed in the literature [53, 68, 29]. In particular, the gradient projection for sparse reconstruction (GPSR) algorithm has been proposed in [29]. GPSR finds a sparse solution to the non-linear, unconstrained minimization problem, where the first term minimizes the euclidean or ℓ_2 distance between the detector measurements, $\mathbf{g}$, and the contribution from the estimate $\boldsymbol{\theta}'$; while the second term is a penalty term that encourages sparsity of the reconstruction in the basis domain and controls the extent to which piecewise smooth estimates are favored. In this formulation, τ is a tuning parameter for the penalty term and higher values yield sparser estimates. An estimate for the data cube, $\tilde{\mathbf{f}}$, for a chosen value of τ, is found by an iterative optimization procedure. This method searches for a data cube estimate with a sparse representation in the chosen basis, i.e. the coefficients in $\boldsymbol{\theta}'$ are mostly zeros numerically.

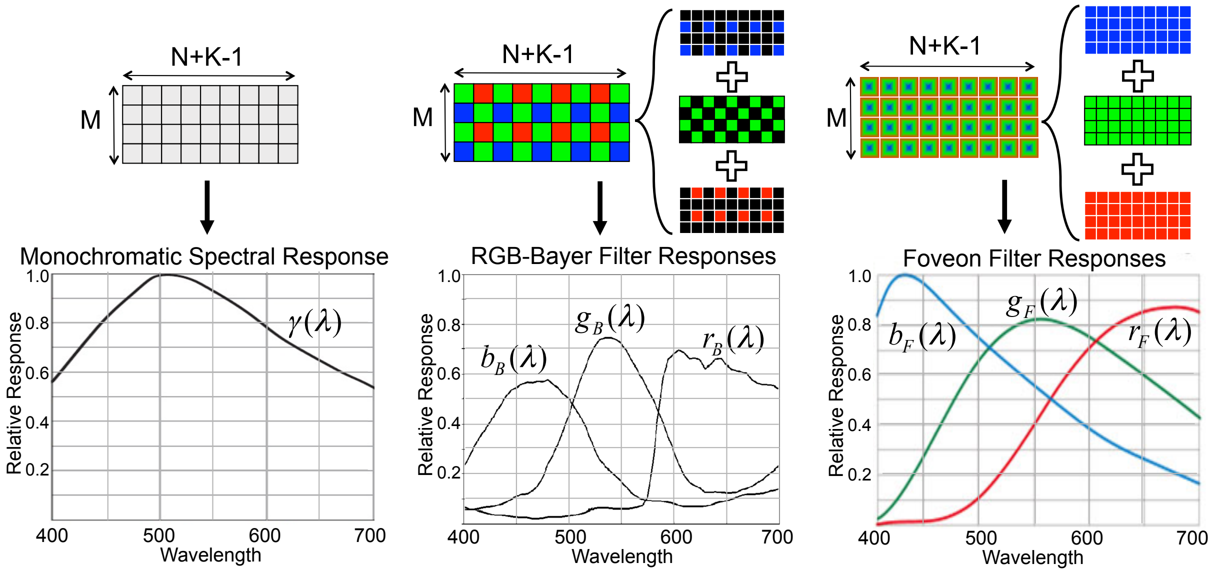


Figure 5.3: FPA detector architectures along with their traditional spectral responses. (left) Monochromatic (Source: Stingray F-033B), (center) RGB-Bayer (Source: TMC) and (right) Foveon (Source: Foveon Quattro sensor).

5.4 Color sensing model

As a means of extending the CS reconstruction to allow for color sensitive sensors, we note that all prior works assumed a monochrome sensor model [30] where $\chi_{m,n}(\lambda) = \gamma(\lambda)$ was constant for all $(m,n)^{th}$ pixels, as depicted in Fig. 5.3(left). In this chapter, we intend to employ two color sensitive sensors: a traditional filter-based RGB Bayer sensor [70] shown in Fig. 5.3(center), and an absorption-based stacked color sensor developed by Foveon [71], presented in Fig. 5.3(right). Color sensors are employed such that each sensor pixel records those wavelengths falling within small spectral ranges corresponding to either the blue, green or red spectra. Figure 5.3 summarizes the 3 different sensors along with their structure and spectral sensitivity curve in the visible window of the electromagnetic spectrum.

Formally, each (m,n) pixel from an RGB-Bayer, or Foveon sensor exhibits a filter function given by,

$$\text{Bayer: } \chi_{n,m}(\lambda) = \begin{cases} b_B(\lambda), & \forall n,m \in \Omega_b \\ g_B(\lambda), & \forall n,m \in \Omega_g \\ r_B(\lambda), & \forall n,m \in \Omega_r, \end{cases} , \text{Foveon: } \chi_{n,m}(\lambda) = \begin{cases} b_F(\lambda), & \forall n,m \\ g_F(\lambda), & \forall n,m \\ r_F(\lambda), & \forall n,m, \end{cases} \tag{5.5}$$

where Ω_b, Ω_g and Ω_r represent the sets of blue, green and red pixels of the Bayer color pattern. Notice that the Foveon sensor entails the absorption of the total light incident in every pixel, whereas the Bayer sensor filters out wavelengths lying outside the spectral-sensitivity of the pixel. Remark that the relative response of the three layers of the Foveon sensor $b_F(\lambda), g_F(\lambda), r_F(\lambda)$ exhibit a broader sensitivity compared to the corresponding RGB filters $b_B(\lambda), g_B(\lambda), r_B(\lambda)$ from the Bayer sensor. Due to the blue on-top-of green on-top-of red layer stack of the Foveon sensor, the blue layer absorbs part of the green and red spectrum, and the green layer will absorb part of the red spectrum as well.

In order to modify our transform functions $\mathbf{H}^{\ell}$, to include variations in $\mathbf{X}$, we will essentially stack three monochrome transforms, as depicted in Fig. 5.4, on top of one another with the first transform weighted for blue, the second weighted for green, and the third weighted for red, according to $b_F(\lambda), g_F(\lambda), r_F(\lambda)$, for the Foveon, and $b_B(\lambda), g_B(\lambda), r_B(\lambda)$ for the Bayer sensor. Due to the stacking process, and in order to be fair in the comparison between the color sensors and the monochrome sensor, 3-monochrome snapshots will be compared against 1-color snapshot. Therefore the compression ratio of the system will be $K : 3L$. Fig. 5.4 shows our proposed color transform models for both Fig 5.4 (left) a spatially multiplexed Bayer sensor, versus Fig 5.4 (right) the stacked-color Foveon sensor. It can be observed that both matrices have the same size, but the Bayer sensor entails completely blanked rows due to the black holes on each Bayer layer compared to the Foveon sensor, which absorbs all the blue, green, and red spectrum in every (m, n) sensor pixel. Further, the Bayer pixel sensitivities can be seen as, $b_B(\lambda)$ filters out the last portion (4 out of 6 bands) of the spectrum, $g_B(\lambda)$ filters out the first (1 out of 6 bands) and the last portion (1 out of 6 bands) of the spectrum, and $r_B(\lambda)$ filters out the first portion (4 out of 6 bands) of the spectrum. Similarly, the Foveon pixel absorption is modeled as, $b_F(\lambda)$ absorbs the first 3/4 of the spectrum in average, due to the broader spectral sensitivity, $g_F(\lambda)$ absorbs the central portion of the spectrum and part of the last portion, and $r_F(\lambda)$ just the last portion of the spectrum. Aside from this change in $\mathbf{H}^\ell$, deriving $\mathbf{f}$ from $\mathbf{g}$ is largely unchanged from the CS process used for monochrome sensors.

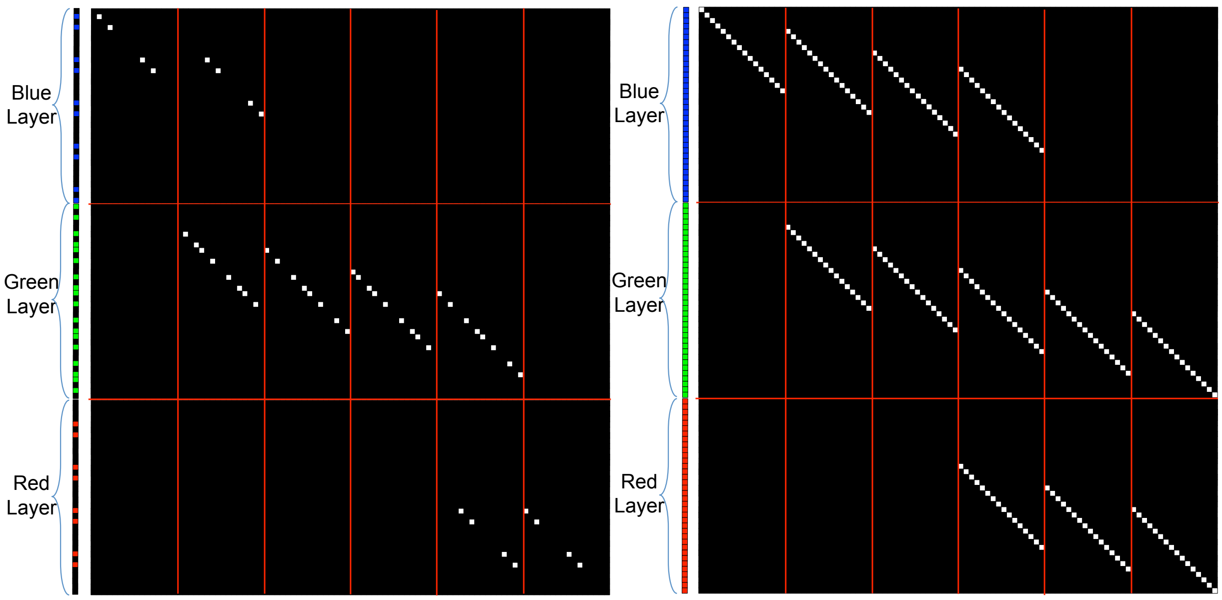


Figure 5.4: Non-zero elements of the system transfer functions, $\mathbf{H}^\ell$, for a $4 \times 4 \times 6$ data cube, using (left) RGB-Bayer and (right) Foveon sensors.

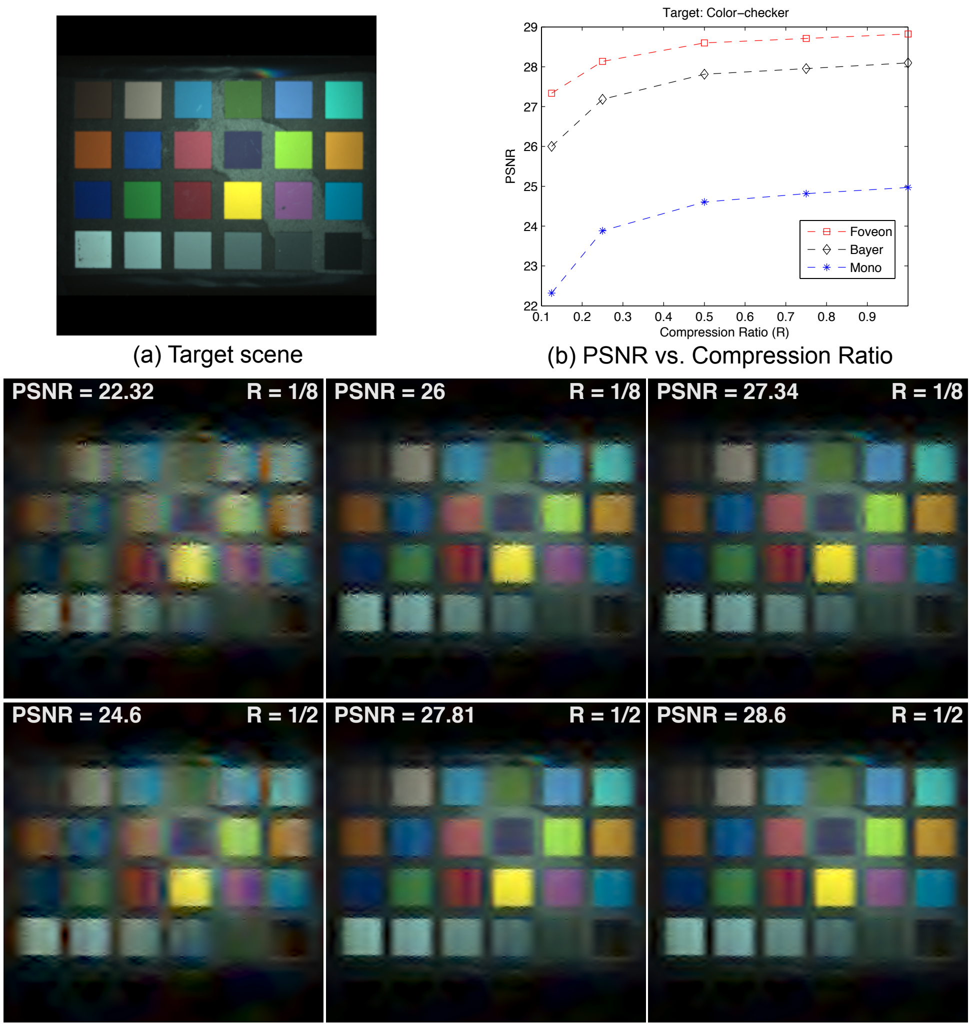
5.5 Simulations and experiments

In this section, we will first evaluate the proposed method through the simulation of the process using the two color-checker patterns shown in Fig. 5.5(a) and Fig. 5.6(a), using discrete versions of the FPA spectral responses described in Fig. 5.3. Afterward, we will collect real data with the DMD-based CASSI imaging system built in our lab [35] using the input target depicted in Fig.5.10(a), and commercially available RGB color filters.

5.5.1 Simulations

In order to study the effect of the different filtering functions entailed by the color sensors under study, a set of compressive measurements is first simulated using the forward model in Eq. (5.3), varying the $\mathbf{X}$ function for a monochrome, an RGB Bayer, and a Foveon sensor according to Fig. 5.3. These measurements are constructed employing two test spectral data cubes acquired by illuminating the color-checker target shown in Fig. 5.5(a) using a broadband Xenon lamp as the light source, and a visible monochromator spanning the spectral range between 450nm and 670nm. The test data cubes, $\mathbf{F}$, have 256×256 pixels of spatial resolution and $K = 24$ spectral bands. Although the coded apertures are a key optical element in the compressive spectral imaging systems, and their design has a great impact on the attained reconstructions quality [30, 36, 52, 41], in this chapter we will use simple random realizations of a Bernoulli random variable with parameter $p = 0.5$; that is, 50% of the coded aperture entries will be 0 and the rest will be 1. The latter is justified because this chapter wants to focus on the impact of color sensors on the quality of the reconstructions, but independently of the effect of the coded apertures. The proper design of the coded apertures will improve even more the results presented in this chapter. Although the color sensors are simulated through the use of a monochrome sensor, they take into account the aliasing effects resulting of demosaicing in the Bayer sensor, and the real absorption/transmission curves of the Foveon sensor. Moreover, the three compressive spectral imaging systems compared in the manuscript are assumed to be ideal in terms of noise, due to the different sources producing it: exposure time, kind of sensor, photon count, quantization, heat, etc. Analysis of the noise characteristics of the sensors lies out of the scope of this chapter.

Figure 5.5(b) shows the reconstructions PSNR of the first test data cube, averaged over all wavebands, as function of the compression ratio ($R = L/K$). Notice that $R = 3/24 = 1/8$ represents a single snapshot for the color sensors, and 3 for the monochrome sensor. In Fig. 5.5(b) it can be seen that the PSNR attained in the reconstructions with the RGB-Bayer sensor improves up to 3dB over the monochrome sensor overall. Similarly, the Foveon sensor improves an extra dB over the RGB-Bayer. For a visual comparison, Fig 5.5(c) depicts the $K = 24$ reconstructed data cube by the three sensors mapped to a pseudo-color map when $R = 1/8$ and $R = 1/2$. The improvement in the spatial quality

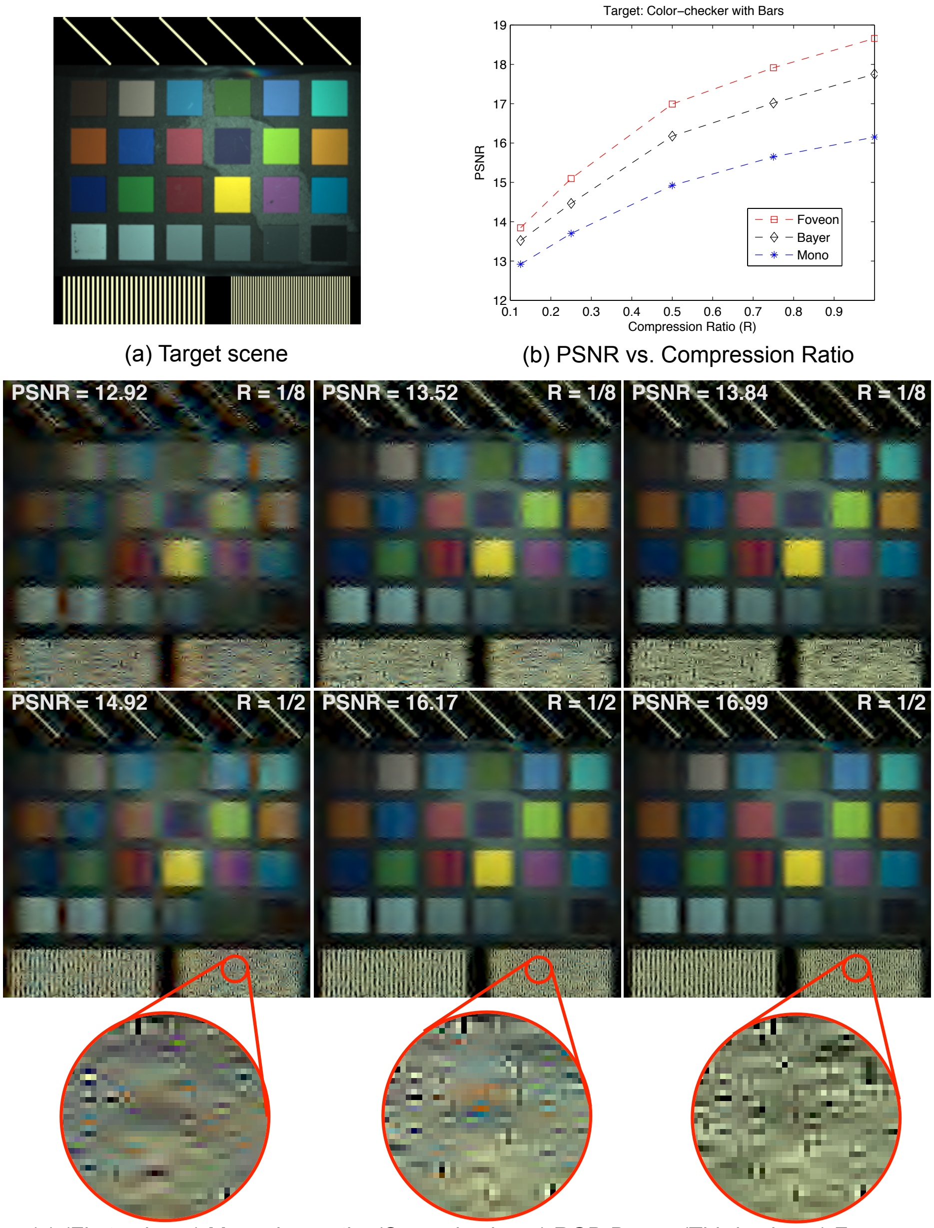


(a) Target scene

(b) PSNR vs. Compression Ratio

(c) (First column) Monochromatic, (Second column) RGB-Bayer, (Third column) Foveon.

Figure 5.5: Simulation results with first target scene. (a) Target scene for the simulations. (b) Averaged PSNR of the reconstructed data cubes as function of the compression ratio (R). (c) Reconstructed datacubes, using (first row) $R = 1/8$ and (second row) $R = 1/2$.



(a) Target scene

(b) PSNR vs. Compression Ratio

(c) (First column) Monochromatic, (Second column) RGB-Bayer, (Third column) Foveon.

Figure 5.6: Simulation results with second target scene. (a) Target scene for the simulations. (b) Averaged PSNR of the reconstructed data cubes as function of the compression ratio (R). (c) Reconstructed datacubes, using (first row) $R = 1/8$ and (second row) $R = 1/2$. Note the color aliasing in the zoomed version of the RGB-Bayer reconstruction.

can easily be observed as L increases (from top to bottom) or as the sensor changes from monochrome to Bayer to Foveon (from left to right).

The second test data cube shown in Fig. 5.6 consisted of a modified version of the first data cube, where white diagonal lines were added to the top region of the color-checker and equally spaced white bars were added to the bottom region. Notice that the left portion of the bottom bars are two pixels wide, whereas the right portion correspond to one pixel width bars. This test wanted to show the reliability of the reconstructions for high-frequency changes, as well as the impact of the demosaicing task perfomed by the Bayer reconstruction. Figure 5.6(b) presents the averaged PSNR of the reconstructions, where the Bayer sensor overcomes the monochrome sensor by about 1.5 dB, and the Foveon maintains an advantage of about 0.5 dB over the Bayer sensor reconstructions overall. Remark that the PSNR is dramatically affected (about 10 dB less) for the three sensors compared with Fig. 5.5(b), when the scene presents high-frequencies changes. This issue arises in most of the algorithms that solve Eq. (5.4). It can be noticed in Fig. 5.6(c) that the diagonals in the top of the scene along with the bars in the bottom left are recovered somehow, but the bars in the bottom right are barely reconstructed. The zoomed versions show large color aliasing artifacts in the reconstruction attained with Bayer sensor unlike Foveon sensor, which presents artifacts, but not color aliasing.

To analyze the reconstruction results at wavelength level, Fig. 5.7 presents the absolute error of the $4^{th}, 8^{th}, 12^{th}, 16^{th}, 20^{th}$ and 24^{th} reconstructed spectral wavebands when

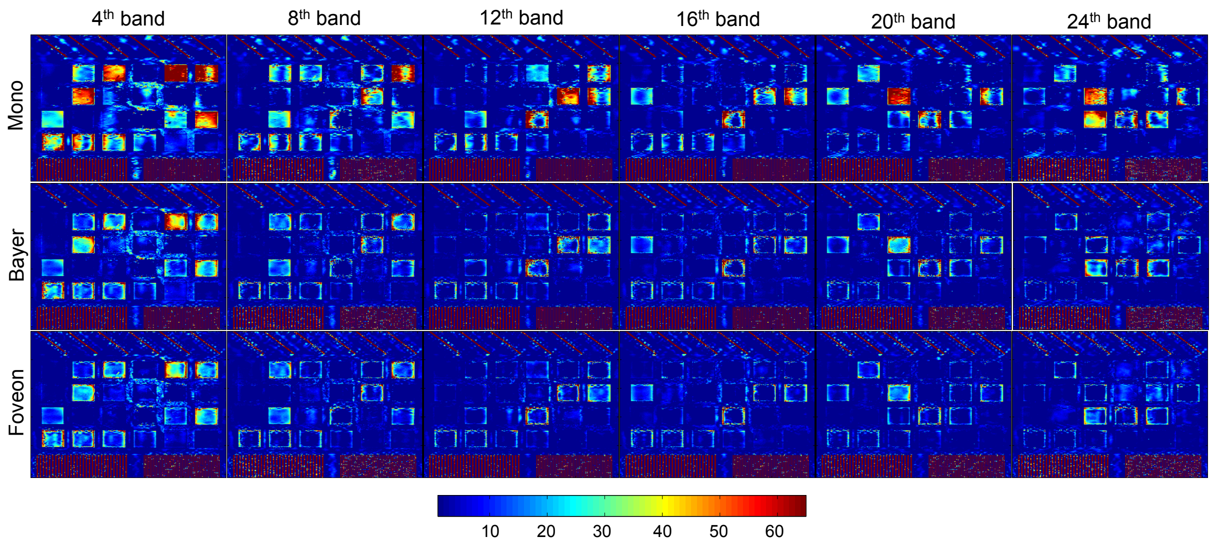


Figure 5.7: Absolute errors of the reconstructions of the $4^{th}, 8^{th}, 12^{th}, 16^{th}, 20^{th}, 24^{th}$ spectral bands, for $R = 1/8$, using (First row) Monochrome sensor, (Second row) Bayer sensor, (Third row) Foveon sensor. Notice that the Foveon sensor attains the best reconstructions between the three sensors, despite the high frequencies (borders) are almost unrecoverable.

a single snapshot is captured ($R = 1/8$). It can be observed that the error of the reconstructions attained with the color sensors is less than that attained with the monochrome sensor along the whole reconstructed spectrum. Remark that the major errors are concentrated along the high-frequencies of the scene, thus confirming the downside of the estimation algorithms.

5.5.2 Experiments

The testbed proposed in [35] formed by an imaging arm and an integration arm, is used to obtain the experimental results in this chapter. In this system, the imaging arm is composed of an objective lens and a DLP spatial light modulator (DMD); the integration arm is composed of a relay lens, a dispersive element, and a CCD camera, as shown in Fig. 7.5. To follow the mathematical model, a target scene is illuminated with its reflected light captured by the objective lens and focused onto the mirrors of the spatial light modulator image plane, which plays the role of the coded aperture. When properly aligned, the mirrors of the modulator reflect light into the integration arm, which relays light through a second lens and then through the prism, such that the dispersed field focuses in the CCD image plane, which imposes its spectral filtering and integration based on its wavelength sensitivity. The DMD coded aperture patterns are the same as described in the simulations section, random realizations of 256×256 pixels with 50% transmittance. The prism used in the testbed is a non-linear double Amici prism, which disperses the visible spectrum between 450 - 670 nm onto 24 sensor pixels. The 24 spectral channels have central wavelengths 455, 460, 465, 470, 476, 482, 488, 494, 500, 506, 512, 519, 527, 536, 545, 555, 565, 576, 587, 599, 612, 626, 643, and 660 nanometers. Notice that the bandwidth of the spectral channels is non-uniform due to the non-linearity of the prism. The spectral resolution of the testbed is 5 nm for the short wavelengths, and about 15 nm for the long wavelengths. Therefore, the attained reconstructions with the testbed will exhibit 256×256 pixels of spatial resolution and 24 spectral channels of spectral resolution.

The three CCD sensors under study are emulated based on their system transfer functions, depicted in Figs. 5.2 and 5.4, while using of a single, monochromatic sensor (Stingray F-033B) with pixel size of $\Delta_d = 9.9\mu m$ and 8 bits of dynamic range. The emulation of the monochrome sensor is straightforward, but taking 3-times more snapshots per each snapshot of the color sensors. To emulate the Bayer sensor, the monochrome sensor captures 3 snapshots imaged through either a red, green, or blue glass plate and then selecting the (m, n) positions corresponding to $\Omega_b, \Omega_g, \Omega_r$ in Eq. (5.5), and blocking the remaining. Similarly, the Foveon sensor is emulated as in the Bayer case, but the selection procedure is not required since every pixel captures all the red, green, and blue channels. Figure 5.9 depicts the emulated compressive measurements attained by each sensor, and their corresponding zoomed versions in order to appreciate their differences.

To evaluate the CS capabilities of each sensor, the target scene depicted in Fig. 5.10(a) is imaged through the system. Fig. 5.10(b) depicts the reconstructed data cubes mapped to a pseudo-color profile when $R = 1/8$ and $R = 1/2$. Without a perfect image to compare for calculating PSNR, visual evaluations confirm that the color sensors improve upon the monochrome, with the Foveon and Bayer sensors, exceptionally so.

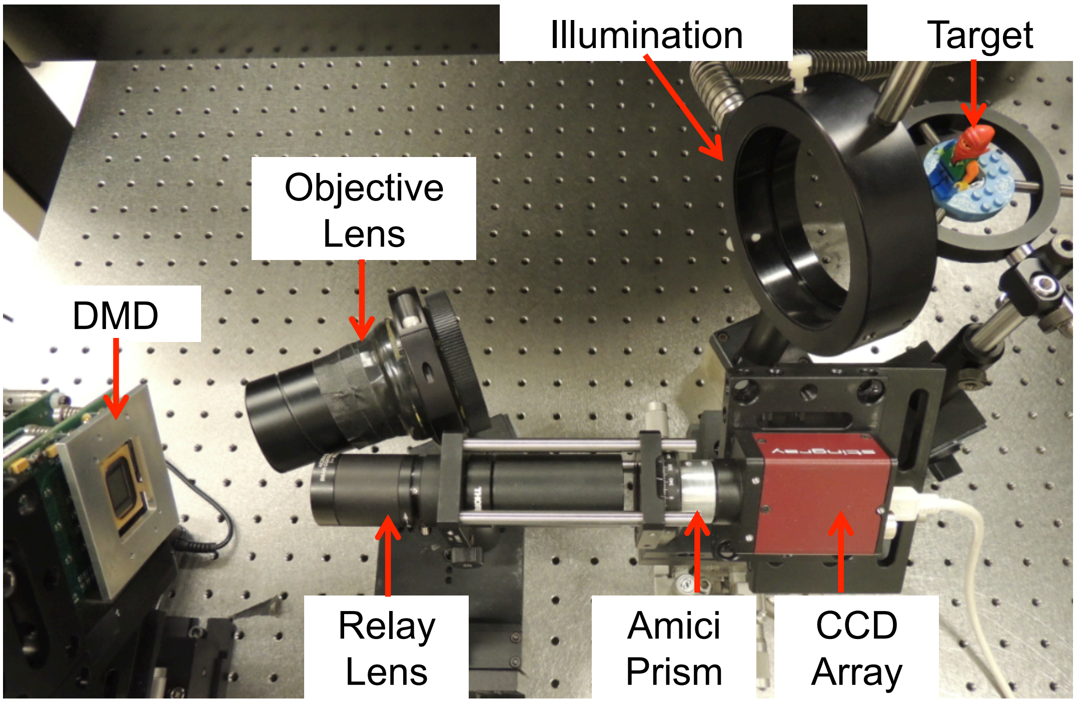


Figure 5.8: DMD-based CASSI testbed setup used in the experiments. The illuminated target scene is imaged onto the image plane of the DMD which plays the role of the coded aperture. Subsequently, the relay lens transmits the coded light through the Amici prism which disperses it onto the image plane of the CCD array.

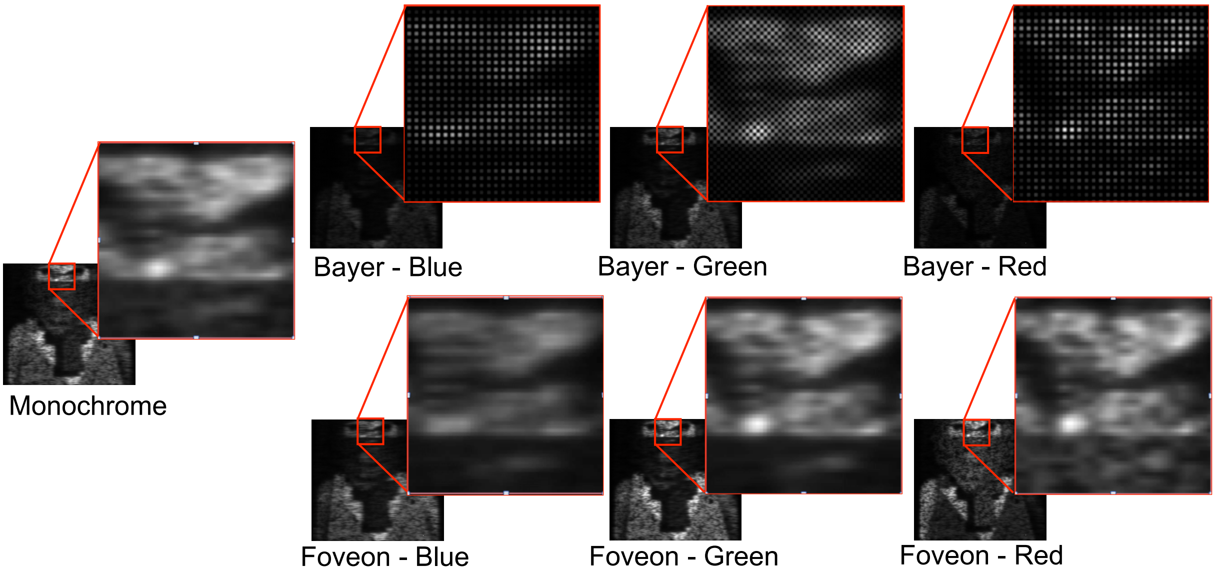
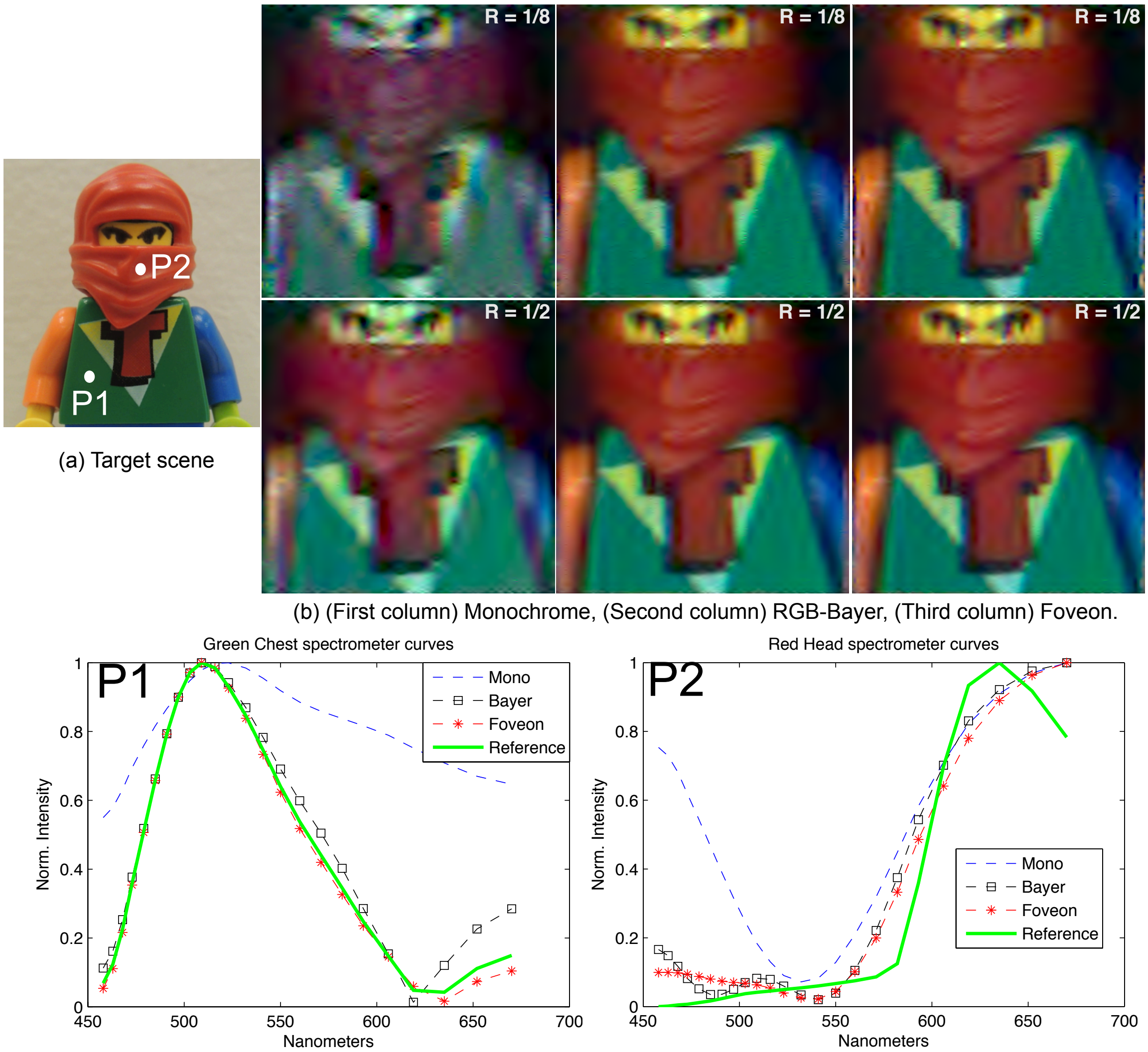


Figure 5.9: Real measurements for a single snapshot of the monochrome, the Bayer and the Foveon sensor. Zoomed versions highlight the sensors differences.

As a means of quantitatively evaluating the various sensors, the spectral signatures of two different spatial points, taken from the target scene (a point from the green chest, and a red point from the helmet), are plotted in Fig. 5.10(c) against a reference signature measured using a commercially available spectrometer assumed as the ground truth. Compared to a traditional monochrome sensor, both the Bayer and Foveon spectra offer substantial improvement, the Foveon sensor especially so at the longer wavelengths. Note that the strong background attained in the reconstructions of the monochrome sensor is the result of the broadband spectral responsivity of the sensor. The strong background reduces its intensity in the color reconstructions signatures due to the spectral filtering and separability imposed by the RGB pixel-wise spectral sensitivity.

5.6 Discussion

Although the color sensors are simulated and emulated through the use of a monochrome sensor, they take into account the aliasing effects resulting of demosaicing in the Bayer sensor, and the real absorption/transmission curves of commercially available Foveon/RGB sensors. Regarding the noise characteristics, they were not taken into account in the manuscript. Several papers have studied the noise characteristics of monochrome, Bayer and previous Foveon sensors [76, 77, 78, 79]. It is important to point



(a) Target scene

(b) (First column) Monochrome, (Second column) RGB-Bayer, (Third column) Foveon.

(c) Spectral signatures comparison for R=1/8 (left) Green chest point. (right) Red helmet point.

Figure 5.10: Experimental reconstruction results. (a) Target scene for the experiments. (b) Real reconstructed datacubes using $R = 1/8$ and $R = 1/2$. (c) Spectral signatures of two reference points measured from the $R = 1/8$ reconstructions.

out that, although the Foveon sensors have been labeled as noisy sensors, recent versions of Foveon, such as the Quattro sensor has reported an alleviation of this issue [80]. To the best of our knowledge a full characterization of the noise characteristics in the Foveon Quattro sensor has not been released nor published yet. Furthermore, the simulations and experimental results assumed single sensor sensing, although the Quattro sensor provides finer resolution due to the top (blue) layer partitions pixels in 4 subpixels that can be sensed independently. Analysis of the noise characteristics as well as the higher resolution capabilities of newer Foveon sensors, such as the Quattro, will be considered and analyzed in a future work.

In terms of the quality of the reconstructions, a 1 dB improvement is attained in average by the Foveon sensor over the Bayer sensor. However, when compared against the monochrome sensor, the color sensors overcome the monochrome sensor by about 2-4 dBs. In addition, and most importantly, the color sensors have the advantage of just requiring a single exposure as opposed to the three-times snapshots required by the monochrome sensor to attain the same amount of compressive measurement

Recent research works have proposed different approaches to recover high resolution spectral images from compressive measurements, such as the use of side information in hybrid camera architectures [81, 82]. These hybrid cameras usually rely on a second image sensor (usually an RGB sensor) to capture the side information of a scene. Although improved results are attained by these recent approaches, the architecture proposed in this manuscript mainly differs in that a single sensor is used. Remark that a number of issues arises that need to be accounted for, in a thorough comparison of these 2 distinct sensing and reconstruction methods. These issues include the registration post-processing and double alignment issues (due to the beam-splitter), and the extra sensor/lens requirement for the side information method.

5.7 Conclusions

Compressive sensing is a powerful signal processing sensing and reconstruction framework to extract clean signals in noisy environments, and for spectral imaging, creates many new opportunities for research. But because of its novelty, few works have been performed to understand how the CS process will merge with existing methods of collecting spectral images. In this chapter, we established both theoretical and experimental evaluations of how using a readily available color sensor improves upon the so far published works that rely on monochrome image sensors. And as more work is done to expand the stacked-color image sensor first pioneered by Foveon to using more colors or a wider range of wavelengths, the CS method described here will take those systems ever higher in spectral resolution.

In closing, let us present Fig. 5.11 as a final evaluation of color sensor abilities to reconstruct multispectral images where we show the 6^{th}, 12^{th}, 18^{th} and 24^{th} wavebands

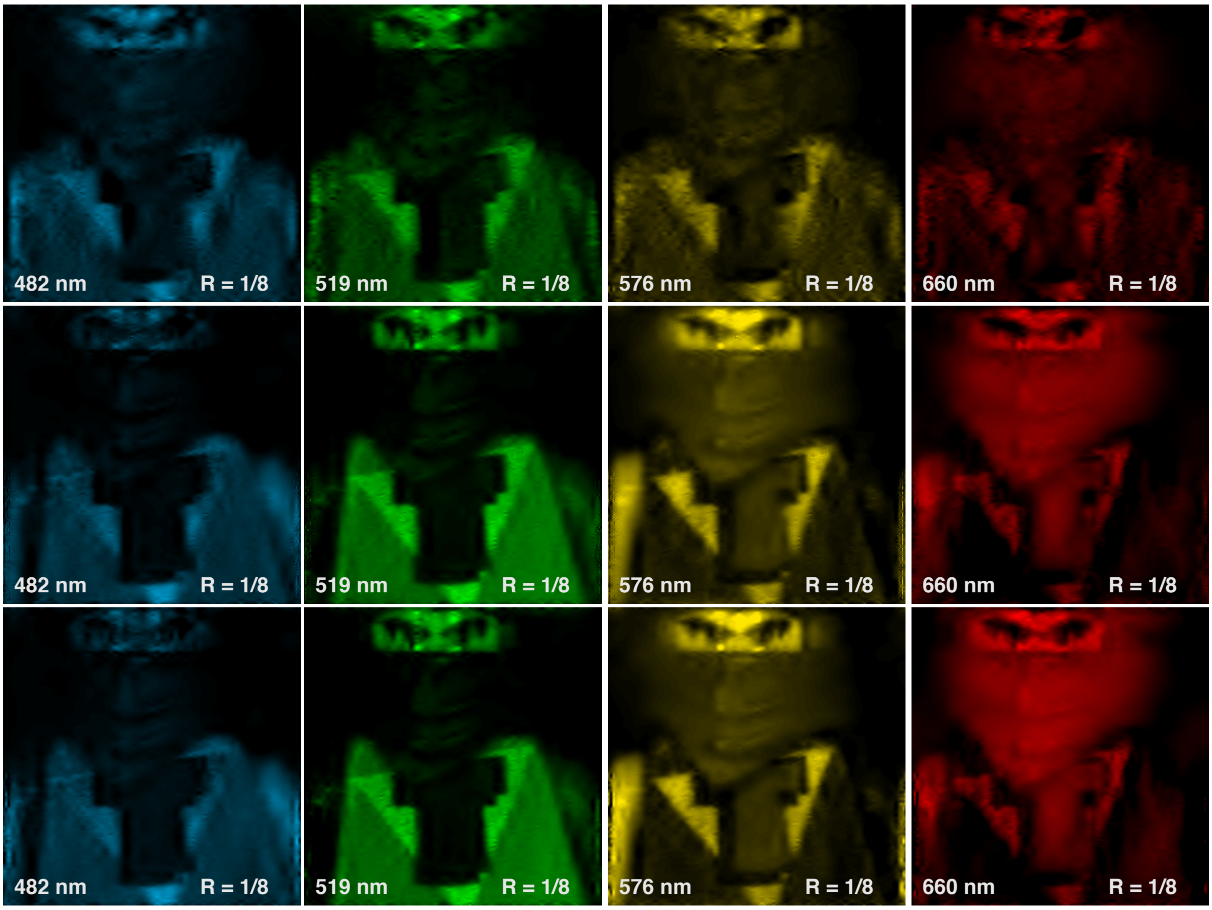


Figure 5.11: Real reconstruction of the 6^{th}, 12^{th}, 18^{th} and 24^{th} spectral bands using just a single snapshot of the (First row) Monochrome sensor, (Second row) Bayer sensor, (Third row) Foveon sensor.

attained with as few as a single snapshot. Here, it can be noticed that the Bayer and Foveon detectors attain better artifact-free reconstructions in their respective wavebands compared with the monochrome sensor, and in particular, Fig. 5.11 shows that using just a single snapshot with the Foveon sensor produces spectral bands with greater fidelity than the monochrome and the Bayer sensor.

Chapter 6

DUAL-ARM VIS/NIR COMPRESSIVE SPECTRAL IMAGER

Compressive spectral imaging (CSI) has demonstrated to be a feasible technique for capturing the 3D spatio spectral information of a scene through less measurements than the Nyquist rate. The coded aperture snapshot spectral imaging (CASSI) is an example of a CSI optical architecture, which has been proposed to work on the visible electromagnetic spectrum. Due to the rich information contained in the infrared spectrum, in this chapter, we mathematically model and demonstrate the implementation of a broadband CASSI system covering the visible and the near infrared spectra between 448 nm to 1436 nm. Particularly, we developed a Digital Micromirror Device-based CSI system, which implements dual-band CS measurement processes of 3D spatio-spectral scenes.

6.1 Introduction

Spectral imaging involves the sensing of a large amount of spatial information across a multitude of wavelengths. Conventional approaches to hyperspectral sensing scan adjacent zones of the underlying spectral scene and merge the results to construct a spectral data cube. Push broom spectral imaging sensors, for instance, capture a spectral cube with one focal plane array (FPA) measurement per spatial line of the scene [67, 7]. Spectrometers based on optical band-pass filters sequentially scan the scene by tuning the band-pass filters in steps. The disadvantage of these techniques is that they require scanning a number of zones linearly in proportion to the desired spatial and spectral resolution. Compressive coded aperture spectral imagers, also known as coded aperture snapshot spectral imagers (CASSI) [30, 1, 36, 67] comprise the new generation of spectral imagers. These naturally embody the principles of compressive sensing (CS) [73, 17]. The remarkable advantage of CASSI imagers is that the entire data cube is sensed with just a few FPA measurements and in some cases with as little as a single FPA shot.

CS dictates that one can recover spectral scenes from far fewer measurements than that required by conventional linear scanning spectral sensors. To make this possible, CS relies on two principles: sparsity, which characterizes the spectral scenes of interest, and incoherence, which shapes the sensing structure [19, 17]. Sparsity indicates that spectral images found in nature can be concisely represented in some basis $\boldsymbol{\Psi}$ with just a small number of coefficients. This is indeed the case in spectral imaging where natural scenes exhibit correlation among adjacent pixels and also across spectral bands [7]. Incoherence refers to the structure of the sampling waveforms used in CS, which, unlike the signals of interest, have a dense representation in the basis $\boldsymbol{\Psi}$ [19]. The remarkable discovery behind CS is that it is possible to design sensing protocols capable of capturing the essential information content in sparse signals with just a small number of compressive measurements. The sensing modality simply correlates incoming signals with a small number of fixed waveforms that satisfy the incoherence principle. The signals of interest are then accurately reconstructed from the small number of compressive measurements by numerical optimization [29, 73, 17, 27, 26, 83].

Due to the rich information contained in both the visible (VIS) and the near infrared (NIR) spectrum [35, 84, 85], a dual-arm VIS-NIR compressive spectral imager is presented in this chapter. Figure 6.1 depicts the optical design of the system being proposed. This architecture aims to cover a wider-band of the electromagnetic spectrum in order to identify, classify and detect objects in the VIS and the NIR domains. Particularly, the proposed system covers the spectrum ranging between 450 nm to 1450 nm. This system uses a VIS-NIR dichroic mirror to split the underlying incoming coded energy into two imaging arms, each one accounting for an independent compressive spectral imager.

6.2 System Model

6.2.1 Optical Model

The sensing physical phenomena in CASSI is strikingly simple, yet it adheres to the incoherence principles required in CS. The proposed dual-arm CASSI measurements are realized optically by a coded aperture, a dispersive element such as a prism, and a focal plane array detector [1, 67]. The coding is applied to the (spatial-spectral) image source

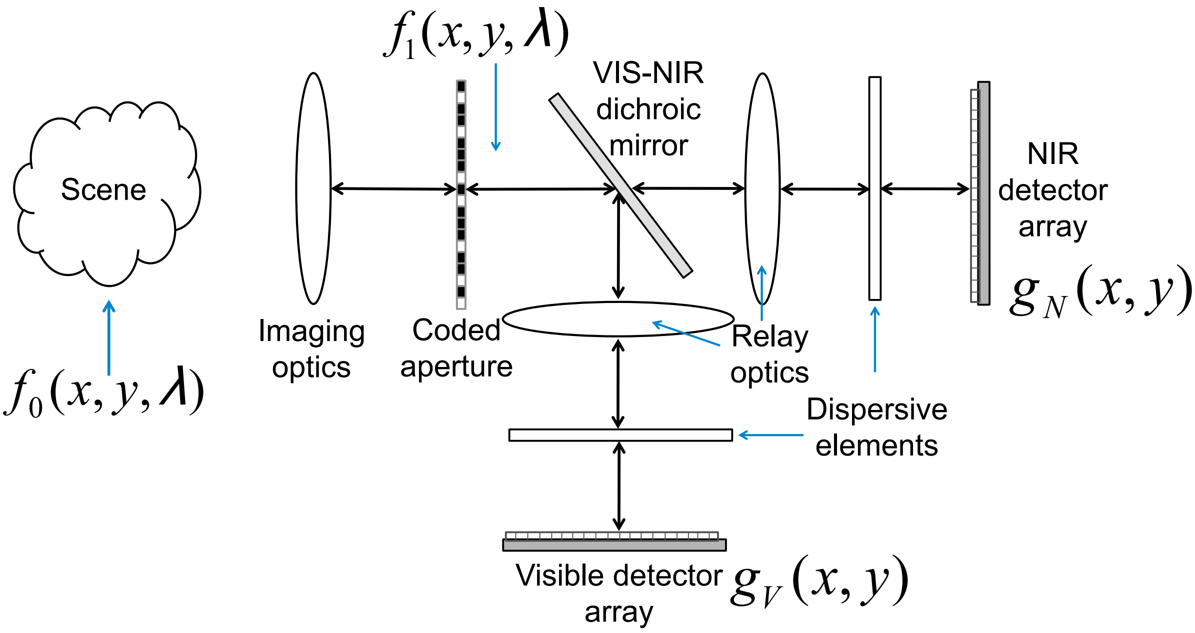


Figure 6.1: Dual-Arm optical design

density $f_0(x, y, \lambda)$ by means of a coded aperture $T(x, y)$ as depicted in Fig. 6.1, where (x, y) are the spatial coordinates and λ is the wavelength [1]. The resulting coded field $f_1(x, y, \lambda) = T(x, y)f_0(x, y, \lambda)$ is subsequently split by means of a VIS-NIR dichroic mirror into the two imaging arms, the visible arm and the near infrared arm. Notice that inside each arm, the coded and split information is subsequently modified by a dispersive element before it impinges onto the corresponding FPA detector. The compressive measurements $g(x, y)$ across the FPA are realized by the integration of the coded, split and dispersed field. Mathematically, let $S_V(\lambda)$ and $S_N(\lambda)$ represent the dispersion function of the VIS and NIR prism, respectively. Similarly, let Λ_V and Λ_N represent the spectral range of the visible (450 nm - 750 nm) and the near infrared (750 nm - 1450 nm), respectively. Therefore the compressive measurements for the VIS imaging arm $g_V(x, y)$, and for the NIR arm $g_N(x, y)$, are given by:

$$g_V(x, y) = \iiint_{\Lambda_V} T(x', y')f_0(x', y', \lambda)h_V(x' - x - S_V(\lambda), y' - y)d\lambda dx'dy' \qquad (6.1)$$

$$g_N(x,y) = \iiint_{\Lambda_N} T(x',y') f_0(x',y',\lambda) h_N(x'-x-S_N(\lambda), y'-y) d\lambda dx' dy', \tag{6.2}$$

where $h_V(x'-x-S_V(\lambda), y'-y)$, and $h_N(x'-x-S_N(\lambda), y'-y)$ are the system impulse responses of the imaging arms, which account for the relay lenses and the dispersion function. Notice that the dispersion is assumed to occur just along the x-axis.

6.2.2 Discrete Model

The sensing mechanism of each of the two VIS-NIR-CASSI arms is illustrated by the discretized model shown in Fig. 6.2, where the spectral data cube $\boldsymbol{\mathcal{F}}$ having L spectral bands and $N \times M$ spatial pixels is first amplitude modulated by a pixelated $N \times M$ coded aperture T [36]. In this case, T is a "block" or "unblock" coded aperture such that the energy along an entire slice of the data cube is punched out when a block coded aperture element is encountered. As the coded field transverses the prism, it is then spatially sheared along one spatial axis. In essence, each coded image plane is shifted along the x-axis where the amount of shifting increases with the wavelength coordinate index. Finally, the coded and dispersed field is collapsed in the spectral dimension by the integration of the energy impinging on each detector element over its spectral range sensitivity. The integrated field is then measured by the FPAs detector elements. It can be shown that if the spectral range of the instrument lies between λ_{V_1} and λ_{V_2} for the visible arm, and between λ_{N_1} and λ_{N_2} for the NIR arm, then the number of resolvable spectral bands is limited by $L = L_V + L_N = \alpha(\lambda_{V_2} - \lambda_{V_1})/\Delta_V + \alpha(\lambda_{N_2} - \lambda_{N_1})/\Delta_N$, where $\alpha\lambda$ is the dispersion induced by any of the prisms and Δ_V and Δ_N are the pixel pitches of the visible and NIR FPA detector, respectively.

Let $T(x,y) = \sum_{i,j} t_{i,j} \text{rect}\left(\frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j\right)$, where Δ_c is the pixel size of the coded aperture, and $f_0(x,y,\lambda) = \sum_{i,j,k} f_{i,j,k} \text{rect}\left(\frac{x}{\Delta_c} - i, \frac{y}{\Delta_c} - j, \frac{\lambda}{\Delta_d} - k\right)$, where $\Delta_d = \{\Delta_V, \Delta_N\}$, and $f_{i,j,k}$ represents the $(i,j,k)^{th}$ discretized source voxel. Note that $i = 0, \ldots, N-1$, and $j = 0, \ldots, M-1$ index the x and y spatial coordinates, whereas $k = 0, \ldots, L-1$ indexes the wavelength. Assuming that the system impulse responses are ideal an space invariant, and that $\Delta_c = \Delta_d$, Eqs. (6.1) and (6.2) can be rewritten in

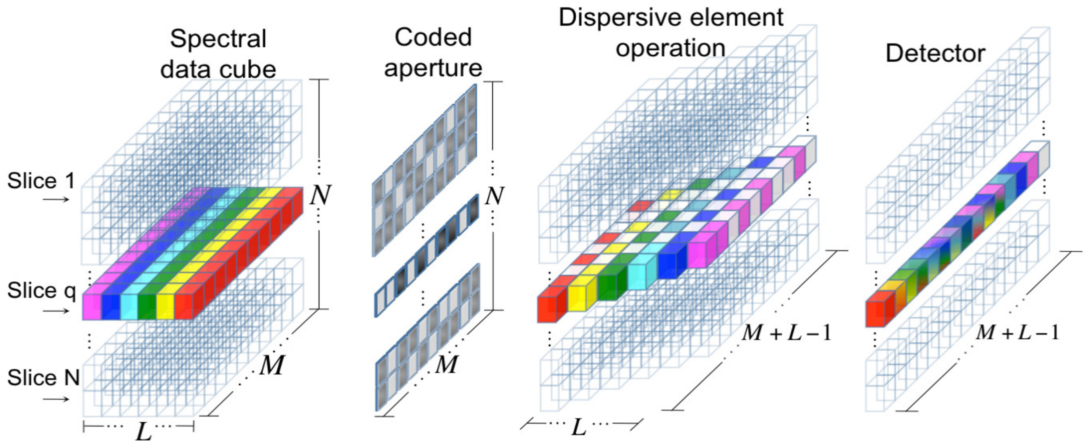


Figure 6.2: Sensing phenomenon performed in each imaging arm

discrete form as,

$$g_{V\,m,n} = \sum_k t_{m,n+k} f_{V\,m,n+k,k} \tag{6.3}$$

$$g_{N\,m,n} = \sum_k t_{m,n+k} f_{N\,m,n+k,k}, \tag{6.4}$$

where $n = 0, \ldots M + L - 1$, and $m = 0, \ldots, N - 1$ index the columns and rows of the FPA detectors; k index the FPA pixels affected by the spreading of $f_0(x, y, \lambda)$ by means of $S_V(\lambda)$ and $S_N(\lambda)$.

Using Eqs. (6.3) and (6.4), the projections in the dual VIS-NIR-CASSI system can be expressed in matrix form as $\mathbf{g}_V = \mathbf{H}_V \boldsymbol{f}_V$, and $\mathbf{g}_N = \mathbf{H}_N \boldsymbol{f}_N$, where $\mathbf{H}_V$ and $\mathbf{H}_N$ are matrices of size $N(M+L_V-1)\times(N\cdot M\cdot L_V)$ and $N(M+L_N-1)\times(N\cdot M\cdot L_N)$, respectively, whose structure is determined by the coded aperture entries and the dispersive element effect, and $\boldsymbol{f}_V$ and $\boldsymbol{f}_N$ and are the visible and near infrared spectral bands of the scene in vector form.

6.2.3 Image Reconstruction

For spectrally rich or very detailed spatial scenes, a single-shot of the dual VIS-NIR-CASSI may not provide a sufficient number of compressive measurements. Thus, increasing the number of measurement shots, each with a distinct coded aperture pattern that remains fixed during the integration time of the detector, will rapidly increase the quality of image reconstruction [23, 42, 43]. Notice, that each new measurement shot adds simultaneously $N(M+L-1)$ compressive projections; thus, the total number of available projections with K shots is $m = KN(M+L-1)$. The time-varying coded apertures are realized by a digital micro-mirror device (DMD) [42, 43]. The DMD is considered a versatile system since the micro-mirrors can tilt at video rates, thus leading to a multi-frame compressive imaging system. Particularly, denoting the ℓ^{th} FPA measurements as $\mathbf{g}_V^\ell = \mathbf{H}_V^\ell \boldsymbol{f}_V$ and $\mathbf{g}_N^\ell = \mathbf{H}_N^\ell \boldsymbol{f}_N$, where $\mathbf{H}_V^\ell$ and $\mathbf{H}_N^\ell$ represent the effects of the ℓ^{th} coded aperture pattern, the sets of the K dual FPA measurements are then assembled as,

$$\mathbf{g}_V = \left[(\mathbf{g}_V^0)^T, \ldots, (\mathbf{g}_V^{K-1})^T\right]^T, \tag{6.5}$$

$$\mathbf{g}_N = \left[(\mathbf{g}_N^0)^T, \ldots, (\mathbf{g}_N^{K-1})^T\right]^T. \tag{6.6}$$

In order to recover an estimation of the input 3D spectral scene, CS supposes that the hyperspectral signal $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{N\times M\times L}$, or its vector representation $\boldsymbol{f} \in \mathbb{R}^{N\cdot M\cdot L}$, is S-sparse on some basis $\boldsymbol{\Psi}$, such that $\boldsymbol{f} = \boldsymbol{\Psi}\boldsymbol{\theta}$ can be approximated by a linear combination of S vectors from $\boldsymbol{\Psi}$ with $S \ll (N\cdot M\cdot L)$. Consequently, the projections in CASSI can be alternatively expressed as $\mathbf{g}_V = \mathbf{H}_V\boldsymbol{\Psi}\boldsymbol{\theta}_V = \mathbf{A}_V\boldsymbol{\theta}_V$, and, $\mathbf{g}_N = \mathbf{H}_N\boldsymbol{\Psi}\boldsymbol{\theta}_N = \mathbf{A}_N\boldsymbol{\theta}_N$, where the matrices $\mathbf{A}_V$ and $\mathbf{A}_N$ are called the visible and the near-infrared compressive sensing matrices. Based on this assumption, CSI allows $\boldsymbol{f}_V$ and $\boldsymbol{f}_N$ to be recovered from m random projections with high probability when $m \gtrsim S\log(N\cdot M\cdot L) \ll (N\cdot M\cdot L)$. In this regard, the underlying input 3D spectral data cubes are separately recovered by solving the minimization problems,

$$\tilde{\boldsymbol{f}_V} = \boldsymbol{\Psi}^{-1}\left(\underset{\boldsymbol{\theta}_V}{\operatorname{argmin}}||\mathbf{g}_V - \mathbf{H}_V\boldsymbol{\Psi}\boldsymbol{\theta}_V||_2 + \tau||\boldsymbol{\theta}_V||_1\right), \tag{6.7}$$

$$\tilde{\boldsymbol{f}_N} = \boldsymbol{\Psi}^{-1}\left(\underset{\boldsymbol{\theta}_N}{\operatorname{argmin}}||\mathbf{g}_N - \mathbf{H}_N\boldsymbol{\Psi}\boldsymbol{\theta}_N||_2 + \tau||\boldsymbol{\theta}_N||_1\right), \tag{6.8}$$

where $\mathbf{H}_V = \left[(\mathbf{H}_V^0)^T, \ldots, (\mathbf{H}_V^{K-1})^T\right]^T$, and, $\mathbf{H}_N = \left[(\mathbf{H}_N^0)^T, \ldots, (\mathbf{H}_N^{K-1})^T\right]^T$; and τ is a regularization constant. Notice that the final estimation of the input scene can be obtained by $\tilde{\boldsymbol{f}} = [(\tilde{\boldsymbol{f}_V})^T, (\tilde{\boldsymbol{f}_N})^T]^T$.

6.2.4 Joint Image Reconstruction

Instead of solving Eqs. (6.7) and (6.8) separately, a joint minimization problem can be solved, in order to achieve better image reconstructions. In consequence, the set of measurements given in Eqs. (6.5) and (6.6) are now concatenated in a single column vector $\mathbf{g} = [\mathbf{g}_V^T, \mathbf{g}_N^T]^T$, and the minimization problem to solve is now given by,

$$\tilde{\boldsymbol{f}} = \boldsymbol{\Psi}^{-1}\left(\underset{\boldsymbol{\theta}}{\operatorname{argmin}}||\mathbf{g} - \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}||_2 + \tau||\boldsymbol{\theta}||_1\right), \tag{6.9}$$

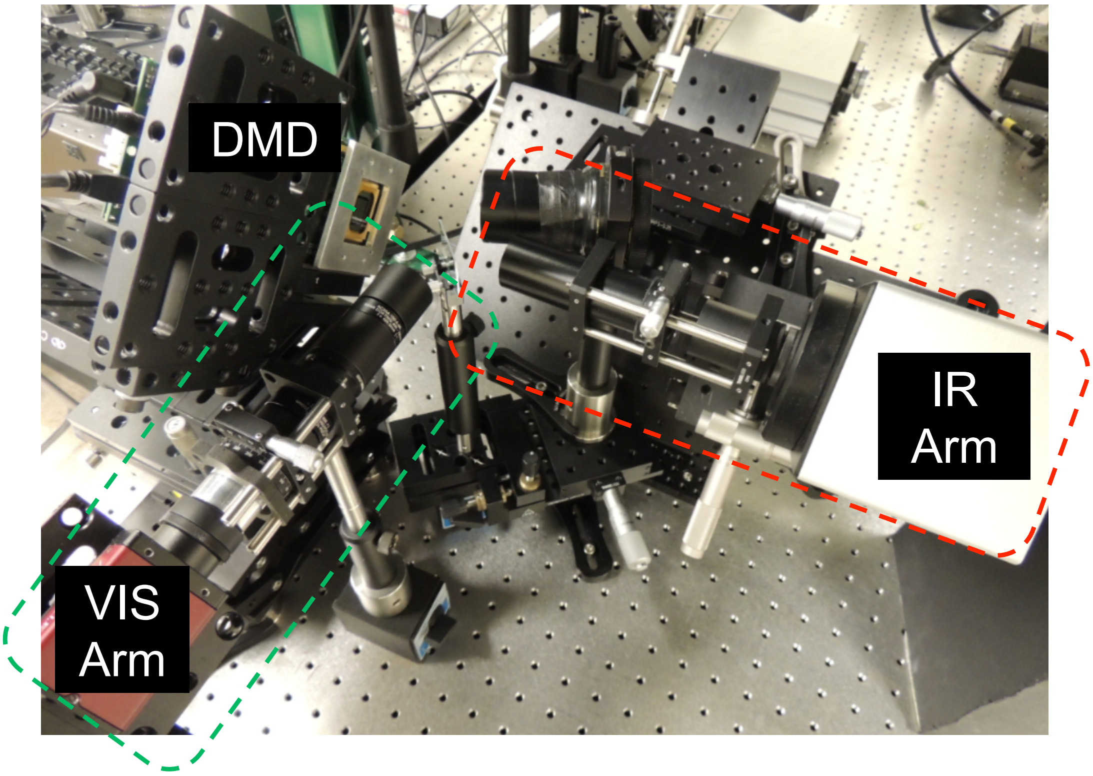
where $\mathbf{H} = \begin{bmatrix} \mathbf{H}_V & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_N \end{bmatrix}$.

6.3 Experimental Setup

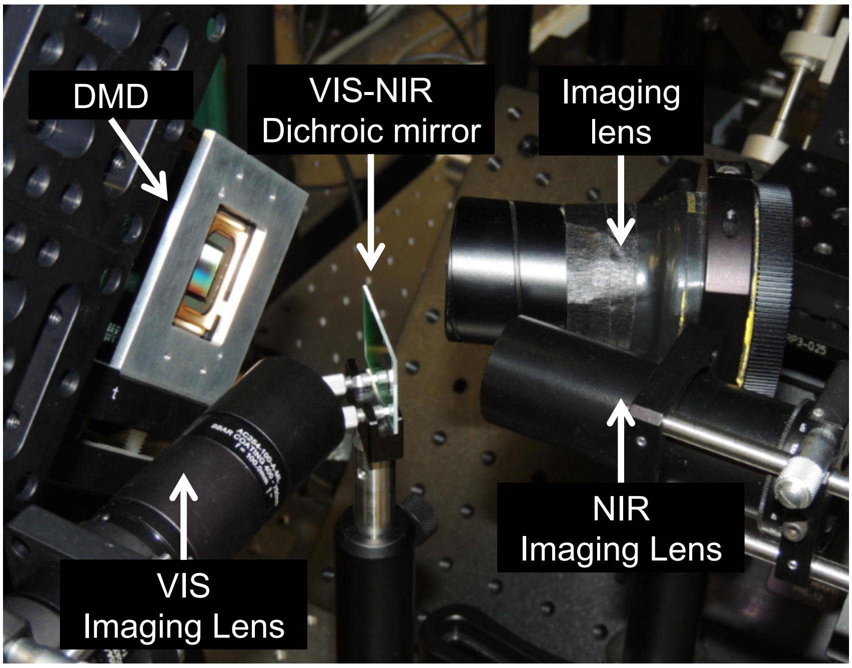
A proof-of-concept optical prototype was constructed to experimentally verify the proposed architecture. This prototype shown in Fig. 6.3 consist of (i) a Leica COLORPLAN-P2 objective lens, (ii) a digital micro-mirror device (DMD) XGA format aluminium mirror array with $\sim$ 92 % reflectivity in the VIS-NIR spectrum, 13.68μm pitch size, and +/- 12-degree tilt angle about the diagonal hinge in the center of the micro-mirrors. (iii) an 875 nm single-edge VIS-NIR dichroic beamsplitter, (iv) the visible imaging arm, and (v) the NIR imaging arm. Each imaging arm is composed by a relay lens, a custom Amici prism, and a monochrome FPA detector.

The visible arm has a double Amici prism which disperses the 448 nm - 662 nm visible light onto 15 different FPA detector pixels, thus entailing the identification of $L_V = 15$ spectral bands. The visible arm FPA detector is a Stingray F-033B camera with pixel size $\Delta_V = 9.9\mu m$. On the other hand, the NIR imaging arm, has a double Amici prism, which disperses the 946 nm - 1436 nm NIR light onto 12 different FPA detector columns, thus entailing the identification of $L_N = 12$ spectral bands. The NIR detector is a XEVA-2508, with pixel size $\Delta_N = 20\mu m$.

To obtain the dual-arm compressive measurements, a fluorescent broadband spectral light source is used to illuminate the target scene shown in Fig. 6.4(a). The DMD



(a)



(b)

Figure 6.3: Dual-arm implementation. (a) Full optical prototype with the two imaging arms. (b) Zoomed version.

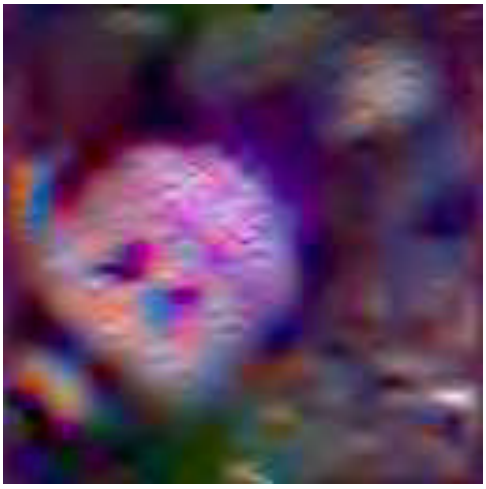
is set to exhibit random coding patterns with 50% of translucent elements. A calibration data cube is captured per each different coding pattern, illuminating a completely white target with a visible-to-NIR monochromator spanning the spectral range between 448nm to 1436nm. The Gradient Projection for Sparse Reconstruction (GPSR) algorithm [29] is used to solve the problem in Eq. 6.9. The representation space $\boldsymbol{\Psi}$ is set to be the Kronecker product of the 2D-Wavelet Symmlet 8 basis and the 1D-DCT basis, due to its sparsifing properties both in the spatial and in the spectral domains.

Figure 6.5 shows the reconstructed spatial/spectral image cube of the target scene. In Fig. 6.5(a) the visible wavelengths are shown, where the Lego target can be easily seen. On the other hand, in Fig. 6.5(b), we can see that LED 1 was reconstructed with strong intensities in spectral channels from 1030 nm to 1110 nm, LED 2 was reconstructed with strong intensities in spectral channels from 1150 nm to 1230 nm, and LED 3 was reconstructed with strong intensities in spectral channels from 1230 nm to 1310 nm. It is important to notice that the reconstructions where attained by capturing just $K = 6$ compressive measurements, that is, less than 50% of the input data information. Note that the FPA measurements are contaminated by optical aberrations, as well as noise and misalignment between the FPAs and the coded apertures, which may be improved in future and extensive experiments.

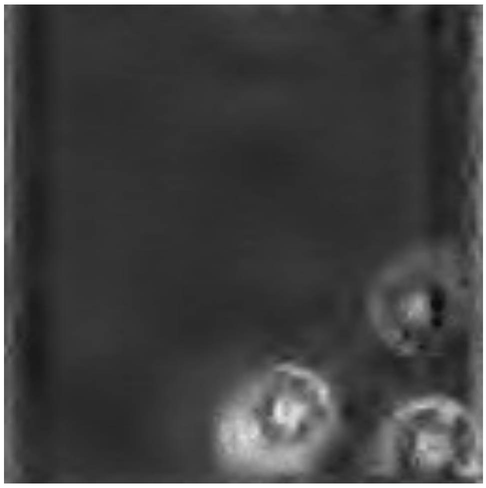
6.4 Conclusions

In this chapter we have presented the optical model and experimental demonstration of a dual-arm compressive sensing spectral imaging system in the broadband visible and near-infrared regime. We have shown that suitable quality estimations of spectral scenes can be attained with far less measurements than the required by the widely-known Nyquist theorem, in both the visible and the near infrared spectrum. Future work will focus on the improvement in the quality of the reconstructions through the design of proper coding patterns.

(a) Original scene

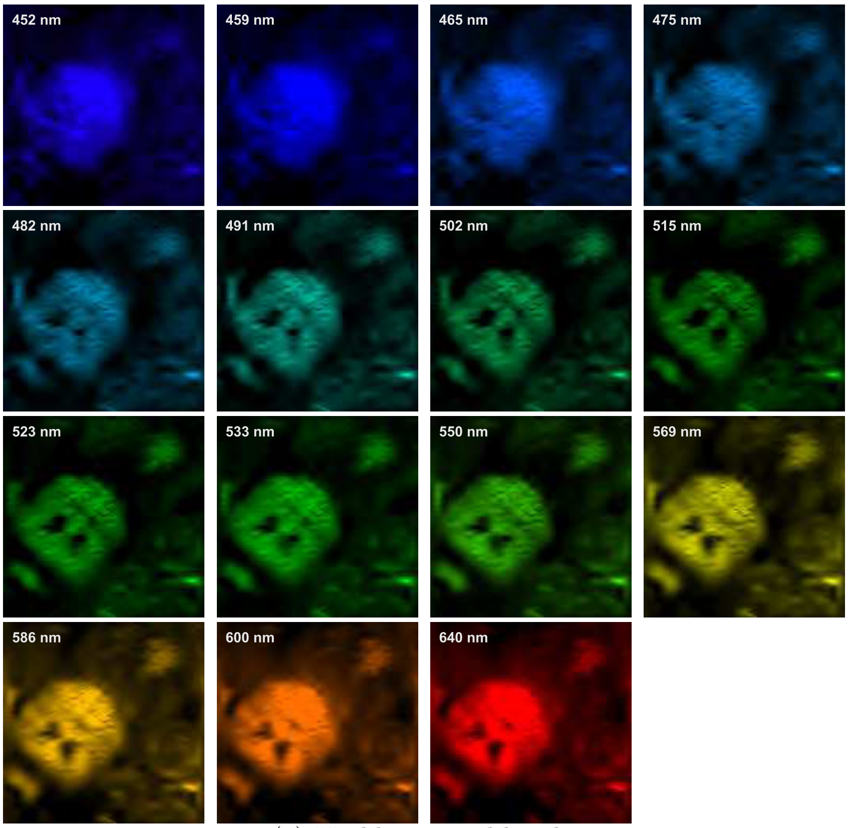


(b) Visible reconstruction

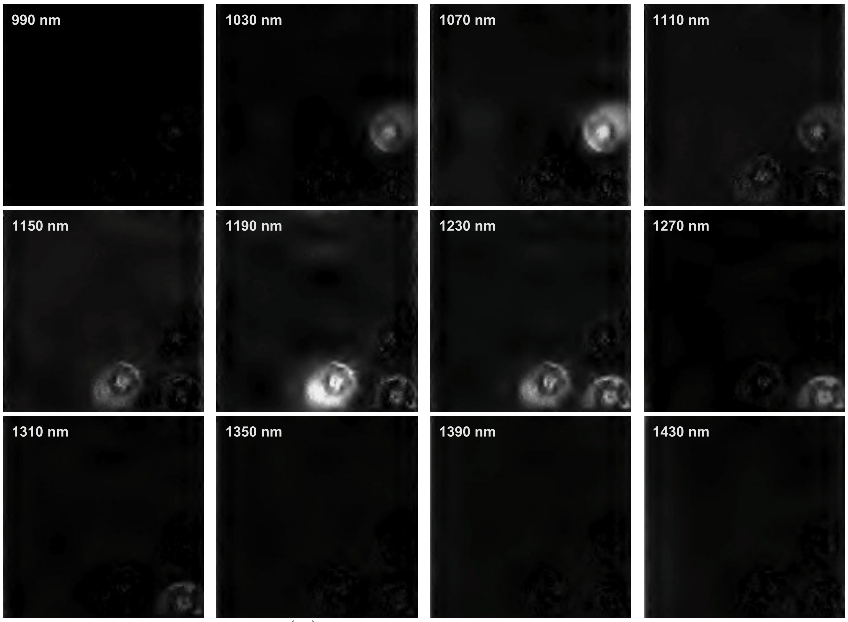


(c) NIR reconstruction

Figure 6.4: Original scene and reconstructed data cubes. (a) Lego toy with 3 LEDs with peak emission wavelengths at 1050 nm, 1200 nm, and 1300 nm. (b) Visible data cube reconstruction, and (c) NIR data cube reconstruction.



(a) Visible spectral bands



(b) NIR spectral bands

Figure 6.5: Monochome reconstructed spectral bands. (a) Visible spectral bands. (b) NIR spectral bands

Chapter 7

SINGLE APERTURE SPECTRAL+TOF COMPRESSIVE CAMERA: TOWARDS HYPERSPECTRAL+DEPTH IMAGERY

Compressive spectral imaging (CSI) has allowed to estimate spectral images with as few as a single coded snapshot. On a different front, 3D ranging imaging often involves scanning along one of the spatial dimensions to estimate the depth of an scene using structured light, or the use of two cameras as required by stereo-imaging techniques. Recently, Time-of-Flight (ToF) snapshot imaging has gained considerable attention, due to its accuracy and speed. To date, however, these imaging modalities (CSI and ToF) have been realized and implemented by separate independent imaging sensors. This chapter presents the development of a single aperture compressive spectral + depth imaging camera that employs a commodity 3D range ToF sensor as the sensing device of a coded-aperture-based compressive spectral imager. The proposed system uses a single aperture/single sensor, thus representing a significant improvement over existing RGB+D cameras that integrate two separate image sensors, one for RGB and another for depth. In addition, the observable wavelength range of the CSI device is expanded from the visible to the near-infrared, resolving up to as many as 16 independent channels. The proposed system allows the addition of side-information by placing a grayscale or RGB camera in the same path of the single-aperture system, such that the quality of the spectral estimation is improved, while maintaining high-frame rates. We demonstrate the proposed ideas through real experimentation conducted on an assembled CSI+ToF testbed camera system.

7.1 Introduction

Imaging spectroscopy, commonly referred to as multispectral or hyperspectral imaging is the process of sensing large amounts of spatial information across a multitude of wavelengths [30]. Compressive spectral imaging (CSI) systems comprise the new generation of spectral imagers, since they record all wavelengths of a scene in a single snapshot, simultaneously [30]. In CSI systems, a ray of light emanating from the scene is passed through a series of optical elements that spread the single ray of light over a lateral sequence of sensor pixels, in a wavelength dependent manner, just like a spectrophotometer would spread light across a linear array. Allowing multiple rays of light onto the sensor means that the camera can resolve a two dimensional image; however, because the spreading of light rays across lateral sequences of pixels means that a single pixel will collect light from multiple sources, compressive sensing techniques are used to decouple the spectral profiles of neighboring pixels [23, 42].

On the other hand, 3-D range imaging, refers to the process of recording a 2-D image of a scene where each pixel is a measure of the distance from the camera sensor to the target surface. There exist two main modalities to estimate 3D imagery: passive illumination such as stereo [86] and light field imaging [87, 88], and active illumination such as structured light imaging (SLI) [89, 90] and time-of-flight (ToF) imaging [91, 92]. One of the big disadvantages of passive systems is that they cannot derive depth without depth cues created by texture; this is known as the depth stereo problem. Passive methods and SLI, rely on triangulation to interpolate depth values in the spaces between edges. The problem with triangulation is the difficulty to work over long ranges since the size of the triangle has to grow proportionally. They also have to deal with problems of occlusion where target surfaces are visible to the camera but not to the projector. ToF range cameras, in contrast, indirectly measure the round-trip propagation delay of the light pulse by strategically triggering each pixel's exposure time such that it captures a portion of the projected pulse. ToF sensors do not suffer of occlusion problems, since the projector and camera have the same line of sight. Moreover, ToF sensors can reach longer ranges by just changing the frequency and amplitude of the modulated light, they are lightweight, power efficient, and can collect full frame images with a single laser pulse. Sensors from Mesa Imaging, Microsoft, SoftKinetic, PMD Tech, among others, rely on custom imaging chips that combine traditional CCD and CMOS light detectors with custom application specific integrated circuits (ASICs). While effective, these techniques tend to produce camera sensors with low pixel counts.

In some cases, the information in both imaging modalities can be combined as in RGB+Depth imaging to infer qualitative features in a scene [93, 94]. To date, however, these imaging modalities have been realized and implemented by separate independent imaging sensors.

7.1.1 Related work

The importance of spectral+3D imaging has been reported recently. Conventional joint spectral and depth imaging usually requires scanning the 4-D scene either spatially or spectrally. Latorre-Carmona, et. al. [95] proposed to combine a lenslet-based integral imaging with a liquid crystal tunable filter that scans each spectral band for spectral integral imaging. Spectral scanning can be avoided through combining a lenslet array with a pixel-wise color filter [96, 97]; however, the spectral resolution is limited by the number of colors in the pixel-wise filter. To overcome scanning and the spectral resolution limitation, compressive spectral integral imagers have been developed employing a liquid crystal or a dispersive element for spectral modulation [98, 99]. These compressive imagers only apply compressive sensing to acquire spectral elemental images. Thus, a second-step integral imaging reconstruction must be followed. Recently, snapshot spectral integral imagers based on lenslet arrays were introduced, where spectro-volumetric images are directly reconstructed from compressed projections [100, 101]. A different approach obtains the spectral and depth information of a scene by adding a grayscale camera on the side of a multi-spectral imager [102], such that a passive stereo system is attained.

In spite of the above 4-D imagers employing passive range finding techniques, active ranging such as laser scanning methods were also applied in spectral imaging. M. Kim, et. al. [103] combines a coded aperture snapshot spectral imager with a conventional laser scanner in their spectral imager design. This imaging system, however, involves a scanning illumination source and two independent detectors. As a replacement of the laser scanning based ranging, ToF cameras acquire the depth information simultaneously on the sensor array. Compressive sensing has been applied with ToF techniques for high resolution depth map reconstructions [104, 105].

7.1.2 Contribution

To our knowledge, spectral imaging based on ToF sensors has been limited to combining a ToF camera with a separate spectral imager or an RGB camera [106]. This chapter introduces a new spectral+TOF imaging camera that measures and reconstructs multispectral and 3D imagery at once. It aims to create super-human sensing that will enable the perception of our world in new and exciting ways. We intend to advance in the state of the art in compressive sensing systems to extract depth while accurately capturing spectral material properties. The applications of such a sensor are self-evident in fields such as computer/robotic vision because they would allow an artificial intelligence to make informed decisions about not only the location of objects within a scene but also their material properties. These goals are made possible after the realization that ToF sensors can exploit both ambient light and modulated light, in order to recover pixelated depth and multispectral information. In particular, we build on this opportunity and integrate the capabilities of a CMOS ToF sensor into a CSI snapshot camera. So unlike a Kinect 2.0 or Prime Sense RGB+D camera, the proposed system uses a single aperture image sensor to capture multispectral+depth (MS+D) images. Moreover, by changing the dispersive element, used in the proposed system, by a diffraction grating, we can move towards hyperspectral+depth (HS+D) imagery, as it will be described shortly.

7.2 Snapshot Single Aperture Compressive ToF+Multispectral Camera

The ability to switch measurement modes to capture both ambient light and ToF depth images in the same sensor array enables a dual-shot single aperture imaging system, where spectral and 3D range depth images are reconstructed from the same ToF sensor. Figure 7.1 depicts a sketch of the proposed single aperture system. It uses a set of lenses for image formation, an LED laser diode illumination board for ToF depth excitation, a digital micromirror device (DMD) to emulate the coded apertures, a dispersive element to spread the ambient illumination, and a ToF sensor to integrate the incoming light field, which can operate in ToF or ambient light mode. The depth image is obtained through processing four discrete pulsed frames measured in the ToF mode, as will be explained next, by turning on all the mirrors of the DMD, while the spectral images will be recovered from compressed projections at the ambient grayscale mode.

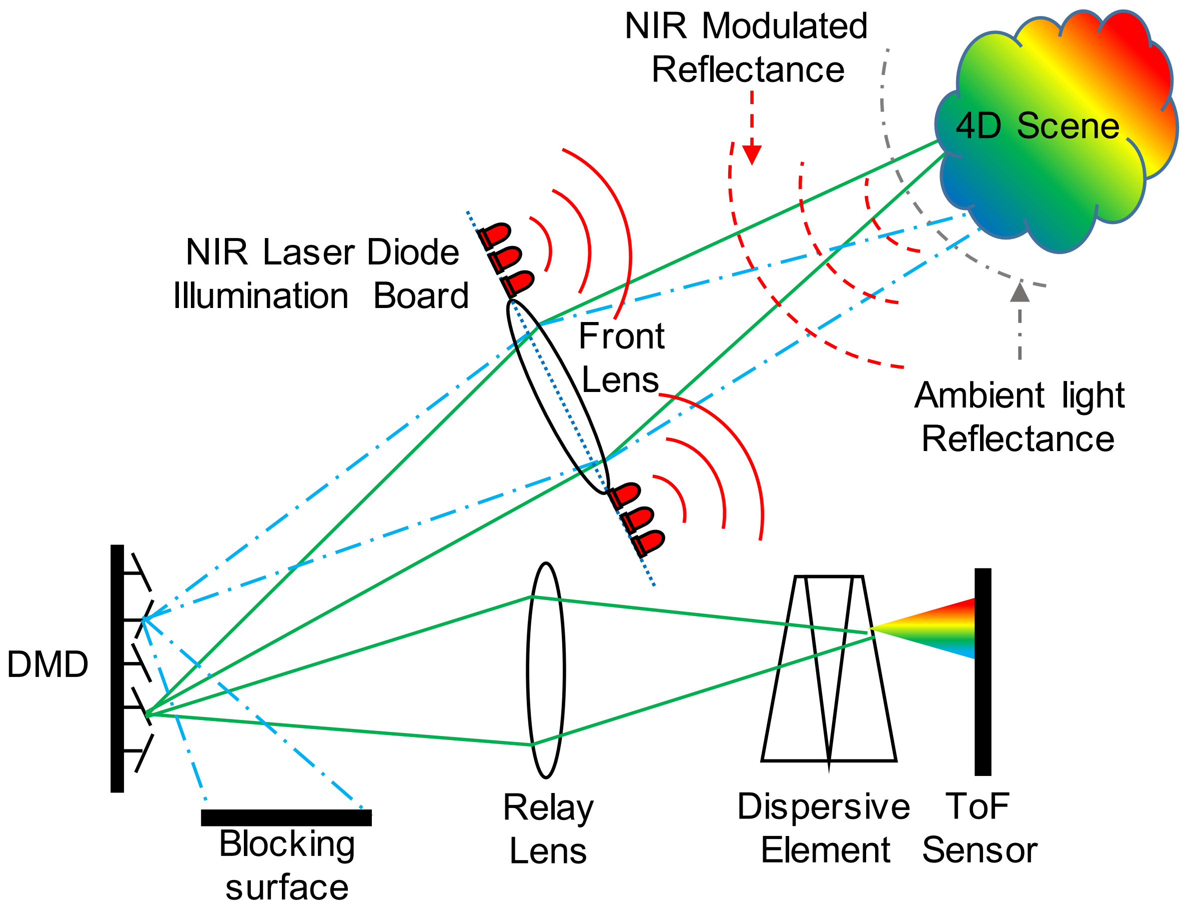


Figure 7.1: Optical design sketch of the single aperture snapshot ToF+Spectral camera. The LED laser diode illumination board is attached to the front lens of the camera which images the 4D scene onto the DMD. Depending on the mode, the DMD will emulate an all-on or a random coded aperture, for the ToF and ambient light capture, respectively. The relay lens transmits the DMD coded and modulated light through a dispersive element, so that monochrome coded, modulated and sheared light is integrated on the ToF sensor.

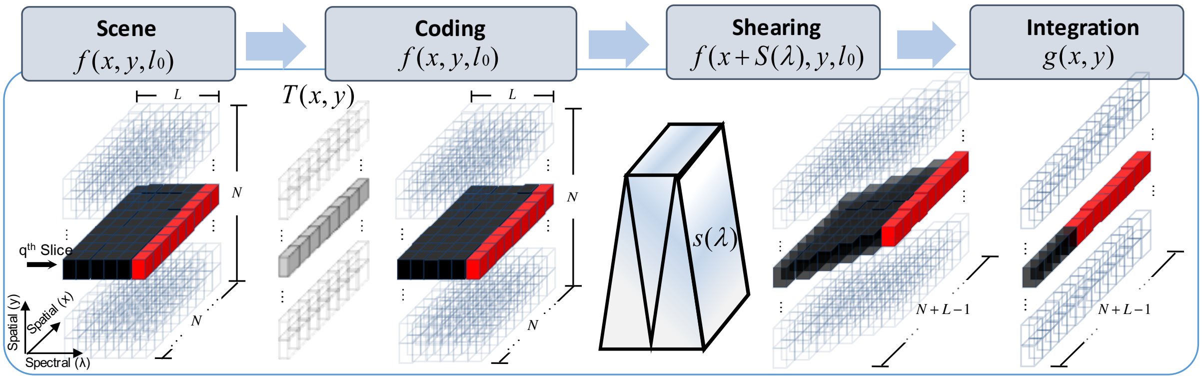
7.2.1 Time-of-Flight Imaging (ToF)

Time-of-flight imaging aims at acquiring the depth image or depth map of the underlying scene using an array of modulated infrared laser diodes as an illumination source. The time delay of the returned light for each spatial location is measured on a two-dimensional (2D) ToF sensor array. According to the laser modulation, there are typically two ways of measuring this time delay [91, 92]. One of them directly measures the round-trip traveling time of the shinned laser light pulses by detection of the first arriving laser photons. Thus the depth value is calculated by $d = c\frac{t}{2}$, where c is the speed of light constant. In this method, the depth measurement accuracy depends on the time measurement resolution. Another ToF imaging method employs a continuous sinusoidal or square-wave laser modulation. Instead of directly measuring the time delay, this method relates the depth information to the phase difference between the emitted and reflected light. In this method, the depth accuracy is determined by the phase measuring capability at a given modulation frequency. Note that the modulation frequency used entails a trade off between depth measurement capability and accuracy; lower frequencies can measure depth at longer distances but with reduced accuracy.

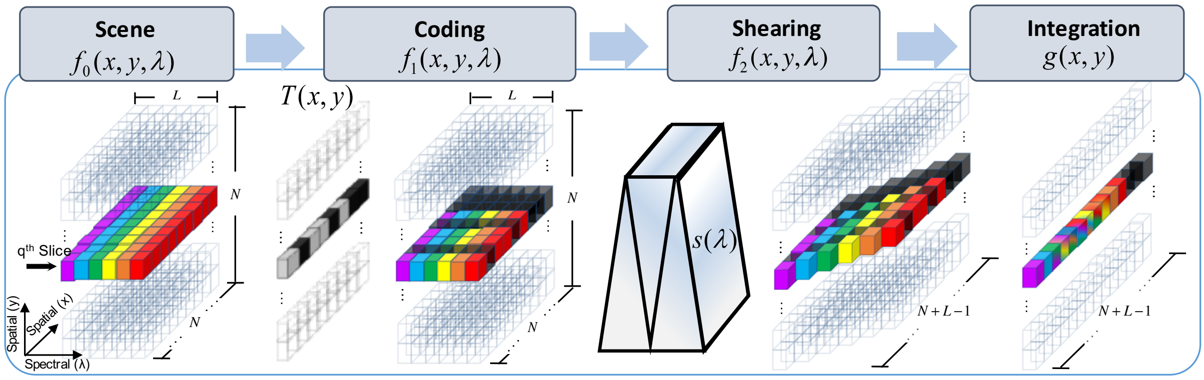
This work uses an Espros EPC660 ToF imaging sensor, where continuous sinusoidal laser modulation is used by default [107]. In phase-based ToF sensors like the Microsoft Canesta [108] or Espros EPC660, distance is measured indirectly as the round-trip phase delay between a modulated light source and synchronized array of gated pixels. To derive a mathematical model for these ToF sensors, we begin by defining $g_{m,n}$ as a single pixel of a traditional $M \times N$ sensor array with discrete row and column coordinate, (m, n), such that that amount of light accumulated over an exposure, τ, is defined according to:

$$g_{m,n} = \sum_{l=0}^{L-1} \frac{1}{\tau} \int_{t=0}^{\tau} A f_{m,n,l} dt = \sum_{l=0}^{L-1} A f_{m,n,l} \tag{7.1}$$

where A represents the amount of ambient light present inside the field of view, and $f_{m,n,l}$ is the reflectance of that light from the target surface, as a function of wavelength l.



(a) ToF operation mode



(b) CSI operation mode

Figure 7.2: Light propagation of the q^{th} slice through the proposed system. (a) In ToF mode, the reflected NIR modulated light is captured after ambient light is subtracted. The single red spectral band represents the reflected NIR light of the LED illumination. The DMD emulates an all-transmissive pattern, such that, no coding is performed. Therefore, propagation through the system causes no multiplexing but shifting in the captured image. (b) In ambient light mode (CSI mode), the incoming light is encoded by the coded aperture pattern $T(x, y)$, then the dispersive element decomposes the light in the corresponding wavelengths, and multiplexing takes place at the sensor.

Adding a modulated light source $B\cos(2\pi ft)$, the captured pixel becomes:

$$g_{m,n} = \sum_{l=0}^{L-1} \frac{1}{\tau} \int_{t=0}^{\tau} (Af_{m,n,l} + Bf_{m,n,l_0} \cos(2\pi ft + \theta))\, dt, \tag{7.2}$$

where θ is the phase delay introduced by the round-trip travel time of the modulated light from source to target and back, and f_{m,n,l_0} is the surface reflectance at the wavelength of the modulated light source (l_0). In our single aperture system shown in Fig. 7.1, the laser illumination board will illuminate the 4D scene with the modulated light, and the reflected signal from the scene will propagate through the optical path of the imaging system as shown in Fig. 7.2(a). Note that all the mirrors of the DMD will be set to on, such that all reflected light from the scene propagates to the sensor.

To represent the gating of pixels, we define the gating function, $r(t)$, as a square wave in the range, $[0, 1]$, with a frequency, f, matched to the modulated light source and with a duty cycle D such that:

$$g_{m,n} = \sum_{l=0}^{L-1} \frac{1}{\tau} \int_{t=0}^{\tau} (Af_{m,n,l} + Bf_{m,n,l_0} \cos(2\pi ft + \theta))\, r(t) dt. \tag{7.3}$$

Denoting the total exposure time as K times the period of one wavelength w (in seconds) of the square wave, then $\tau = K * w$. Given that, $f = 1/w$, then $f = K\tau^{-1}$, such that there are an integer, K, periods of $\cos(2\pi ft)$ and $r(t)$ inside of τ. Therefore, by integrating over one period instead of all of τ, we simplify Eq. (7.3) to:

$$g_{m,n} = \sum_{l=0}^{L-1} \frac{K}{\tau} \int_{t=0}^{\frac{D\tau}{K}} (Af_{m,n,l} + Bf_{m,n,l_0} \cos(2\pi ft + \theta))\, dt. \tag{7.4}$$

Assuming a small D, $g_{m,n}$ can be approximated as:

$$g_{m,n} \approx \sum_{l=0}^{L-1} DAf_{m,n,l} + DBf_{m,n,l_0} \cos(\theta). \tag{7.5}$$

By itself, there is no way to extract θ separate from B, but if, as demonstrated in Fig. 7.3, we collect a sequence of 4 images indexed by ℓ with a phase shift between the

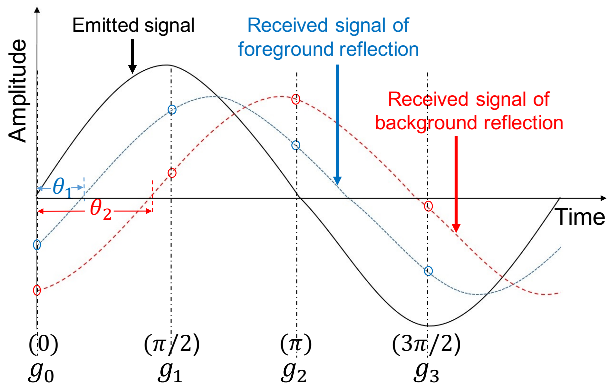


Figure 7.3: ToF sensor raw data g_0 to g_3 are captured in every single $(m,n)^{th}$ pixel. The corresponding phase delays θ_0 to θ_3 are then calculated to obtain the depth map.

modulated light source and camera images such that

$$(g_\ell)_{m,n} \approx \sum_{l=0}^{L-1} DAf_{m,n,l} + DBf_{m,n,l_0} \cos\left(\frac{\pi\ell}{2} + \theta\right), \tag{7.6}$$

then we can define the difference of complements, $g_{0,2}$ and $g_{3,1}$, according to:

$$\begin{aligned}(g_{0,2})_{m,n} &= (g_0)_{m,n} - (g_2)_{m,n}, \\ &= 2DBf_{m,n,l_0} \cos\theta \end{aligned} \tag{7.7}$$

and

$$\begin{aligned}(g_{1,3})_{m,n} &= (g_1)_{m,n} - (g_3)_{m,n}, \\ &= 2DBf_{m,n,l_0} \sin\theta \end{aligned} \tag{7.8}$$

and, finally, extract θ by treating $(g_{0,2})_{m,n}$ and $(g_{1,3})_{m,n}$ as quadrature components such that:

$$\begin{aligned}\theta &= \tan^{-1}\left(\frac{(g_{1,3})_{m,n}}{(g_{0,2})_{m,n}}\right), \\ &= \tan^{-1}\left(\frac{2DBf_{m,n,l_0}\sin\theta}{2DBf_{m,n,l_0}\cos\theta}\right), \\ &= \tan^{-1}\left(\frac{\sin\theta}{\cos\theta}\right). \end{aligned} \tag{7.9}$$

The depth value, at every $(m,n)^{th}$ sensor pixel, $d_{m,n}$, is then calculated as

$$d_{m,n} = \frac{c}{2}\frac{\theta}{2\pi f}, \tag{7.10}$$

where, for the case of the Espros EPC660 sensor, the laser modulation frequency, f, can be selected from 0.625 MHz to 24 MHz, resulting in an unambiguous distance ranging of up to $d_{m,n} = 6.25$ meters for $f = 24$ MHz and up to $d_{m,n} = 240$ meters for $f = 0.625$ MHz.

Separate from range, we can extract the surface reflectance from the modulated light, f_{m,n,l_0}, by looking at the magnitude of the quadrature components such that:

$$f_{m,n,l_0} = \frac{\sqrt{(g_{1,3})^2_{m,n} + (g_{0,2})^2_{m,n}}}{2DB}, \tag{7.11}$$

given that $2DB$ is known a priori. Alternatively, to extract the surface reflectance under ambient light illumination, we can either use $(g_0)_{m,n}$ and $(g_2)_{m,n}$ which are 180^o out-of-phase, as are $(g_1)_{m,n}$ and $(g_3)_{m,n}$, by evaluating their sum,

$$(g_0)_{m,n} + (g_2)_{m,n} = 2\sum_{l=0}^{L-1} DAf_{m,n,l} \tag{7.12}$$

such that:

$$\sum_{l=0}^{L-1} Af_{m,n,l} = \frac{(g_0)_{m,n} + (g_2)_{m,n}}{2D}. \tag{7.13}$$

This last equation demonstrates that phase-based ToF sensors like the Espros EPC660 have an inherent ability to separately measure both ambient (DC) and modulated (AC) light images, and given this inherent ability, the integration of ToF sensors into a

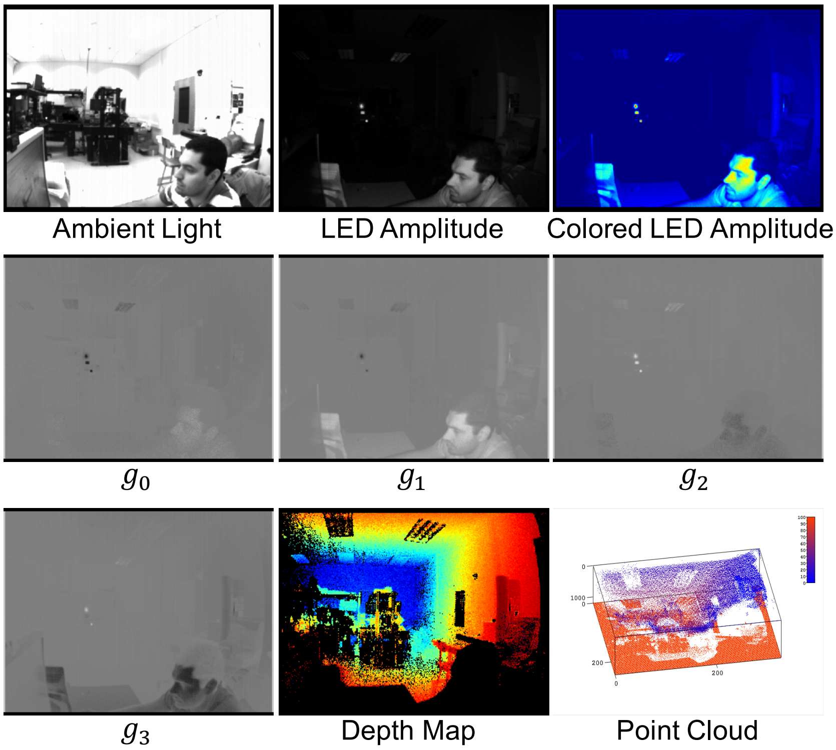


Figure 7.4: Ambient image is recorded in grayscale mode. Four successive frames g_0, g_1, g_2 and g_3 are captured for the depth map and point cloud estimation.

CSI system becomes almost compulsory since the unused DC image can now be used for multispectral imaging while the modulated image reserved for ToF.

To illustrate the entire process, Fig. 7.4 shows the 9 different images provided by the ToF sensor. The original ambient light image of Eq. (7.2) is included as the (top-left) grayscale image. The amplitude of the quadrature components in Eq. (7.11) is illustrated as a (top-center) grayscale and (top-right) pseudo-color image. The sequence of phase shifted images of Eq. (7.6) are illustrated as signed grayscale images as the sequence from (middle-left) to (bottom-left). Finally, the depth estimation is illustrated as a (bottom-center) pseudo-color and (bottom-right) point cloud.

7.2.2 Compressive Spectral Imaging

Compressive spectral imaging (CSI) encompass techniques which sense the spatio-spectral information of a scene by using 2D coded projections, and as few as a single coded projection. The underlying spectral 3D data cube is then recovered using compressive sensing (CS) reconstruction algorithms which assume that the hyperspectral images are sparse in some representation basis. The advantage of CSI is that the required number of measurements needed for reconstruction is far less than that required by traditional scanning methods [1, 58, 109]. In the CSI literature different strategies to attain the coded projections have been proposed. The review articles in [30, 60, 15, 16] provide a number of architectures that attain coded projections. One of the main families in the state-of-the-art CSI architectures are those based on coded apertures. These use a set of lenses, a coded aperture as coding element, a dispersive element and a 2D focal plane array (FPA) sensor, to capture compressive multiplexed random projections of the scene.

As depicted in Fig. 7.2(b) the front-end lens in these architectures effectively creates a light field in front of the coded aperture plane, such that $f_0(x, y, \lambda)$ represents the light ray incident upon the coded aperture at 2-D coordinate (x, y) and having wavelength λ. As light approaches the sensor, the coded aperture blocks or allows the rays of light from reaching the sensor. This aperture effect can be accounted in our model by means of the binary modulation function $T(x, y) = \sum_{m,n} T_{m,n}\text{rect}\left(\frac{x}{\Delta_c} - n, \frac{y}{\Delta_c} - m\right)$, where $T_{m,n} \in [0, 1]$. In order to separate the individual wavelengths of the incoming signal, the system places a dispersive element in front of the image sensor such that $f_1(x, y, \lambda)$ is split into discrete bands spread across L consecutive pixels along the lateral or n dimension in a wavelength dependent manner, indicated by the dispersion function $S(\lambda)$. Accumulating all of the light incident upon a discrete pixel of the sensor with row and column coordinate $[m, n]$ and defining $\chi_{m,n,l}$ as the spectral sensitivity of the sensor for that same discrete pixel, the monochrome pixel, $g_{m,n}$, of the resulting digital image is given by:

$$g_{m,n} = \sum_{l} \chi_{m,n,l} T_{m,n-l} f_{m,n-l,l} + \omega_{m,n} \tag{7.14}$$

where $l = 0, \ldots, L-1$ indexes the pixels affected by the spreading of $f_1(x, y, \lambda)$ as a result of $S(\lambda)$. One would expect $\chi_{m,n,l}$ to be either constant for all pixels and absorb light in the visible light range from 350 to 750nm or selectively absorb red, green, or blue light in a traditional Bayer pattern.

In order to establish a matrix representation consistent with CS, we define $\mathbf{f}$ as a column vector containing $f_{m,n,l}$ for all spectral components l and all pixels m and n ordered first by row, then column, and finally by wavelength. Assuming we have a coded aperture with $M \times N$ pixels ($\mathbf{T} \in \mathcal{R}^{M\times N}$), $\mathbf{f}$ will be MNL in length ($\mathbf{f} \in \mathcal{R}^{MNL}$). Then, we define $\mathbf{g}^k$ as a column vector holding all of the recorded pixel values, $g_{m,n}$, from the monochrome image sensor for the k^{th} image in our sequence of $K \leq L$ snapshots. Multiple snapshots allow to obtain different information from the same scene as different coded patterns can be used. Noting the lateral dispersion of light, the length of $\mathbf{g}^k$ is $M(N + L - 1)$, such that all light is accounted for on the sensor. So while the coded aperture has an $M \times N$ pixel array, the sensor has an extra set of $L - 1$ pixel columns. Assuming that $\mathbf{f}$ stays constant over the K snapshots, we can relate $\mathbf{g}^k$ to $\mathbf{f}$ according to:

$$\mathbf{g}^k = \mathbf{PXT}^k\mathbf{f} = \mathbf{H}^k\mathbf{f}, \tag{7.15}$$

where $\mathbf{H}^k$ represents the combined effects of the dispersion effect $\left(\mathbf{P} \in \mathbb{R}^{M(N+L-1)\times MNL}\right)$, the sensor pixel-wise spectral sensitivity in matrix form $\left(\mathbf{X} \in \mathbb{R}^{MNL\times MNL}\right)$, and the block-diagonal representation of the k^{th} coded aperture $\left(\mathbf{T}^k \in \mathbb{R}^{MNL\times MNL}\right)$.

Alternatively, the sensing process can be expressed as $\mathbf{g} = \mathbf{Hf} = \mathbf{H\Psi}\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\theta}$, such that $\mathbf{g} = \left[(\mathbf{g}^0)^T, \ldots, (\mathbf{g}^{K-1})^T\right]^T$, $\mathbf{H} = \left[(\mathbf{H}^0)^T, \ldots, (\mathbf{H}^{K-1})^T\right]^T$, $\mathbf{A} = \mathbf{H\Psi}$ is the compressive sensing matrix, and $\boldsymbol{\theta} = \mathbf{\Psi}^T\mathbf{f}$ is a vector coordinate of $\mathbf{f}$ in an orthogonal basis $\mathbf{\Psi}$. Conventional basis functions include the Wavelet ($\mathbf{\Psi}_{W2D}$) and discrete cosine transforms, ($\mathbf{\Psi}_{DCT}$) as well as pre-trained dictionaries [75], and bases which simultaneously exploit the sparsity properties along a multidimensional space by doing the Kronecker product of bases used for each of its dimensions [45, 52, 24]. The basis transformation assumes $\mathbf{f}$ to be sparse on $\mathbf{\Psi}$, so that, a small subset of the basis vectors can accurately reconstruct $\mathbf{f}$ with small distortions. An estimation of the spectral input data cube can be attained by solving the regularization problem,

$$\hat{\mathbf{f}} = \mathbf{\Psi}\left(\underset{\boldsymbol{\theta}'}{\operatorname{argmin}}\|\mathbf{g} - \mathbf{A}\boldsymbol{\theta}'\|_2^2 + \tau\|\boldsymbol{\theta}'\|_1\right), \tag{7.16}$$

where $\boldsymbol{\theta}'$ is the estimation of the sparse coefficients, $\|\cdot\|_2^2$ represents the ℓ_2 norm, $\|\cdot\|_1$ the ℓ_1 norm, and τ is a regularization parameter which penalizes the fact of $\boldsymbol{\theta}'$ being sparse, while reducing the error between the estimated $\mathbf{A}\boldsymbol{\theta}'$, and the captured set of compressive measurements $\mathbf{g}$. A large variety of solvers for problem Eq. (7.16) has been proposed, including, the two-step iterative shrinkage/thresholding (TwIST) using the total-variation (TV) as the regularization function [53], the gradient projection for sparse reconstruction (GPSR) [29], Gaussian mixture models (GMM) [54], and denosing algorithms based on approximate message passing (AMP) [33].

7.2.3 Joint Compressive Spectral+ToF Imaging

The previous subsections described the reconstruction of spectral and 3D ranging imagery, performed separately and independently. However, the correlations between the spectral bands and the ToF quadrature components g_0 to g_3, encourages the joint reconstruction of spectral+depth imagery. In addition, side information can be used to improve the reconstruction quality as demonstrated in [3, 102, 82]. The collection of side information usually involves mounting an additional grayscale [3, 102] or RGB camera [82], either side-by-side or along the same path of the multi-spectral system by means of a beam-splitter. The reconstruction algorithm then exploits the correlation between the multi-spectral measurement and the grayscale, or RGB side image. In the proposed single-aperture system, and as a first approach, side information can be extracted from the ToF quadrature components with no additional camera needed. In particular, the surface reflectance of the target scene at the LED modulation wavelength (l_0) can be calculated according to Eq. (7.11). The extracted monochrome image can then be viewed as the ground truth of the scene at l_0-nanometers. Mathematically, we assume that the spectrum of the collected surface reflectance falls into the range of the last reconstructed spectral band. Denote $\mathbf{f}_r$ as the vector form of the collected image band containing the side information. Then the joint sensing process for spectral imaging can be reformulated as,

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{f}_r \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}_b \end{bmatrix} \mathbf{f}, \tag{7.17}$$

where $\mathbf{H}_b = [\mathbf{0}, \mathbf{I}]$, such that $\mathbf{0}$ is an $MN \times MN(L-1)$ zero matrix and $\mathbf{I}$ is an $MN \times MN$ identity matrix.

In addition, since compressive sensing reconstruction algorithms tend to converge to smoothed estimations, the edge information from the ToF amplitude image can also be incorporated as side information across the spectrum, since each spectral band preserves most of the borders in the reflectance image. To this end, an edge detector is first applied to the modulated surface reflectance image to extract a binary edge image $\mathbf{E}$. Then, by noting that the grayscale values of the edge pixels vary with the spectrum wavelengths, the edge image needs to be scaled by means of a calibration data cube captured at each specific wavelength $\mathbf{C}_\ell$. This calibration data cube needs to be captured only once, and off-line, so that it does not affect the frame-rate of the proposed system. Mathematically, the side-information from the edge image can be incorporated in the sensing process, by rewriting Eq. (7.17) as, $[\mathbf{g}^T, \mathbf{f}_e^T]^T = [\mathbf{H}^T, \mathbf{H}_e^T]^T \mathbf{f}$, where $\mathbf{H}_e = [\mathbf{H}_E, \mathbf{I}]$, with $\mathbf{H}_E$ being the concatenation of $L-1$, $MN \times MN$ diagonal matrices, with elements being the vectorized point-wise multiplication between the edge image $\mathbf{E}$ and the corresponding calibrated image $\mathbf{C}_\ell$; $\mathbf{f}_e$ is the estimated grayscale image with the scaled border information.

Side information via a grayscale or an RGB camera can be incorporated at the input of the single-aperture proposed system by use of a beam-splitter. The beam-splitter allows the same path to be shared by the two cameras, so that, registration between the 2 images can be easily achievable. Note that by including this additional camera to the proposed single-aperture ToF architecture, higher flexibility is attained since CSI+ToF+side information can now be jointly captured. Mathematically, this approach to side-information sensing, can be written as $[\mathbf{g}^T, \mathbf{g}_s^T]^T = [\mathbf{H}^T, \mathbf{H}_s^T]^T \mathbf{f}$, where $\mathbf{g}_s \in \{\mathbf{g}_g, \mathbf{g}_c\}$, such that $\mathbf{g}_g \in R^{MN \times 1}$ is the captured grayscale side image in vector form, and $\mathbf{g}_c \in R^{3MN \times 1}$ is the RGB side image in vector form. In addition, $\mathbf{H}_s \in \{\mathbf{H}_g, \mathbf{H}_c\}$, where $\mathbf{H}_g$ is the concatenation of L, $MN \times MN$ all-one diagonal matrices representing the grayscale side sensing matrix, and $\mathbf{H}_c$ represents the RGB side-information sensing matrix, which concatenates L, $3MN \times MN$ diagonal matrices. The entries of $\mathbf{H}_c$ correspond to the discrete response of the filters from the Bayer pattern of the RGB camera.

Side information can also be used to improve the depth map estimation. Typical denoising techniques can be used to smooth the depth imagery, but filtering tends to generate smooth images with blurred target borders. Note however that object borders can be easily obtained either from the quadrature components, the grayscale or RGB side information. With this prior border information, a center weighted median filter (CWMF) [110] can then be applied to the initial depth map. The advantage of using CWMF is that a median filter is applied only on surfaces inside the borders, while the values on the borders are unmodified. This, in turn, guarantees a smoother depth estimation without blurring edges.

Other possible approaches to process the ToF surface reflectance image as side information in spectral imaging include extracting a sparse spatial basis through dictionary learning, as described in [102]. Joint reconstruction of compressive spectral and depth imagery is a promising topic which requires more future research for significant improvements on the reconstruction quality in both spectral and depth dimensions.

7.3 Compressive ToF+Multispectral Testbed Imaging System

The proof-of-concept optical prototype presented in Fig. 7.5 was assembled in our lab. In this system, the imaging arm is composed of an 8-LED laser diode board attached to a VIS-NIR matched achromatic doublet objective lens, and a digital micromirror device; a beam-splitter can be used to enable side information from a secondary sensor as proposed in [3]. The integration arm is composed of a VIS-NIR doublet as the relay lens, a custom VIS-NIR double Amici prism as the dispersive element, and an Espros EPC660 ToF sensor. A target scene is illuminated with a broadband illumination lamp source, and with the modulated LED light with frequencies ranging between $f = 0.625$ MHz and $f = 24$ MHz. The reflected light from the scene is captured by the objective lens and focused onto the mirrors of the DMD image plane, which plays the role of the coded aperture, and onto the secondary sensor, if the beam-splitter is present. When properly aligned, the mirrors of the DMD reflect light into the integration arm, which relays light through a second lens and then through the prism. The dispersed field focuses in the ToF image plane, which imposes its spectral filtering and integration based on its wavelength sensitivity that spans the range between 450 and 1000 nm.

The DMD pixel size is $13.68um$, and the size of the ToF sensor pixels is $20um$. A 3×3 macro-pixel windowing is done in the DMD, and a $2{\times}2$ macro-pixel windowing is realized on the ToF sensor, such that an approximately one-to-one correspondence between the two elements (DMD macro-pitch $= 41.04um$, and ToF macro-pitch $= 40um$) is attained. In order to have dyadic dimensions to perform Wavelet-based representation recovery, and given the DMD and ToF sensor resolution, the coded aperture patterns

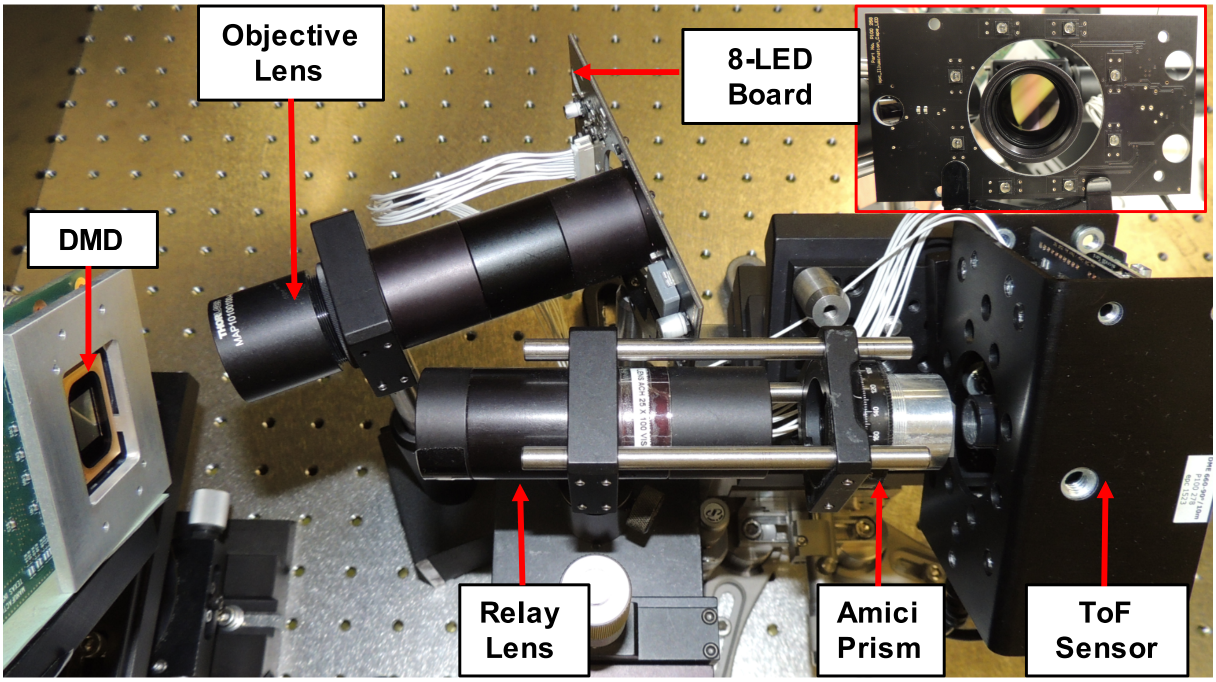


Figure 7.5: Compressive single-aperture ToF+CSI testbed. An objective lens with an attached LED illumination board images the light field of the target scene onto the DMD. After the DMD, a VIS-NIR relay lens transports the coded light through the dispersive element, which decomposes the light field in the corresponding wavelengths being integrated in the ToF sensor device. Side information can be obtained by incorporating a beam-splitter and a secondary sensor as proposed in [3].

exhibit a resolution of 128×128 pixels (384×384 DMD pixels, and 256×256 ToF sensor pixels). More on resolution improvements will be discussed in Section IV.

The prism used in the testbed is a custom designed non-linear double Amici prism, which disperses the wavelength spectrum between 486 - 766 nm onto 16 sensor pixels. If a photomask or a DMD with the same pixel size as that of the sensor is used instead, we would be able to recover 16 spectral band datacubes. However, we are able to recover up to 8 spectral channels, due to the 2×2 macro-pixel windowing. The cut-off wavelengths of these 8 channels are 486, 508, 532, 550, 580, 636, 676, and 766 nanometers. The bandwidth of the spectral channels is non-uniform due to the non-linearity of the prism. The attained reconstruction with the testbed will therefore exhibit $N \times N = 256 \times 256$ pixels of spatial resolution and $L = 8$ spectral channels of spectral resolution. Note that, the spectral resolution can be increased significantly by the use of a diffraction grating.

7.3.1 Calibration of the System

To test the multispectral capabilities of our system, we use the grayscale mode of the ToF sensor in order to capture the raw ambient light data. Each coded aperture used, should be calibrated to determine the experimental impulse response of the system at each wavelength. The calibration is done using a high-reflective white plate as the target scene, illuminating it with monochromatic light at the cut-off wavelengths of the dispersive element. The monochromatic light is obtained from a Tunable Xenon Arc Light source, which provides nanometer resolution. A calibration data cube is thus captured for each coded aperture. Since the testbed setup can reconstruct up to 8 spectral channels, we can capture up to 8 snapshots to estimate the underlying input scene, while still satisfying compression. Therefore, we analyzed the use of $K = 1$, $K = 2$, and $K = 4$ snapshots for the spectral reconstructions, each one using a different realization of a random coded aperture with 50% of transmittance (50% of the elements are 1, and the other are 0). The calibration data cube for 1 of the 4 coded apertures used in the experiments is shown in Fig. 7.6; each image represents the PSF attained on the ToF sensor at the corresponding monochromatic illumination wavelength. These calibrated coded apertures are then used in the reconstruction algorithm as the $\mathbf{H}$ matrix.

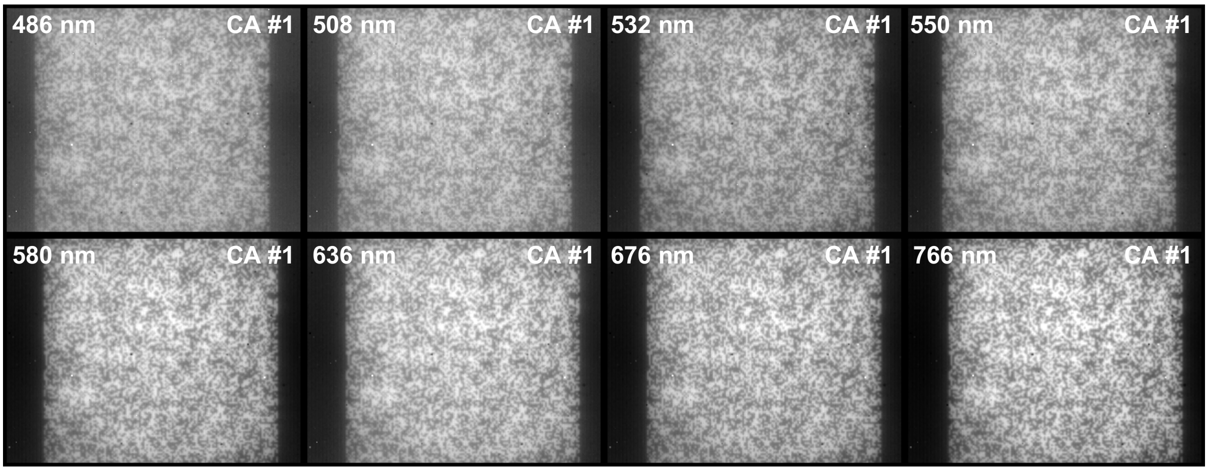


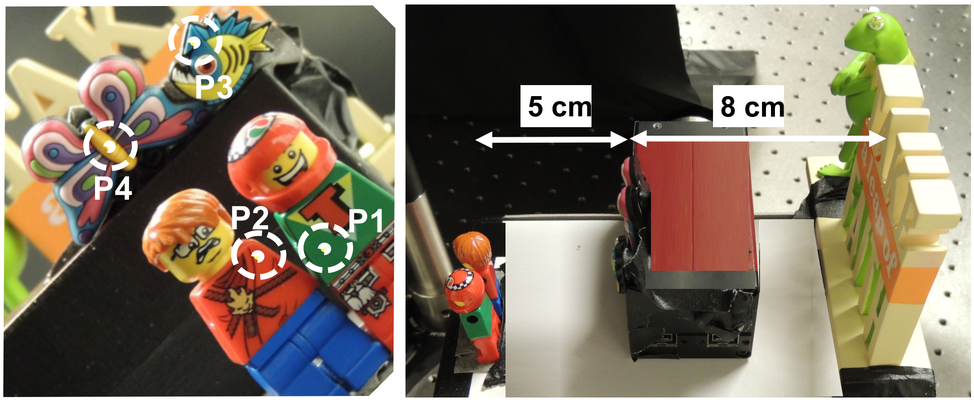
Figure 7.6: Coded aperture calibration patterns as seen by the ToF sensor when illuminated with monochrome light at the given wavelengths. The coded aperture is the result of a binary random variable with 50% transmittance.

7.3.2 Capturing Compressive CSI and ToF Measurements

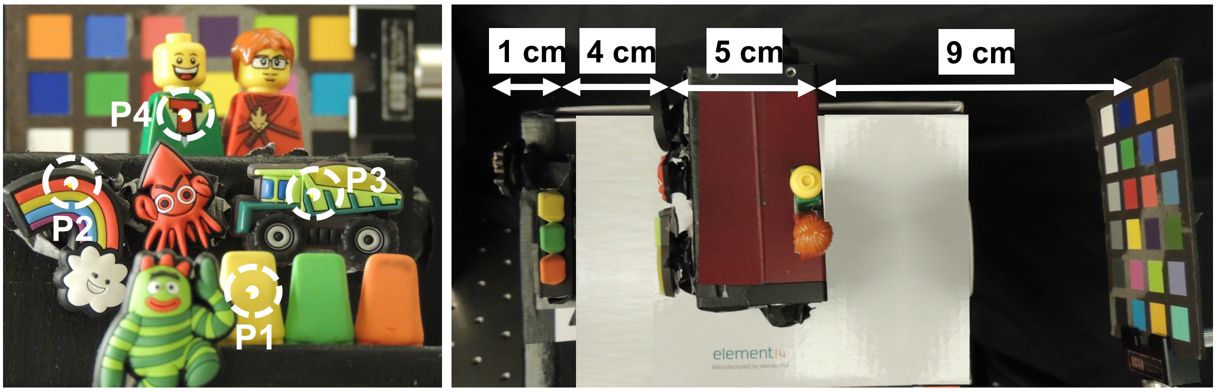
Compressive measurements of a real target use a broadband white illumination source instead of the monochromatic light used in the calibration. To test the efficiency of the images attained with the testbed, we evaluate 2 target scenes, depicted in Fig. 7.7, placed at around 40 and 50 centimeters away from the front-end lens, respectively. The first target scene exhibits 3 depth planes within a range of 13 centimeters, whereas the second scene exhibits 4 depth planes within a range of 20 centimeters.

Figure 7.8 shows an example of the compressive measurements attained with the testbed setup, under the ambient light reading mode, for the two target scenes. Note that the different snapshots integrate approximately the same amount of intensity, restricted by the transmittance of the coded aperture, which was fixed to be 50%. Also note that, the higher the number of snapshots, the higher quality in the reconstruction is expected.

Similarly, in the ToF reading mode, we turn on all the mirrors of the DMD, the LED laser diode illumination board shines the sinusoidal function as explained in Section II. A., the broadband illumination source remain on, and the system then captures the four consecutive frames g_0, g_1, g_2 and g_3 in a single period of exposure time. Figure 7.9 shows the 4 frames captured using a modulation frequency of $f = 12$ MHz, for the 2 target scenes. In this figure, it can be noticed the reciprocity between g_0 and g_2, and g_1 and g_3.



(a) Target scene 1



(b) Target scene 2

Figure 7.7: Target scenes used in the experiments. The first scene is placed around 40 centimeters away from the objective lens, and the second scene is placed around 50 centimeters away. Spectral signatures of the four highlighted points (P1-P4) will be compared in Figs. 7.14 and 7.15.

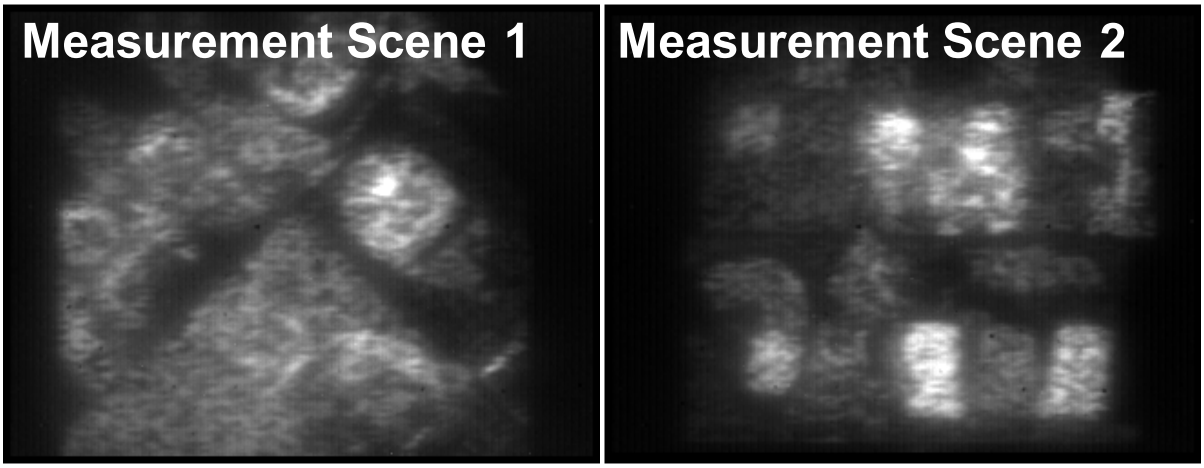
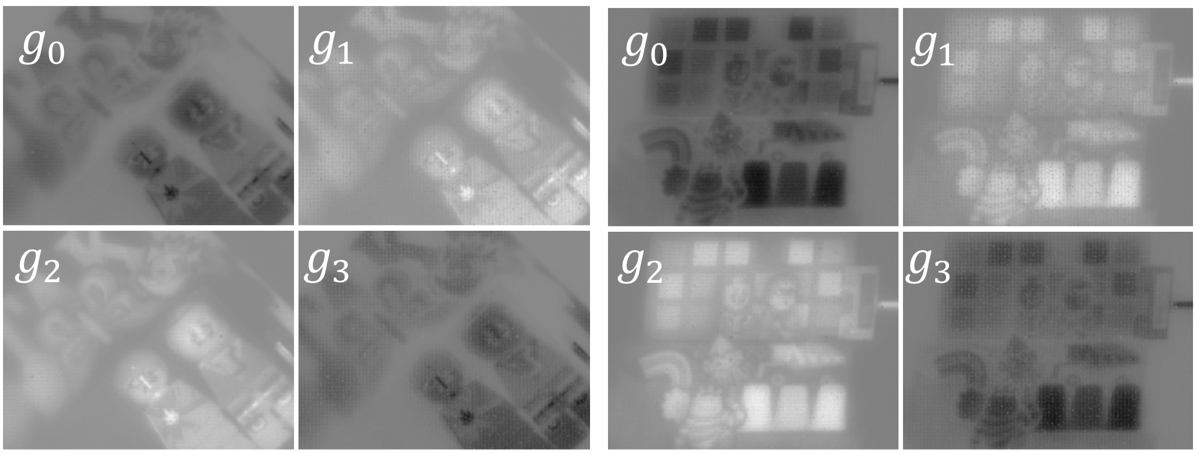


Figure 7.8: Compressive measurements as seen by the ToF sensor when the target scenes are illuminated with broadband fluorescent white light and the DMD exhibit a random coded pattern.



(a) Scene 1 (b) Scene 2

Figure 7.9: Sequence of four images g_0, g_1, g_2, g_3 captured with the testbed, in ToF mode using $f = 12$ MHz, for (a) Scene 1, and (b) Scene 2.

7.3.3 Reconstruction Results for CSI

To reconstruct the spectral data cubes from the coded projections, the compressive-sensing-based gradient projection for sparse reconstruction (GPSR) algorithm [29] was used to solve the optimization problem in Eq. 7.16, with and without the use of side information. In our experiments, we set $\boldsymbol{\Psi} = \boldsymbol{\Psi}_{W2D} \otimes \boldsymbol{\Psi}_{DCT}$ to be the Kronecker product of the 2D Symlet 8 Wavelet transform and the 1D DCT. The grayscale or RGB side information images are emulated by following the same procedure used in the calibration, replacing the white plate with the corresponding scene, and the DMD exhibiting a full-transmissive pattern. We then register and map the attained data cube to the spectral responses $\chi_{m,n,l}$ of the Stingray F033B and F033C cameras, to generate the 256×256 grayscale, and $256 \times 256 \times 3$ color side images.

Figures 7.10 and 7.11 presents the reconstructed data cubes, for the two target scenes in Fig. 7.7, using $K = 1, 2$ and 4 snapshots, mapped to an RGB profile, to test the spatial fidelity. It can be noticed that the reconstruction quality improves with additional snapshots from the single-aperture system, or by incorporating side-information. Note that although the ToF side information impacts the final results, the grayscale and RGB side information have a higher impact with as low as a single snapshot.

The performance gain achieved with the use of side information is illustrated in Figs. 7.12 and 7.13, where 4 out of the 8 reconstructed spectral channels are shown, when $K = 1$ is used. The corresponding band/color discrimination can be contrasted with the RGB reconstructions presented in Figs. 7.10 and 7.11. For instance, for scene 1, the green chest of the target on the right is well characterized at the 532 nm wavelength, whereas its red helmet clearly appears at the 676 nm wavelength. For scene 2, the yellow eraser at the bottom is well reconstructed at 580 nm, the red octopus is discriminated by the 676 nm wavelength, and the green chest of the Lego toy is again well characterized at the 532 nm wavelength. Here, it can be confirmed the boost in quality entailed by the use of side information.

To evaluate the spectral reconstruction results, Figs. 7.14 and 7.15 present the spectral footprint of 4 different points of the 2 target scenes, respectively, denoted as P1, P2, P3 and P4 in Fig. 7.7, which are compared against reference spectra measured with

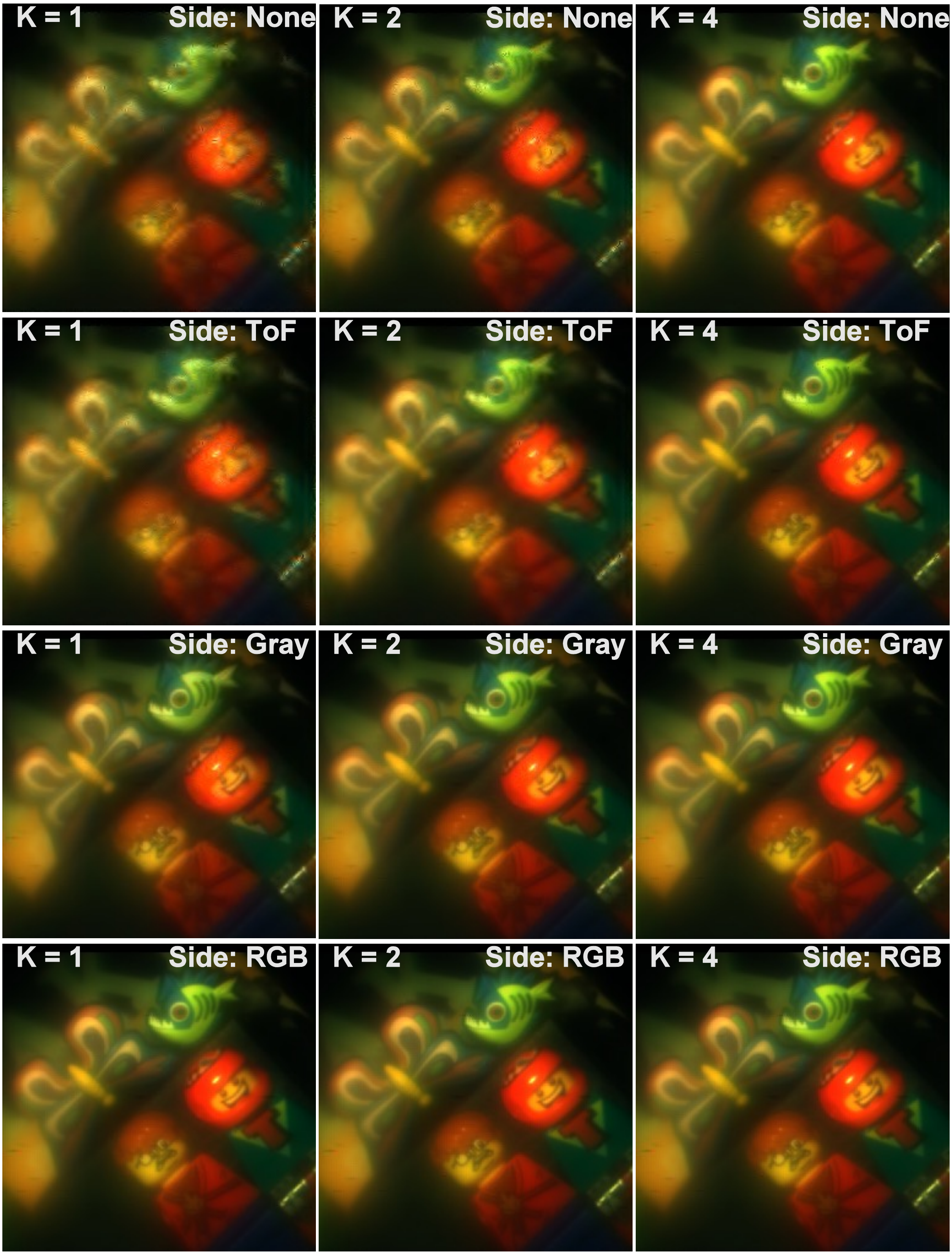


Figure 7.10: Reconstructed data cubes mapped to RGB, when $K = 1, 2$ and 4 CSI snapshots are captured, using none, grayscale or RGB side information for Scene 1.

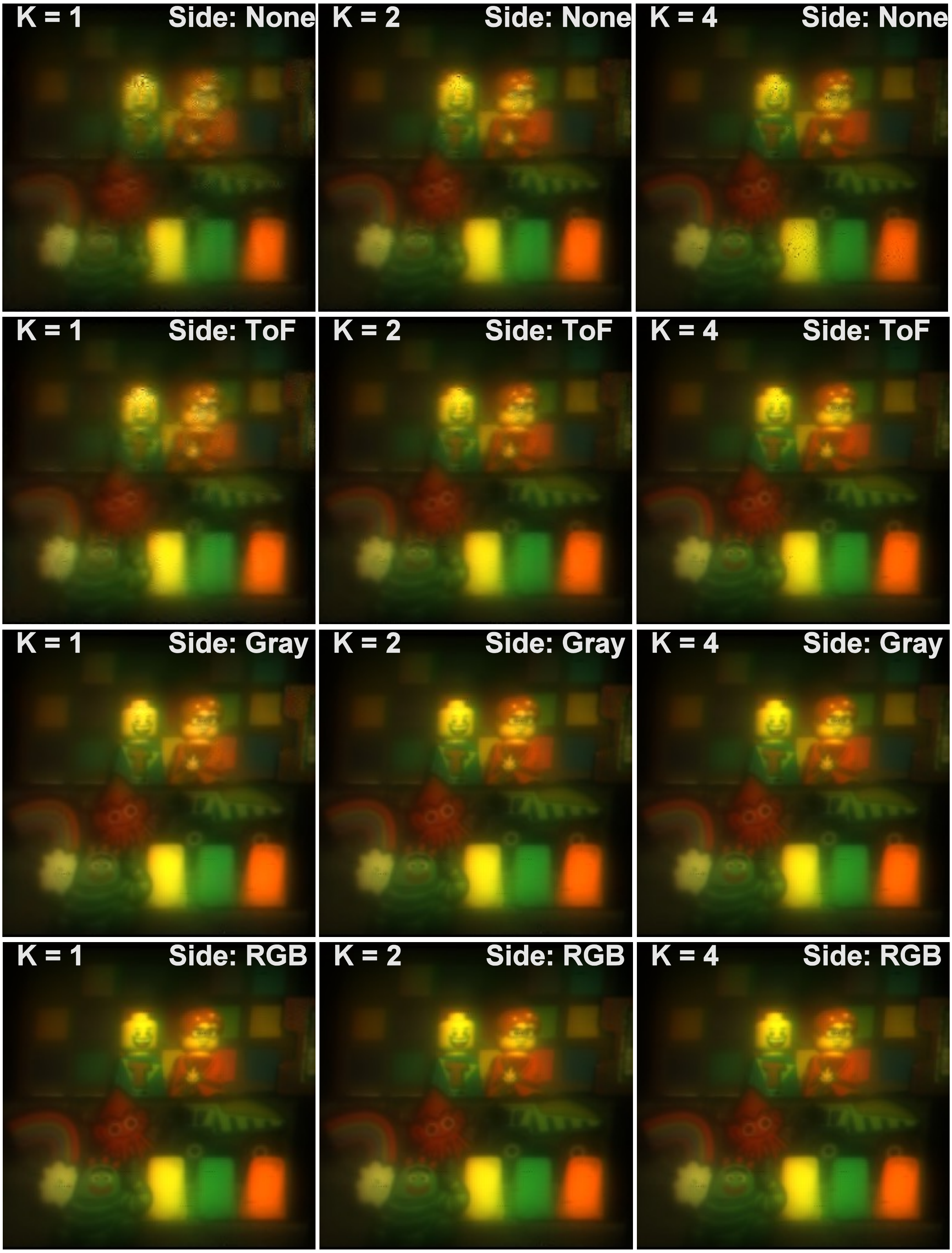


Figure 7.11: Reconstructed data cubes mapped to RGB, when $K = 1, 2$ and 4 CSI snapshots are captured, using none, grayscale or RGB side information for Scene 2.

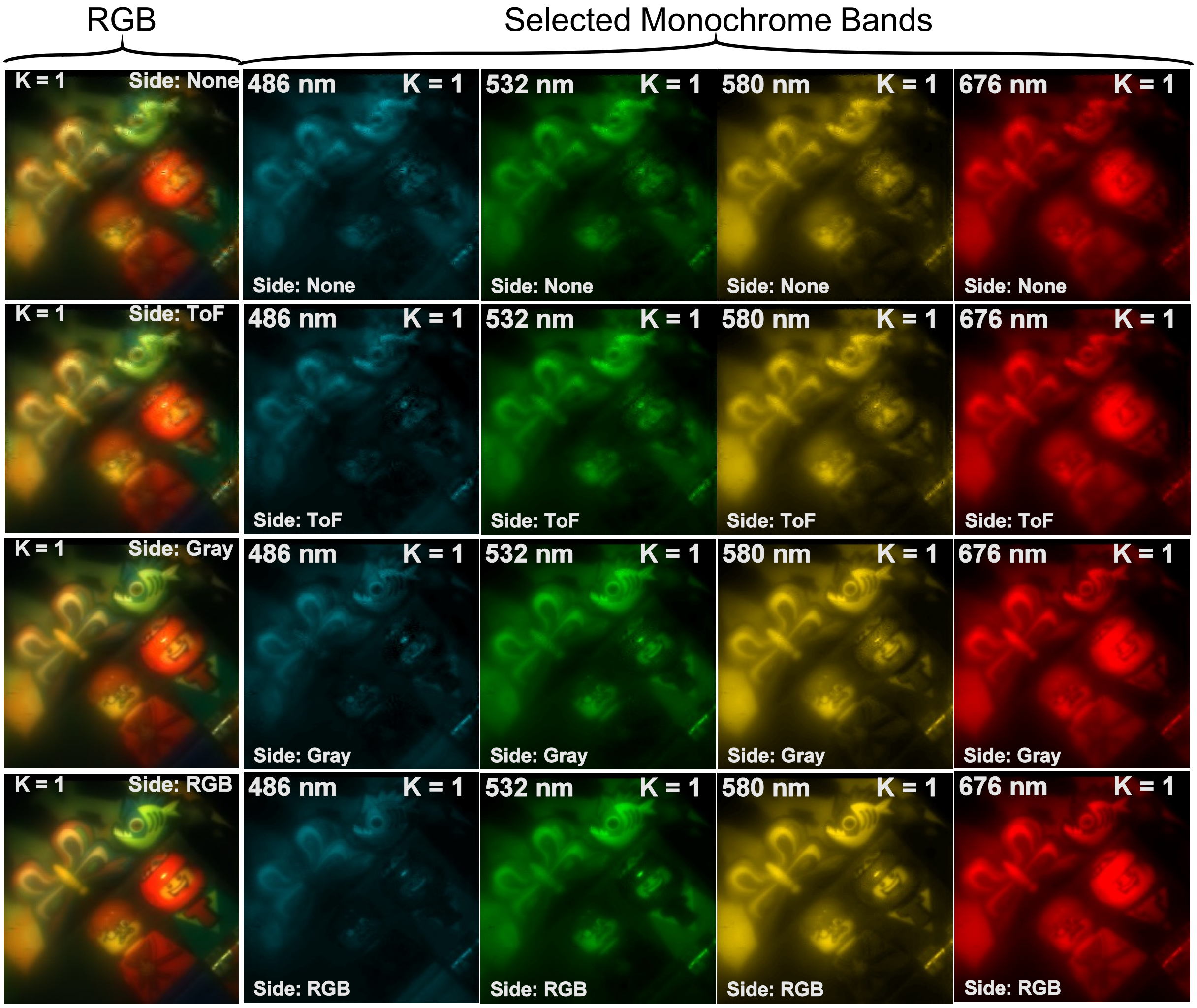


Figure 7.12: Reconstruction results of 4 out of the 8 spectral bands, for a single snapshot, using none, grayscale or RGB side information for Scene 1.

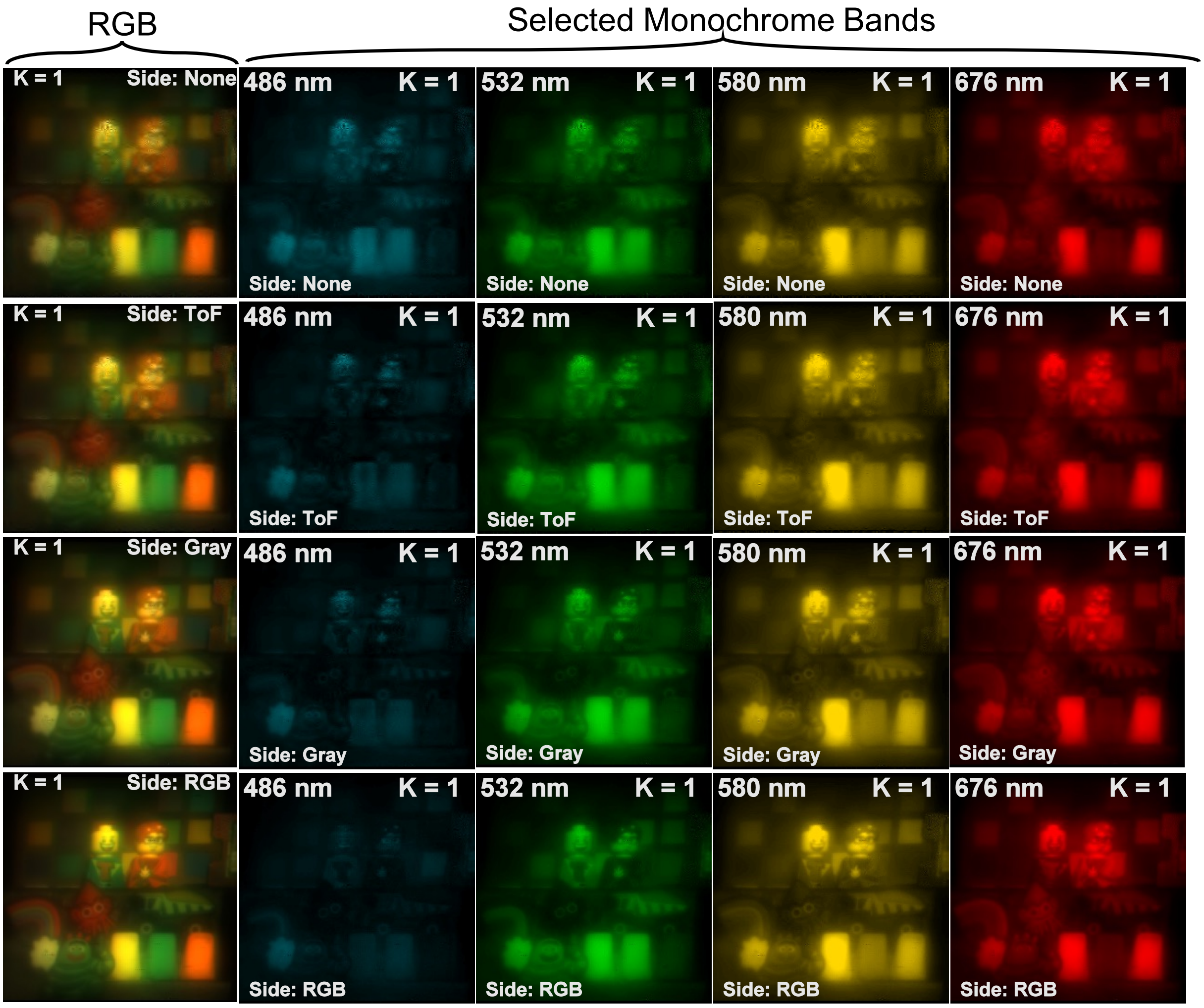


Figure 7.13: Reconstruction results of 4 out of the 8 spectral bands, for a single snapshot, using none, grayscale or RGB side information for Scene 2.

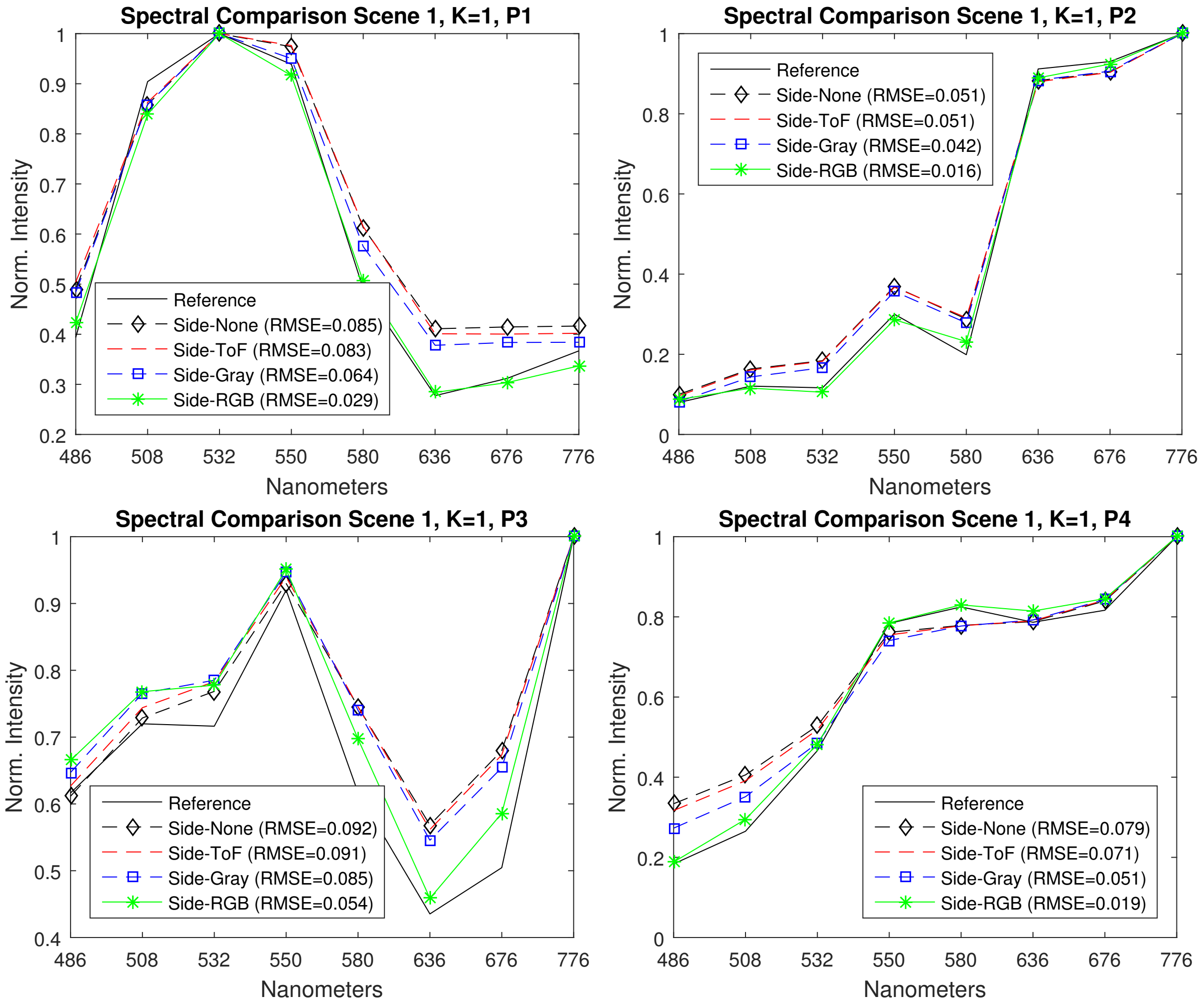


Figure 7.14: Comparison of the spectrum reconstructions when $K = 1$ snapshots are captured with no side information, and with side information from ToF, grayscale or RGB camera. Four different points (P1 to P4) from the first target scene in Fig. 7.7(a) were measured by an spectrometer and compared against the reconstructed data cubes.

an spectrometer (Ocean Optics USB2000+). All the spectral curves are averaged in a 5×5 window, and normalized to the maximum value in their respective curves. The root-mean-square error (RMSE) between the reconstructed and the reference spectra are included in each subplot for ease of interpretation. It can be noticed that the curves from the use of RGB side information attain the best fit to the reference spectrum, followed by the curves from the reconstructions that use a grayscale side image, and these closely followed by the ones that use ToF, and the ones that do not use any, side information.

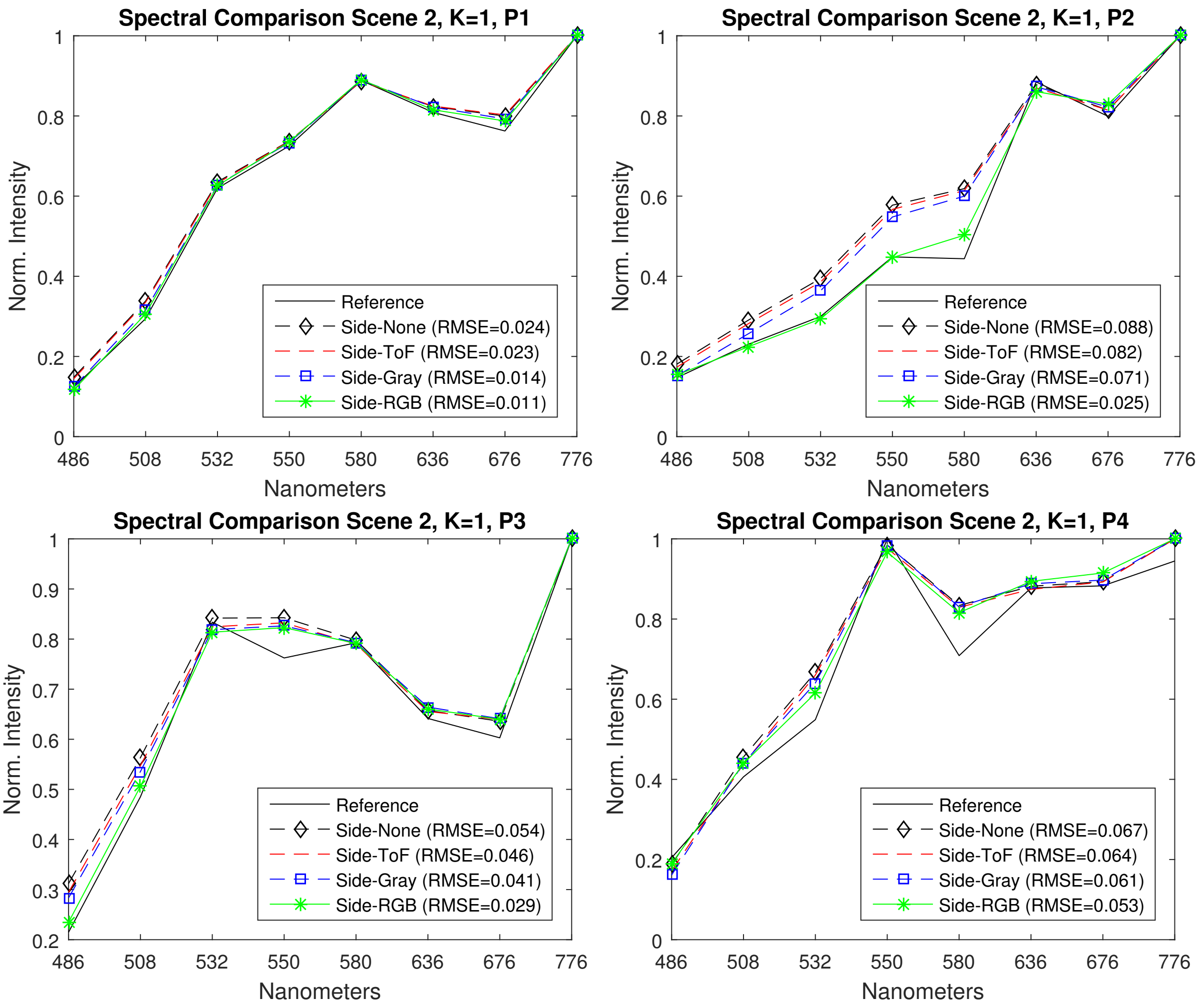


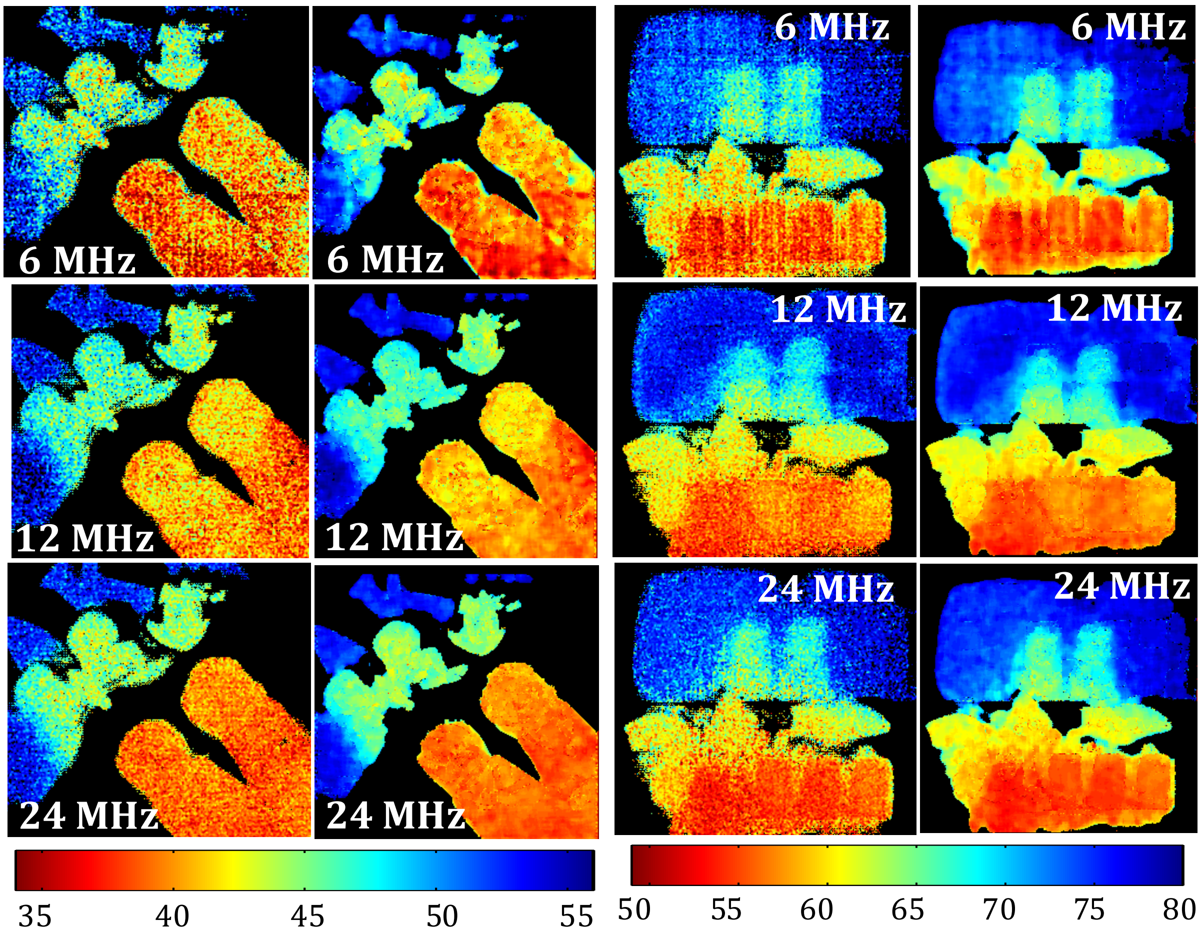
Figure 7.15: Comparison of the spectrum reconstructions when $K = 1$ snapshots are captured with no side information, and with side information from ToF, grayscale or RGB camera. Four different points (P1 to P4) from the second target scene in Fig. 7.7(b) were measured by an spectrometer and compared against the reconstructed data cubes.

7.3.4 Reconstruction Results for ToF

To evaluate the 3D ranging estimation of the proposed testbed camera, the depth map is estimated using Eq. (7.10), and the sequence of 4 frames g_0 to g_3 captured in the ToF mode, shown in Fig. 7.9. Given the reduced range in which the toys are placed within the 2 scenes, we use different frequencies f for the modulation LED laser diode light, in order to evaluate the precision of the system. Recall that the target scenes are placed at around 40 and 50 cm away from the objective lens, respectively. Figure 7.16 depicts the estimated depth maps using three modulation frequencies, $f = 6$ MHz, $f = 12$ MHz and $f = 24$ MHz, for the 2 target scenes, comparing the use of the CWMF described in Section II.C. Note that lower frequencies can reach longer ranges, but the depth precision will diminish. In Fig. 7.16(a) it can be seen that objects in the target scene 1 can be distinguished at 3 different depth planes, but $f = 24$ MHz attains a cleaner depth map than the other frequencies. Similarly in Fig. 7.16(b) the objects from the target scene 2 can be distinguished along 4 different depth planes, and again $f = 24$ MHz attains the most precise depth map. By comparing first against second column within Figs. 7.16(a) and 7.16(b), it can be noticed the boost in quality entailed by the use of the CWMF, since cleaner, and more precise, depth maps are achieved.

7.4 Comparison with alternative solutions and limitations of the current system

Passive and active illumination 3D ranging systems have, indeed, their own pros and cons. Current off-the-shelf ToF sensors have low pixel count. However, these sensors are still in its early stages, and improving rapidly, as the technology becomes more popular for consumer applications. The proposed methods in this manuscript are general and will be applicable on ToF sensors as they continue to improve. Compared with stereo or structured light based depth cameras, ToF cameras exhibit several advantages. A big disadvantage of passive systems is that they cannot derive depth without depth cues created by texture; this is known as the depth stereo problem. Passive ranging methods, such as stereo imaging, and structured light imaging, rely on triangulation to interpolate depth values in the spaces between edges. The problem with triangulation is that it is very difficult to work over long ranges since the size of the triangle has to grow proportionally.



(a) (Left) Depth from raw data. (Right) Depth using CWMF

(b) (Left) Depth from raw data. (Right) Depth using CWMF

Figure 7.16: Estimated depth maps using frequencies $f = 6$ MHz, $f = 12$ MHz, and $f = 24$ MHz. (a) Scene 1. (b) Scene 2. Units are in centimeters.

They also have to deal with problems of occlusion where target surfaces are visible to the camera but not to the projector. ToF sensors, in contrast, do not suffer of occlusion problems, since the projector can wrap around the lens, thus the projector and camera have basically the same line of sight; and also do not require any texture or clue to estimate the depth. Moreover, ToF sensors can reach longer ranges by just changing the frequency and amplitude of the modulated light, they are lightweight, power efficient, and can collect full frame images with a single laser pulse, thus admitting very high frame-rates, and requiring no moving parts.

Regarding the single-aperture/single-sensor architecture, we expect that multi-mode (spectral + range) image sensors, such as the ToF used in our manuscript, will enable smaller form factors, lower power consumption and lighter weight imaging systems. This will lead to miniaturization and, in turn, to promote their use in commercial platforms.

While the benefits of the system in this chapter are limited by the low pixel count of the ToF sensor, it does not, in reality, suffer from frame-rate reduction since the ambient light image and the modulated light amplitude can be jointly estimated from the same ToF quadrature components, using Eqs. (7.7) - (7.13). Note however, that although we can estimate the 2 images from the quadrature components, we require at least 2 snapshots with complementary coded apertures, such that a pure (non-codified) modulated amplitude can be attained, in order to estimate the depth map per pixel. Note also that complementary snapshots will also boost the reconstruction quality of the spectral data cube estimation.

7.5 Conclusions and Future work

A single aperture compressive snapshot multispectral+3D ToF ranging camera is proposed using a commodity ToF sensor, such as the Espros EPC660. The ambient light and modulated light dual-mode reading capability of the ToF sensor is exploited to jointly extract the characteristics of a target scene along the depth, space and spectral dimensions. It was shown that increasing the number of snapshots or by jointly reconstructing spectral and 3D ranging using side information, greatly improves the quality of reconstruction.

Given the pixel count of the current ToF sensor, multispectral+depth (MS+D) datacubes with 256×256 pixels of spatial resolution, 8 spectral channels and depth estimation at centimeter level, were recovered using compressive snapshots such that no scanning processes were required. The spatial resolution of the system can be further improved by using super-resolution techniques and high-resolution coded apertures [30, 111]. The spectral resolution can be enhanced by changing the dispersive element by a diffraction grating, so that hyperspectral+depth (HS+D) imagery could be attained. Further improvements along the spatial, spectral and depth dimensions can be achieved by optimizing the entries of the coded apertures [36], or by using fusion techniques that further exploit the side information.

BIBLIOGRAPHY

[1] A. A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47:B44–B51, 2008.

[2] Cave multispectral database. (accessed 05-01-2017). available: http://www.cs.columbia.edu/CAVE/databases/multispectral/.

[3] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 54(4):848–858, Feb 2015.

[4] D. C. Heinz and C. Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):529–545, Mar 2001.

[5] R. M. Levenson and J. R. Mansfield. Multispectral imaging in biology and medicine: Slices of life. *Cytometry Part A*, 69A(8):748–758, 2006.

[6] R. A. Schultz, T. Nielsen, J. R. Zavaleta, R. Ruch, R. Wyatt, and H. R. Garner. Hyperspectral imaging: A novel approach for microscopic analysis. *Cytometry*, 43(4):239–247, 2001.

[7] M. Eismann. *Hyperspectral Remote Sensing*. SPIE, SPIE Press, 2012.

[8] A. Plaza et. al. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113:S110 – S122, 2009.

[9] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36, June 2013.

[10] G. D. Boreman R. G. Sellar. Classification of imaging spectrometers for remote sensing applications. *Optical Engineering*, 44:44 – 44 – 3, 2005.

[11] M. E. Schaepman, S. L. Ustin, A. J. Plaza, T. H. Painter, J. Verrelst, and S. Liang. Earth system science related imaging spectroscopyan assessment. *Remote Sensing of Environment*, 113(Supplement 1):S123 – S137, 2009. Imaging Spectroscopy Special Issue.

[12] F. A. Kruse, J. W. Boardman, and J. F. Huntington. Comparison of airborne hyperspectral data and eo-1 hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1388–1400, June 2003.

[13] K. Hege, D. O'Connell, W. Johnson, S. Basty, and E. Dereniak. Hyperspectral imaging for astronomy and space surviellance. In *Proc. SPIE's 48th Annu. Meeting Opt. Sci. Technol*, volume 5159, pages 380 – 391, 01 2004.

[14] N. Gat. Imaging spectroscopy using tunable filters: a review. In *Proc. SPIE*, volume 4056, pages 50–64, 2000.

[15] N. Hagen, R. T. Kester, L. Gao, and T. S. Tkaczyk. Snapshot advantage: a review of the light collection improvement for parallel high-dimensional measurement systems. *Optical Engineering*, 51(11):111702–1–111702–7, 2012.

[16] N. Hagen and M. W. Kudenov. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):090901–090901, 2013.

[17] D. L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52:1289–1306, 2006.

[18] S. Gleichman and Y. C. Eldar. Blind compressed sensing. *IEEE Trans. Inf. Theory*, 57(10):6958–6975, 2011.

[19] E. J. Candes and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[20] R. Willett, R. Marcia, and J. Nichols. Compressed sensing for practical optical imaging systems: a tutorial. *Optical Engineering*, 50(7), July 2011.

[21] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Trans. on Inform. Theory*, 54(5):2210–2219, May 2008.

[22] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Opt. Express*, 15(21):14013–14027, 2007.

[23] D. Kittle, K. Choi, A. A. Wagadarikar, and D. J. Brady. Multiframe image estimation for coded aperture snapshot spectral imagers. *Appl. Opt.*, 49(36):6824–6833, Dec 2010.

[24] M. F. Duarte and R. G. Baraniuk. Kronecker compressive sensing. *IEEE Trans. on Image Processing*, 21(2):494–504, 2012.

[25] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single pixel imaging via compressive sampling. *IEEE signal processing magazine*, March 2008.

[26] J. Shihao, X. Ya, and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008.

[27] J. L. Paredes and G. R. Arce. Compressive sensing signal reconstruction by weighted median regression estimates. *IEEE Transactions on Signal Processing*, 59(6):2585–2601, June 2011.

[28] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale l1_regularized least squares. *IEEE J. of Selected Topics in Signal Processing*, pages 606–617, 2007.

[29] M. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. of Selected Topics in Signal Processing*, 1(4):586–597, 2007.

[30] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2014.

[31] D. Needell and J. Tropp. Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[32] S. Wright, R. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions Signal Processing*, 57(8):2479–2493, 2009.

[33] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. R. Arce. Compressive hyperspectral imaging via approximate message passing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):389–401, 2016.

[34] A. Pelagotti, A. D. Mastio, A. D. Rosa, and A. Piva. Multispectral imaging of paintings. *IEEE Signal Processing Magazine*, 25(4):27–36, July 2008.

[35] Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather. Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Opt. Lett*, 36:2692–2694, 2011.

[36] H. Arguello and G. R. Arce. Rank minimization code aperture design for spectrally selective compressive imaging. *IEEE Transactions on Image Processing*, 22(3):941–954, March 2013.

[37] J. M. Eichenholz and J. Dougherty. Ultracompact fully integrated megapixel multispectral imager. In *Proc.SPIE*, volume 7218, pages 7218 – 7218 – 10, 2009.

[38] P. A. Stupar, R. L. Borwick III, J. F. DeNatale, P. H. Kobrin, and W. J. Gunning. Mems tunable fabry-perot filters with thick, two sided optical coatings. *Solid-State Sensors, Actuators and Microsystems Conference, 2009. TRANSDUCERS 2009. International*, pages 1357 – 1360, June 2009.

[39] Y. Bando, B. Chen, and T. Nishita. Extracting depth and matte using a color-filtered aperture. *ACM Trans. Graph.*, 27(5):134:1–134:9, December 2008.

[40] S. Kim, E. Lee, M. H. Hayes, and J. Paik. Multifocusing and depth estimation using a color shift model-based computational camera. *IEEE Transactions on Image Processing*, 21(9):4152–4166, Sept 2012.

[41] H. Arguello and G. R. Arce. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908, 2014.

[42] H. Arguello, H. Rueda, Y. Wu, D. W. Prather, and G. R. Arce. Higher-order computational model for coded aperture spectral imaging. *Applied Optics*, 52(10), 2013.

[43] H. Arguello, C. V. Correa, and G. R. Arce. Fast lapped block reconstructions in compressive spectral imaging. *Applied Optics*, 52(10), 2013.

[44] T.-H. Tsai, P. Llull, X. Yuan, L. Carin, and D. J. Brady. Spectral-temporal compressive imaging. *Opt. Lett.*, 40(17):4054–4057, Sep 2015.

[45] Y. August, C. Vachman, Y. Rivenson, and A. Stern. Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains. *Appl. Opt.*, 52(10):D46–D54, Apr 2013.

[46] D. Knipp, R. A. Street, H. Stiebig, M. Krause, J.-P. Lu, S. Ready, and J. Ho. Vertically integrated thin film color sensor arrays for imaging applications. *Opt. Express*, 14(8):3106–3113, Apr 2006.

[47] P. L. P. Dillon, A. T. Brault, J. R. Horak, E. Garcia, T. W. Martin, and W. A. Light. Fabrication and performance of color filter arrays for solid-state imagers. *IEEE Journal of Solid-State Circuits*, 13(1):23–27, Feb 1978.

[48] S. J. Hood and S. G. Chamberlain. Color-filter arrays for silicon solid-state image arrays sensors. *Canadian Journal of Physics*, 69:543–548, 1991.

[49] D. Hays, A. Zribi, S. Chandrasekaran, S. Goravar, S. Maity, L. R. Douglas, K. Hsu, and A. Banerjee. A hybrid mems-fiber optic tunable fabry-perot filter. *Journal of Microelectromechanical Systems*, 19(2):419–429, April 2010.

[50] P. Buchsbaum. Method of making dichroic filter arrays, 1998. US Patent 5,711,889.

[51] P. Buchsbaum. Dichroic filter detector arrays for spectroscopic imaging, 2003. US Patent 6,638,668.

[52] H. Rueda, H. Arguello, and G. R. Arce. Dmd-based implementation of patterned optical filter arrays for compressive spectral imaging. *J. Opt. Soc. Am. A*, 32(1):80–89, Jan 2015.

[53] J. Bioucas-Dias and M. Figueiredo. A new twist: Two-step iterative shrinking/thresholding algorithms for image restoration. *IEEE Transactions Image Processing*, 16(12):2992–3004, 2007.

[54] A. Rajwade, D. Kittle, T.-H. Tsai, D. Brady, and L. Carin. Coded hyperspectral imaging and blind compressive sensing. *SIAM Journal on Imaging Sciences*, 6(2):782–812, 2013.

[55] H. Arguello and G. R. Arce. Code aperture optimization for spectrally agile compressive imaging. *J. Opt. Soc. Am. A*, 28(11):2400–2413, November 2011.

[56] Y. Rivenson and A. Stern. Compressed imaging with a separable sensing operator. *IEEE Signal Process. Lett.*, 16(6):449–452, 2009.

[57] C. V. Correa, H. Arguello, and G. R. Arce. Snapshot colored compressive spectral imager. *J. Opt. Soc. Am. A*, 32(10):1754–1763, 2015.

[58] H. Rueda, D. Lau, and G. R. Arce. Multi-spectral compressive snapshot imaging using rgb image sensors. *Opt. Express*, 23(9):12207–12221, May 2015.

[59] G. Lu and B. Fei. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, 19(1):010901:1–23, 2014.

[60] G. R. Arce, H. Rueda, C. V. Correa, A. Ramirez, and H. Arguello. *Snapshot Compressive Multispectral Cameras*, pages 1–22. John Wiley & Sons, Inc., 2017.

[61] H. Rueda, H. Arguello, and G. R. Arce. Compressive spectral testbed imaging system based on thin-film color-patterned filter arrays. *Appl. Opt.*, 55(33):9584–9593, Nov 2016.

[62] C. V. Correa, H. Arguello, and G. R. Arce. Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging. *J. Opt. Soc. Am. A*, 33(12):2312–2322, Dec 2016.

[63] Z. T. Harmany, R. F. Marcia, and R. M. Willett. Compressive coded aperture keyed exposure imaging with optical flow reconstruction. *CoRR*, abs/1306.6281, 2013.

[64] R. F. Marcia and R. M. Willett. Compressive coded aperture superresolution image reconstruction. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 833–836, March 2008.

[65] J. M. Duarte-Carvajalino and G. Sapiro. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Transactions on Image Processing*, 18(7):1395–1408, July 2009.

[66] X. Lin, Y. Liu, J. Wu, and Q. Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. on Graphics*, 33(6):233, 2014.

[67] D. J. Brady. *Optical Imaging and Spectroscopy.* Wiley, John and Sons, 2009.

[68] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, March 2011.

[69] D. L. Lau and R. Yang. Real-time multispectral color video synthesis using an array of commodity cameras. *Real-Time Imaging*, 11(2):109–116, April 2005.

[70] B. E. Bayer. Color imaging array, 1976. U.S. Patent 3,971,065.

[71] P. M. Hubel, J. Liu, and R. J. Guttosch. Spatial frequency response of color image sensors: Bayer color filters and foveon x3. In *Proc.SPIE*, volume 5301, pages 5301 – 5301 – 6, 2004.

[72] R. B. Merrill. Color separation in an active pixel cell imaging array using a triple-well structure, 1999. U.S. Patent 5,965,875.

[73] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52:489–509, 2006.

[74] Z. Wang and G. Arce. Variable density compressed image sampling. *IEEE Trans. Image Processing*, 19(1):264, 2010.

[75] X. Yuan, P. Llull, D. J. Brady, and L. Carin. Tree-structure bayesian compressive sensing for video. *arXiv:1410.3080*, 2014.

[76] G. Riutort-Mayol, Á. Marqués-Mateu, A. E. Seguí, and J. L. Lerma. Grey level and noise evaluation of a foveon x3 image sensor: A statistical and experimental approach. *Sensors*, 12:10339–10368, 2012.

[77] A. Foi, S. Alenius, V. Katkovnik, and K. Egiazarian. Noise measurement for raw-data of digital imaging sensors by automatic segmentation of nonuniform targets. *IEEE Sensors Journal*, 7(10):1456–1461, Oct 2007.

[78] M. Dinguirard and P. N. Slater. Calibration of space-multispectral imaging sensors: A review. *Remote Sensing of Environment*, 68(3):194 – 205, 1999.

[79] Á. Marqués-Mateu, J. L. Lerma, and G. Riutort-Mayol. Statistical grey level and noise evaluation of foveon x3 and cfa image sensors. *Optics & Laser Technology*, 48(Supplement C):1 – 15, 2013.

[80] SIGMA Corporation. Foveon quattro technology, 2016. Accessed: March, 2015. Available: http://www.sigma-global.com/en/cameras/dp-series/technology/.

[81] C. Ma, X. Cao, R. Wu, and Q. Dai. Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system. *Opt. Lett.*, 39(4):937–940, Feb 2014.

[82] X. Yuan, T. H. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin. Compressive hyperspectral imaging with side information. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):964–976, Sept 2015.

[83] J. A. Tropp and S. J. Wright. Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE*, 98(6):948–958, June 2010.

[84] Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather. Study of an nir digital micromirror device-based snapshot spectral imaging system. In *Proc.SPIE*, volume 8254, pages 8254 – 8254 – 8, 2012.

[85] B. M. Kaylor, A. Ashok, E. M. Seger, C. J. Keith, and R. R. Reibel. Dynamically programmable, dual-band computational imaging system. In *Imaging and Applied Optics Technical Papers*, page CM4B.3. Optical Society of America, 2012.

[86] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47(1-3):7–42, apr 2002.

[87] I. Ihrke, J. Restrepo, and L. Mignard-Debise. Principles of light field imaging: Briefly revisiting 25 years of research. *IEEE Signal Processing Magazine*, 33(5):59–69, Sept 2016.

[88] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, PP(99):1–1, 2017.

[89] J. Geng. Structured-light 3d surface imaging: a tutorial. *Adv. Opt. Photon.*, 3(2):128–160, Jun 2011.

[90] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recogn.*, 43(8):2666–2680, aug 2010.

[91] S. Foix, G. Alenya, and C. Torras. Lock-in time-of-flight (tof) cameras: A survey. *IEEE Sensors Journal*, 11(9):1917–1926, Sept 2011.

[92] A. Bhandari and R. Raskar. Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-d imaging. *IEEE Signal Processing Magazine*, 33(5):45–58, Sept 2016.

[93] S. Gupta, P. Arbelez, and J. Malik. Perceptual organization and recognition of indoor scenes from rgb-d images. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 564–571, June 2013.

[94] K. Lai, L. Bo, X. Ren, and D. Fox. *RGB-D Object Recognition: Features, Algorithms, and a Large Scale Benchmark*, pages 167–192. Springer London, London, 2013.

[95] P. Latorre-Carmona, E. Sanchez-Ortiga, X. Xiao, F. Pla, M. Martinez-Corral, H. Navarro, G. Saavedra, and B. Javidi. Multispectral integral imaging acquisition and processing using a monochrome camera and a liquid crystal tunable filter. *Opt. Express*, 20(23):25960–25969, Nov 2012.

[96] R. Horisaki, X. Xiao, J. Tanida, and B. Javidi. Feasibility study for compressive multi-dimensional integral imaging. *Opt. Express*, 21(4):4263–4279, Feb 2013.

[97] I. Muniraj, B. Kim, and B.-G. Lee. Encryption and volumetric 3d object reconstruction using multispectral computational integral imaging. *Appl. Opt.*, 53(27):G25–G32, Sep 2014.

[98] W. Feng, H. Rueda, C. Fu, C. Qian, and G. R. Arce. Compressive spectral integral imaging using a microlens array. In *Proc.SPIE*, volume 9857, pages 9857 – 9857 – 9, 2016.

[99] V. Farber, Y. Oiknine, I. August, and A. Stern. Compressive 4d spectro-volumetric imaging. *Opt. Lett.*, 41(22):5174–5177, Nov 2016.

[100] J. Wu, B. Xiong, X. Lin, J. He, J. Suo, and Q. Dai. Snapshot hyperspectral volumetric microscopy. *Scientific reports*, 6(24624), 2016.

[101] W. Feng, H. Rueda, C. Fu, G. R. Arce, W. He, and Q. Chen. 3d compressive spectral integral imaging. *Opt. Express*, 24(22):24859–24871, Oct 2016.

[102] L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu. Simultaneous depth and spectral imaging with a cross-modal stereo system. *IEEE Transactions on Circuits and Systems for Video Technology*, PP(99):1–1, 2016.

[103] M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady. 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Trans. Graph.*, 31(4):38:1–38:11, July 2012.

[104] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal. Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Opt. Express*, 19(22):21485–21507, Oct 2011.

[105] M. J. Sun, M. Edgar, G. M. Gibson, B. Sun, N. Radwell, R. Lamb, and M. J. Padgett. Single-pixel three-dimensional imaging with time-based depth resolution. *Nature Communications*, 7(12010), 2016.

[106] M. Van den Bergh and L. Van Gool. Combining rgb and tof cameras for real-time 3d hand gesture interaction. In *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 66–72, Jan 2011.

[107] Espros Photonics Corporation. Datasheet - epc660 3d tof imager 320 x 240 pixel, 2016. Accessed: Nov, 2016. Available: ftp://ftp.espros.com/01_Chips/Datasheet_epc660-V1.08.pdf.

[108] S. Hsu, S. Acharya, A. Rafii, and R. New. *Performance of a Time-of-Flight Range Camera for Intelligent Vehicle Safety Applications*, pages 205–219. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[109] C. Fu, M. L. Don, and G. R. Arce. Compressive spectral imaging via polar coded aperture. *IEEE Transactions on Computational Imaging*, 3(3):408–420, Sept 2017.

[110] K. Barner and G. R. Arce. *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*. CRC Press, Inc., Boca Raton, FL, USA, 2003.

[111] M. L. Don, C. Fu, and G. R. Arce. Compressive imaging via a rotating coded aperture. *Appl. Opt.*, 56(3):B142–B153, Jan 2017.

Appendix A

DISCLAIMER

Some target objects used in this dissertation were not endorsed by the trademark owners and they are used here as fair use to illustrate the quality of reconstruction of compressive spectral image measurements. LEGO is a trademark of the LEGO Group, which does not sponsor, authorize or endorse the images in this dissertation. The LEGO Group. All Rights Reserved. http://aboutus.lego.com/en-us/legal-notice/fair-play/.

Honors, Awards and Publications

Honors and Awards

1. **University Dissertation Fellow Award**. Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware, USA, (2016).

2. **Signal Processing Communications Graduate Faculty Award**. Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware, USA, (2016).

3. **Fulbright-Colciencias Scholarship**. Scholarship to pursue Ph.D. studies abroad. Francisco Jose de Caldas Institute for the Development of Science and Technology Colciencias, Bogota D.C., Colombia. (2012-2017)

Book Chapters

1. G. R. Arce, **H. Rueda**, C. V. Correa, A. Ramirez and H. Arguello, "Snapshot Compressive Multispectral Cameras", *Wiley Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons, Inc., (2017).

Journal Papers

1. **H. Rueda**, H. Arguello and G. R. Arce, "Color Coded Aperture Optimization in Compressive Spectral Imaging via Blue Noise Shaping", Submitted to IEEE Transactions on Image Processing, (November 2017).

2. **H. Rueda**, C. Fu, D. Lau and G. R. Arce, "Single Aperture Spectral+ToF Compressive Camera: Towards Hyperspectral+Depth Imagery", IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 7, pp. 992 - 1003, (2017).

3. **H. Rueda**, H. Arguello and G. R. Arce, "Compressive spectral testbed imaging system based on thin-film color-patterned filter arrays', Applied Optics, vol. 55, pp. 9584-9593, (2016).

4. W. Feng, **H. Rueda**, C. Fu, G. R. Arce, W. He and Q. Chen, "3D compressive spectral integral imaging', Optics Express, vol. 24, pp. 24859-24871, (2016)

5. J. Tan, Y. Ma, **H. Rueda**, D. Baron and G. R. Arce, "Compressive Hyperspectral Imaging via Approximate Message Passing", IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 2, pp. 389-401, (2016).

6. **H. Rueda**, D. Lau and G. R. Arce, "Multi-spectral compressive snapshot imaging using RGB image sensors", Optics Express, vol. 23, no. 9, pp. 12207-12221, (2015).

7. **H. Rueda**, H. Arguello and G. R. Arce, "DMD-based implementation of patterned optical filter arrays for compressive spectral imaging", Journal of the Optical Society of America A, vol. 32, no. 1, pp. 80-89, (2015).

8. H. Arguello, **H. Rueda**, Y. Wu, I. Mirza, D. Prather and G. R. Arce, "Higher-order computational model for coded aperture spectral imaging", Applied Optics, vol. 52, no. 10, pp. D12-D21, (2013).

Conference Papers

1. **H. Rueda**, H. Arguello, and G. R. Arce, "High-dimensional Optimization of Color Coded Apertures for Compressive Spectral Cameras", Proceedings of European Signal Processing Conference (EUSIPCO '17), Kos Island, Greece, (August 2017).

2. **H. Rueda**, H. Arguello, and G. R. Arce, "Optimal Colored Coded Apertures for Compressive Spectral Imaging Systems", Proceedings of Imaging and Applied Optics 2017 (COSI '17), OSA Technical Digest (online) (Optical Society of America, 2017), San Francisco, CA., USA, (June 2017).

3. **H. Rueda**, D. Lau, and G. R. Arce, "Spectral+Depth Imaging with a Time-of-Flight Compressive Snapshot Camera ", Proceedings of Imaging and Applied Optics 2017 (COSI '17), OSA Technical Digest (online) (Optical Society of America, 2017), San Francisco, CA., USA, (June 2017).

4. D. Lau, Y. Zhang, T. Hastings, **H. Rueda**, and G. R. Arce, "Light Field Modeling for Coded Aperture Systems", Proceedings of Imaging and Applied Optics 2017 (COSI '17), OSA Technical Digest (online) (Optical Society of America, 2017), San Francisco, CA., USA, (June 2017).

5. **H. Rueda**, H. Arguello, and G. R. Arce, "Development of a Compressive Spectral Testbed based on Thin-film Color patterned Filter Array", Proceedings of Imaging and Applied Optics 2016 (COSI '16), OSA Technical Digest (online) (Optical Society of America, 2016), paper CW5D.3, Heidelberg, Germany, (July 2016).

6. W. Feng, **H. Rueda**, C. Fu, Q. Chen and G. R. Arce, "Compressive spectral integral imaging using a microlens array", Proceedings of SPIE Conference on Commercial + Scientific Sensing and Imaging (DCS '16), Baltimore, MD, USA, (April 2016).

7. **H. Rueda**, H. Arguello, and G. R. Arce, "Colored Coded Aperture Compressive Spectral Imaging: Design and Experimentation", Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP '15), Orlando, FL., USA, (December 2015).

8. **H. Rueda**, D. Lau, and G. R. Arce, "RGB Detectors on Compressive Snapshot Multi-spectral Imagers", Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP '15), Orlando, FL., USA, (December 2015).

9. J. Tan, Y. Ma, **H. Rueda**, D. Baron and G. R. Arce, "Application of Approximate Message Passing in Coded Aperture Snapshot Spectral Imaging", Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP '15), Orlando, FL., USA, (December 2015).

10. **H. Rueda**, H. Arguello, G. R. Arce, "Dual-Arm VIS/NIR Compressive Spectral Imager", Proceedings of IEEE International Conference on Image Processing (ICIP '15), Quebec City, Canada, (September 2015).

11. **H. Rueda**, H. Arguello, G. R. Arce, "Experimental demonstration of a colored coded aperture-based compressive spectral imaging system", Proceedings of Imaging and Applied Optics (COSI '14), OSA Technical Digest (online) paper CTu2C.6, Kohala Coast, HI, USA, (June 2014).

12. **H. Rueda**, H. Arguello, G. R. Arce, "Compressive spectral imaging based on colored coded apertures", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '14), Florence, Italy, (May 2014).

13. **H. Rueda**, A. Parada, H. Arguello, Y. Wu, D. W. Prather, and G. R. Arce, "Demonstration of a higher-order discretization model for compressive spectral imaging", Proceedings of Imaging and Applied Optics (COSI '13), OSA Technical Digest (online) paper CM4C.4, Arlington, VA, USA, (June 2013).

14. **H. Rueda**, H. Arguello and G. R. Arce. "On super-resolved coded aperture spectral imaging", Proceedings of SPIE Conference on Defense, Security and Sensing (DSS '13), Baltimore, MD, USA, (May 2013).

15. H. Arguello, **H. Rueda** and G. R. Arce. "High-precision discretization model for coded aperture-based compressive spectral imaging", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '13), Vancouver, CA, (May 2013).

16. H. Arguello, **H. Rueda** and G. R. Arce. "Spatial super-resolution in code aperture spectral imaging", Proceedings of SPIE Conference on Defense, Security and Sensing (DSS '12), Baltimore, MD, USA, (April 2012).

Biography

Hoover F. Rueda-Chacon received the B.Sc. and M.Sc. degrees in Computer Science from Universidad Industrial de Santander, Colombia, in 2009 and 2012, and the M.Sc. degree in Electrical and Computer Engineering from University of Delaware, DE, USA, in 2015. He is currently pursuing the Ph.D. in Electrical and Computer Engineering at the University of Delaware. His main research areas are computational optical imaging, compressed sensing and high dimensional signal processing.

To date, Hoover has co-authored over 40 combined publications, and a book chapter. He interned with Disney Research Zurich, Switzerland, during the summer 2016, where he developed mathematical models and computational algorithms for multispectral projector/camera systems. He received a Colciencias/Fulbright scholarship to pursue his Ph.D. studies abroad in 2012, the Signal Processing Communications Graduate Faculty Award and the University Dissertation Fellow Award from the University of Delaware both in 2016.