SISTEMA ÓPTICO-COMPUTACIONAL PARA LA ESTIMACIÓN DE LA INFORMACIÓN ESPECTRAL Y DE PROFUNDIDAD DE UNA ESCENA MEDIANTE EL USO DE LA TEORÍA DE MUESTREO COMPRESIVO

MIGUEL ANGEL MARQUEZ CASTELLANOS

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTY OF SCIENCES
PHYSICS SCHOOL
BUCARAMANGA
2022

SISTEMA ÓPTICO-COMPUTACIONAL PARA LA ESTIMACIÓN DE LA INFORMACIÓN ESPECTRAL Y DE PROFUNDIDAD DE UNA ESCENA MEDIANTE EL USO DE LA TEORÍA DE MUESTREO COMPRESIVO

MIGUEL ANGEL MARQUEZ CASTELLANOS

DOCTORAL THESIS TO QUALIFY FOR THE TITLE OF:
DOCTOR OF PHILOSOPHY IN PHYSICS

ADVISOR:
HENRY ARGUELLO
PH.D IN ELECTRICAL ENGINEERING
CO-ADVISOR:
HOOVER RUEDA CHACON
PH.D IN ELECTRICAL ENGINEERING

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTY OF SCIENCES
PHYSICS SCHOOL
BUCARAMANGA
2022

DEDICATORIA

I dedicate my dissertation to my family, friends, and parents. Thank you for your unconditional support.

CONTENIDO

pág.

1. Introduction **15**

1.1. Hypothesis 19

1.2. Scope of the thesis 20

1.2.1. Organization of the thesis 21

2. Fundamentals of Spectral and Depth Imaging **24**

2.1. Spectral Imaging 24

2.1.1. Line scan/Pushbroom 25

2.1.2. Filter-based approach 29

2.2. Depth imaging 32

2.2.1. Depth from defocus (DFD) 32

2.2.2. Structured light (SL) 38

2.2.3. Light field (LF) 46

2.3. Compressive imaging 53

2.3.1. Compressive spectral and depth imaging 54

3. CASSI-VSI: Coded Aperture Snapshot Spectral Imaging via Virtual Side Information **62**

3.1. Continuous and discrete sensing model 65

3.1.1. Continuous Sensing Model 66

3.1.2. Discrete Sensing Model 70

3.2. Reconstruction algorithm 73

3.3. Simulation results 76

3.3.1. Beam-splitting-ratio Theoretical Analysis 78

3.3.2. Beam-splitting-ratio Reconstruction Performance 81
3.3.3. General Performance Metric 84
3.3.4. Comparison with the State-of-the-art Imaging Systems 85
3.3.5. Comparison with the State-of-the-art Algorithms 86
3.4. Proof of concept experiments 89
3.5. Conclusions 97

4. CASSDI-DFA: Compressive Spectral Depth Imager via Depth from Aberrations 98
4.1. Continuous and discrete sensing model 101
4.1.1. Continuous sensing model 101
4.1.2. Discrete sensing model 104
4.2. Reconstruction algorithm 107
4.2.1. Reconstructing $\mathbf{f}_g$ 107
4.2.2. Reconstructing $\mathbf{f}_f$ 109
4.2.3. Reconstructing $\mathbf{f}_s$ 112
4.3. Simulation results 113
4.3.1. Reconstruction performance for given aberrations 115
4.3.2. Depth and spectral performance 120
4.4. Proof of concept experiments 121
4.5. Conclusions 126

5. C^{3}SLFI: Color-coded Compressive Spectral Light Field Imager 128
5.1. Continuous and tensor model 130
5.1.1. Continuous model 132
5.1.2. Tensor model 135
5.2. Reconstruction algorithms 136
5.2.1. Conventional reconstruction problem 136

5.2.2. 5D to 3D spectral light field parameterization 137
5.2.3. Online image representation 139
5.2.4. Proposed reconstruction algorithm 141
5.3. Simulation results 147
5.3.1. WDCT and DL versus OTD representation basis 148
5.3.2. Performance evaluation for different compression ratios and additive noise scenarios 150
5.3.3. Spatial, angular, and spectral performance 151
5.4. Proof of concept experiments 153
5.5. Conclusions 159

6. CS^{2}LI: Compressive Spectral Structured Light Imager 161
6.1. Continuous and discrete sensing model 163
6.1.1. Continuous sensing model 165
6.1.2. Discrete sensing model 167
6.2. Reconstruction algorithm 168
6.2.1. Traditional ADMM optimization problem 169
6.2.2. Hierarchical-based PnP-ADMM reconstruction algorithm 170
6.2.3. Depth map recovery via N-step least-square PSA principle 174
6.3. Simulation results 175
6.3.1. Performance Evaluation for Different Additive Noise Scenarios 176
6.3.2. Color coded aperture design 178
6.4. Proof of concept experiments 182
6.5. Conclusions 189

7. Mu-CoSDI: Multimodal Compressive Spectral Depth Imager 191
7.1. Continuous and discrete sensing model 193
7.1.1. Continuous sensing model 195

7.1.2. Discrete sensing model 199
7.2. Reconstruction algorithm 200
7.3. Simulation results 203
7.4. Proof of concept experiments 205
7.5. Conclusions 218

8. Mu-CoSDI: Multimodal Compressive Spectral Depth Imager 220
8.1. Continuous and discrete sensing model 222
8.1.1. Continuous sensing model 224
8.1.2. Discrete sensing model 227
8.2. Reconstruction algorithm 229
8.3. Simulation results 232
8.4. Proof of concept experiments 234
8.5. Conclusions 246

9. Conclusions and Future Work 248

BIBLIOGRAPHY 251

LISTA DE FIGURAS

pág.

Figura 1.	Doctoral thesis sketch	22
Figura 2.	Pushbroom implementation	26
Figura 3.	Pushbroom's spectrogram characterization	27
Figura 4.	Pushbroom's raw measurement	28
Figura 5.	Wavelenght-scan implementation	30
Figura 6.	Blur effect in a simple lens system	33
Figura 7.	Deformable mirror-based depth-from-defocus system.	35
Figura 8.	Structured light optical system	45
Figura 9.	Epipolar plane	48
Figura 10.	Light field imagers	50
Figura 11.	Light field implementation	51
Figura 12.	Lenslet array modalities	52
Figura 13.	Compressive sensing model	53
Figura 14.	Compressive spectral imagers	56
Figura 15.	Compressive depth imagers	60
Figura 16.	CASSI-VSI sketch	64
Figura 17.	CASSI-VSI sketch	67
Figura 18.	Sensing matrices example	74
Figura 19.	Splitting ratio ablation	81
Figura 20.	SNR analysis in CASSI-VSI	82
Figura 21.	CASSI-VSI reconstruction results	83
Figura 22.	SNR ablation analysis	84

Figura 23. CASSI, CASSI-VSI and dual-arm CASSI simulation results 87
Figura 24. CASSI, CASSI-VSI and dual-arm CASSI spectral results 88
Figura 25. Lambda-net-based CASSI-VSI reconstruction results 89
Figura 26. Lambda-net testing results with SNR=20dB 90
Figura 27. Lambda-net testing results with SNR=30dB 90
Figura 28. (A) Testbed implementation of the CASSI-VSI. (B) Geometric sketch of the optical system. 91
Figura 29. CASSI-VSI raw compressed measurements: Boat and Church scenes 93
Figura 30. CASSI-VSI experimental results: Boat scene 94
Figura 31. CASSI-VSI experimental results: Church scene 95
Figura 32. CASSI-VSI implementation 96

Figura 33. CASSDI-DFA sketch 102
Figura 34. CASSDI-DFD reconstruction pipeline. 111
Figura 35. Zernike polynomials cross-validation 116
Figura 36. CASSDI-DFD depth results: Cars scene 118
Figura 37. CASSDI-DFD spectral results: Cars scene 119
Figura 38. CASSDI-DFD implementation 122
Figura 39. CASSDI-DFD depth experimental results: Flowers scene 124
Figura 40. CASSDI-DFD spectral experimental results: Flowers scene 125
Figura 41. CASSDI-DFA spider graph 127

Figura 42. C^3SLFI sketch 131
Figura 43. C^3SLFI: Light field image parametrization 139
Figura 44. (A) C^3SLFI's emulation via spectral-angular raster scanner 146
Figura 45. C^3SLFI: Sparse representation analysis 149
Figura 46. C^3SLFI's spatial reconstruction performance: Cars scene 152
Figura 47. C^3SLFI's depth reconstruction performance: Cars scene 153

Figura 48. C^{3}SLFI's spectral reconstruction performance: Cars scene 154
Figura 49. C^{3}SLFI setup 155
Figura 50. Examples of the complementary CCAs 156
Figura 51. C^{3}SLFI raw measurement 156
Figura 52. C^{3}SLFI spectral result 157
Figura 53. C^{3}SLFI spectral signature analysis 158
Figura 54. C^{3}SLFI spider graph 160

Figura 55. CS^2L sketch 164
Figura 56. CS^2L's spectral reconstruction performance: Boxplot 177
Figura 57. Structured light depth results: RGB and depth map 178
Figura 58. Coded aperture ablation: Spectral results 180
Figura 59. CS^{2}LI setup 181
Figura 60. Images of experimental fringes 183
Figura 61. Wrapped and unwrapped phase 183
Figura 62. CS^{2}LI's spectral and depth reconstruction performance: Kids scene 185
Figura 63. CS^{2}LI's spectral and depth reconstruction performance: Toys scene 186
Figura 64. CS^{2}LI's spectral and depth reconstruction performance: Tayrona scene 187
Figura 65. CS^{2}LI spectral signature reconstruction performance 188
Figura 66. CS^{2}SLFI chart graph 190

Figura 67. Mu-CoSDI sketch 196
Figura 68. CNN-based CS^{2}LI 200
Figura 69. Mu-CoSDI system 207
Figura 70. Coded aperture characterization 208
Figura 71. LF-based Mu-CoSDI results: Penguin scene 209
Figura 72. LF-based Mu-CoSDI results: Church scene 211
Figura 73. LF-based Mu-CoSDI results: Astronaut scene 212

Figura 74. SL-based Mu-CoSDI results: Leaf scene 214
Figura 75. pure-CNN-based Mu-CoSDI results: Students scene 215
Figura 76. pure-CNN-based Mu-CoSDI results: Creator scene 216
Figura 77. pure-CNN-based Mu-CoSDI results: Palms scene 217
Figura 78. Mu-CoSDI spider graph 218

Figura 79. Mu-CoSDI sketch 225
Figura 80. CNN-based CS^{2}LI 229
Figura 81. Mu-CoSDI system 235
Figura 82. Coded aperture characterization 237
Figura 83. LF-based Mu-CoSDI results: Penguin scene 238
Figura 84. LF-based Mu-CoSDI results: Church scene 239
Figura 85. LF-based Mu-CoSDI results: Astronaut scene 240
Figura 86. SL-based Mu-CoSDI results: Leaf scene 242
Figura 87. pure-CNN-based Mu-CoSDI results: Students scene 243
Figura 88. pure-CNN-based Mu-CoSDI results: Creator scene 244
Figura 89. pure-CNN-based Mu-CoSDI results: Palms scene 245
Figura 90. Mu-CoSDI spider graph 246

Figura 91. Spider graph summarize 250
Figura 92. Double Amici prism ray diagram 273
Figura 93. Left-prism ray diagram 274
Figura 94. Middle-prism ray diagram 275
Figura 95. Right-prism ray diagram 276
Figura 96. Refractive indices analysis 278
Figura 97. ϕ_1 and ϕ_3 ablation analysis 279
Figura 98. Prism deviation analysis 281
Figura 99. Customized double Amici prism 282

Figura 100. Signal-to-noise ratio analysis between the split signal intensities and the total sensor noise. 285
Figura 101. Absolute error analysis between the intensity ratios $\frac{(\mathbf{g}_T)_i}{(\mathbf{g}_R)_i}$ and $\frac{(\mathbf{g}_R)_i}{(\mathbf{g}_T)_i}$. 286
Figura 102. A-VSI reconstruction results 288
Figura 103. Error estimation 296

RESUMEN

TÍTULO: Sistema óptico-computacional para la estimación de la información espectral y de profundidad de una escena mediante el uso de la teoría de muestreo compresivo [*]

AUTOR: Miguel Angel Marquez[1]

PALABRAS CLAVE: Muestreo compresivo, codigos de apertura, imágenes espectrales, imágenes de profundidad, imágenes computacionales de alta dimensionalidad.

DESCRIPCIÓN:

Esta tesis doctoral propone diseñar, simular, desarrollar e implementar un sistema CoSDI eficiente y robusto capaz de registrar la información de profundidad espectral de las escenas, haciendo uso de un único detector. Específicamente, el paradigma de detección de imágenes espectrales instantáneas de apertura codificada (CASSI, por sus siglas en inglés) se estudia como base para transferir muchas propiedades ventajosas integradas en sus modelos matemáticos a sistemas de imágenes de profundidad. Este primer estudio condujo al desarrollo de un sistema CASSI basado en información lateral que depende de un solo sensor de imagen, denominado información lateral virtual CASSI (CASSI-VSI). Este nuevo generador de imágenes está inspirado en un interferómetro de doble trayectoria, donde el brazo del haz de muestra corresponde al frente de onda codificado-y-disperso (CaD, de sus siglas en inglés) [denominado medición CASSI] y el trayecto de referencia al frente de onda codificado espacialmente. El estudio y contribución en el marco de imágenes espectrales de compresión, permitio la extrapolación de la geometría de detecciín de CaD a tres técnicas tradicionales de imágenes de profundidad: profundidad desde el desenfoque, campo de luz y luz estructurada. En consecuencia, desarrollaremos cuatro sistemas CoSDI basados en CaD de sensor único que permiten realizar tareas de codificación y decodificación en cubos de datos de cuatro dimensiones. Además, como componente indispensable en CoSDI, se proponen cuatro métodos avanzados de reconstrucción de imágenes inspirados en los paradigmas matemáticos: método de dirección alterna de multiplicadores (ADMM), ADMM basado en tensores, ADMM plug-and-play, y red neuronal convolucional basada en ADMM.

---

[*] Tesis de doctorado

[1] Facultad de Ciencias. Escuela de Física., Director Henry Arguello. Co-director Hoover Rueda

ABSTRACT

TITLE: Compressive spectral-depth imaging [1]

AUTHOR: MIGUEL ANGEL MARQUEZ CASTELLANOS [**]

KEYWORDS: Compressive sensing, coded apertures, spectral imaging, depth imaging, computational high dimensional imaging.

DESCRIPTION:

This doctoral thesis proposes designing, simulating, developing, and implementing an efficient and robust CoSDI system capable of recording the scenes' spectral-depth information of a scene of interest, using just a single sensor. Specifically, the coded aperture snapshot spectral imaging (CASSI) sensing paradigm is studied as a baseline to transfer many advantageous properties embedded in its mathematical models into depth imaging systems. This first study led to the development of a side-information-based CASSI system that relies on a single image sensor, termed CASSI virtual side information (CASSI-VSI). This new imager is inspired by the dual-dispersive compressive spectral imager, where the sample beam path corresponds to the coded and dispersed (CaD) wavefront (referred to as CASSI measurement) and the reference path to the spatially coded wavefront (referred to as encoded grayscale measurement). Having studied and contributed to the compressive spectral imaging framework, we proceeded to extrapolate the CaD sensing geometry to three traditional depth imaging techniques: depth-from-defocus, light field, and structured light.

Consequently, we developed four single-sensor CaD-based CoSDI systems that offer encoding - decoding tasks for four-dimensional datacubes. Alongside these systems, we proposed four advanced computational image reconstruction methods inspired by the mathematical algorithms: alternating direction method of multipliers (ADMM), tensor-based ADMM, plug-and-play ADMM, and convolutional neural network based on the ADMM. The resulting optical systems – one for spectral and four for spectral-depth – are experimentally validated in the HDSP optical laboratory.

---

[1] Doctoral thesis

[**] Faculty of Sciences. School of Physics. Advisor Henry Arguello. Co-advisor Hoover Rueda

1. Introduction

In the last two decades, the light flux interpretation as a high-dimensional space has become increasingly popular to describe the scene's associated brightness information Pierre Pellat-Finet. *Optique de Fourier: Théorie métaxiale et fractionnaire*. Springer Science & Business Media, 2009. In particular, the brightness of all light rays observed in a static scene is represented by a six-dimensional function, which is a special case of the seven dimensional plenoptic function $f(x, y, z, \gamma, \theta, \lambda, t)$ that represents the space (x, y), depth (z), angle (γ, θ), wavelength (λ) and time (t) Marc Levoy. "Light fields and computational imaging". En: *Computer* 39.8 (2006), págs. 46-55; Edward H Adelson, James R Bergen y col. *The plenoptic function and the elements of early vision*. Vision y Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991. Efforts to measure more than three dimensions of the plenoptic function has resulted in the emergence of the spectral depth imaging (SDI) research area, aiming to capture or estimate the four dimensions (x, y, z, λ) related with the scene.

In this direction, and given the increasing availability of spectral and depth imagers, multiple works based on scanning sampling protocols have been resported in the SDI framework Abderrahim Halimi y col. "Joint reconstruction of multitemporal or multispectral single-photon 3D LiDAR images". En: *2019 Sensor Signal Processing for Defence Conference (SSPD)*. IEEE. 2019, págs. 1-5; Julián Tachella y col. "Bayesian 3D Reconstruction of Subsampled Multispectral Single-photon Lidar Signals". En: *arXiv preprint arXiv:1904.02583* (2019). Conventional scan-based SDI systems are composed of two disjoint optical imagers: one that acquires the spectral information and another onte that measures the depth. In particular, the scanning-based spectral imagers are classified into three categories: Whiskbroom, Pushbroom Rajiv Gupta y Richard I Hartley. "Linear pushbroom cameras". En: *IEEE Transactions on Pattern*

Analysis and Machine Intelligence 19.9 (1997), págs. 963-975; Luis Gómez-Chova y col. “Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images”. En: *Applied Optics* 47.28 (2008), F46-F60, and Filter-based scan Andres Castellanos-Gomez y col. “Spatially resolved optical absorption spectroscopy of single-and few-layer MoS2 by hyperspectral imaging”. En: *Nanotechnology* 27.11 (2016), págs. 115705-115710; Mohammad Torabzadeh y col. “Hyperspectral imaging in the spatial frequency domain with a supercontinuum source”. En: *Journal of Biomedical Optics* 24.7 (2019), págs. 071614-071623; Peter F Favreau y col. “Excitation-scanning hyperspectral imaging microscope”. En: *Journal of Biomedical Optics* 19.4 (2014), págs. 046010-046021; Gyanendra Sheoran y col. “Swept-source digital holography to reconstruct tomographic images”. En: *Optics Letters* 34.12 (2009), págs. 1879-1881. In all them, the diffraction grating and electronic positioning devices (e.g., linear translation stage or filter wheels) are the essential elements to decompose and acquire the light’s constituent wavelengths. On the other hand the scan-based depth imagers are classified into three representative sensing paradigms: depth-dependent geometrical aberration (referred to as depth-from-defocus (DFD)) Huixuan Tang y col. “Depth from Defocus in the Wild”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 2740-2748; Murali Subbarao y Gopal Surya. “Depth from defocus: A spatial domain approach”. En: *International Journal of Computer Vision* 13.3 (1994), págs. 271-294; Julie Chang y Gordon Wetzstein. “Deep Optics for Monocular Depth Estimation and 3D Object Detection”. En: *arXiv preprint arXiv:1904.08601* (2019); RenÃ© Ranftl y col. “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.3 (2022), págs. 1623-1637. DOI: 10.1109/TPAMI.2020.3019967, phase deformation (referred to as structured light (SL)) Stefan Heist y col. “5D hyperspectral imaging: fast and accurate measurement of surface

shape and spectral characteristics using structured light". En: *Optics Express* 26.18 (2018), págs. 23366-23379; Joan Batlle, E Mouaddib y Joaquim Salvi. "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey". En: *Pattern Recognition* 31.7 (1998), págs. 963-982, and multi-parallax views (referred to as light field (LF)) Hae-Gon Jeon y col. "Accurate depth map estimation from a lenslet light field camera". En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555; Jeon y col., "Accurate depth map estimation from a lenslet light field camera"; Yang Zhao y col. "Heterogeneous camera array for multispectral light field imaging". En: *Optics express* 25.13 (2017), págs. 14008-14022. Specifically, the DFD is a passive-illumination approach based on expensive high-precision linear positioning devices or tunable focal lenses. The SL is an active-illumination approach that relies on illuminating the scene with 2D spatially varying fringe patterns. Finally, the LF is a passive-illumination approach divided into three fundamental categories: camera-array systems, linear positioning devices, and lenslet arrays. Although, the scan-based SDI systems allow high spectral and depth resolution, they are unsuitable for rapidly changing scenarios because of their time-consuming sensing protocols. Moreover, SDI systems coupling independent spectral and depth optical systems imply registration and high-complex calibration processes.

To circumvent the scanning-based time requirements, researchers have proposed to incorporate the ideas from compressive sensing theory into spectral and depth imaging systems, giving rise to the framework of compressive spectral depth imaging (CoSDI). In CoSDI, coding and multiplexing schemes are exploited to compress the high dimensional spectral and depth data into a single two-dimensional intensity measurement. State-of-the-art CoSDI systems rely on the dual-arm paradigm– two sensors coupled by a beam splitter– to alleviate the complexity of the four-dimensional datacube retrieval. Specifically, one of the arms corresponds to a com-

pressive spectral imager (CSI) ME Gehm y col. “Single-shot compressive spectral imaging with a dual-disperser architecture”. En: *Optics express* 15.21 (2007), págs. 14013-14027; Ashwin Wagadarikar y col. “Single disperser design for coded aperture snapshot spectral imaging”. En: *Applied optics* 47.10 (2008), B44-B51; Seung-Hwan Baek y col. “Single-shot Hyperspectral-Depth Imaging with Learned Diffractive Optics”. En: (2021), págs. 2651-2660; Daniel S Jeon y col. “Compact snapshot hyperspectral imaging with diffracted rotation”. En: *ACM Transactions on Graphics (TOG)* 38.4 (2019), págs. 1-13; Yingming Lai y col. “Compressed ultrafast tomographic imaging by passive spatiotemporal projections”. En: *Optics Letters* 46.7 (2021), págs. 1788-1791, while the other one is a compressive depth imager (CDI) Edward H Adelson y John Y. A. Wang. “Single lens stereo with a plenoptic camera”. En: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 14.2 (1992), págs. 99-106; Eric Krotkov. “Focusing”. En: *International Journal of Computer Vision* 1.3 (1988), págs. 223-237; Todor Georgiev y col. “Lytro camera technology: theory, algorithms, performance analysis”. En: 8667 (2013), 86671J-86681J. Both arms work by first encoding a high-dimensional scene with one or more coding elements (such as coded apertures, spatial light modulators (SLM) or deformable mirrors), followed by an auxiliary arbitrary disambiguating operation (e.g., spectral shearing or wavefront relaying process), and finally, integrating the encoded scene on a focal plane array. Despite the increasing system complexity compared to conventional imaging modalities, CoSDI allows using widely available two-dimensional (2D) focal plane arrays (e.g., CCD and CMOS cameras) to collect multiplexed snapshots and then recover the underlying four-dimensional datacube using computational methods. Leveraging this encoder-decoder approach, CoSDI sparks the development of many high-dimensional imaging systems with low bandwidth/memory requirements, low costs, and low power consumption.

Although CSI and CDI optical schematics share similar concepts, their propaga-

tion models and coder-decoder processes significantly differ, implying the use of post-processing registration, alignment, and fusion algorithms. These shortcomings are overshadowed by the computational and assembly complexity that implies developing a single sensor-based CoSDI system. Because of this, and motivated by the deep-learning-based CSI and CDI retrieval algorithms, CoSDI approaches have settled into the dual-arm paradigm. In this path, few works have tackled the single-sensor CoSDI approach by relying purely on convolutional neural network (CNN) capabilities to design compact systems at the expense of reconstruction accuracy Ranftl y col., "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer"; *End-to-end hyperspectral-depth imaging with learned diffractive optics*. IEEE/CVF International Conference on Computer Vision (ICCV), 2020. DOI: https://doi.org/10.1109/iccv48922.2021.00265.

Therefore, inspired by the close link between compressive depth imaging protocols and compressive spectral imaging protocols, this thesis aims to answer the following research question: ***Is it possible to develop a compressive sensing theory-based optical-computational system capable of simultaneously acquiring and reconstructing spectral depth images relying on a single 2D sensor? If affirmative, which optical elements, propagation model, and computational algorithm are required to build such a system?***

1.1. Hypothesis

Considering that:

- Previous studies have been carried out on the design of optical elements that allow exploiting the spectral and depth dependence of the point-spread-function as a coding element in the multiplexing of spatial-spectral or spatial-depth information.

- The CSI framework has addressed the limitation of the acquisition and corresponding reconstruction of three or more dimensions of a real scene from two-dimensional projections.
- Depth estimation and spectral compressive sensing systems, spatial light modulating elements, or diffractive optical elements are used in the coding and multiplexing process.
- In structured light systems, spatially variable wavelength illumination patterns have been studied to establish a one-to-one correspondence between the projection angle of a light plane and a specific wavelength.

The hypothesis of this thesis claims that:
It is feasible to design a compressive optical-computational imaging system to acquire and reconstruct spatial, spectral, and depth information of a scene using light modulators as coding elements and a single 2D image sensor.

1.2. Scope of the thesis

This Ph.D. dissertation aims to broaden the knowledge of compressive spectral-depth acquisition protocols, exploiting the intrinsic relationship between the optical elements used in the traditional compressive spectral and depth imagers and their sampling protocols. In the development of this work, the following results are expected:

- To develop an optical and mathematical model for acquiring spatial- spectral and depth information of a scene from light coding and multiplexing processes.
- To simulate a compression protocol for the simultaneous acquisition of spectral and depth information of a scene.

- To develop an optimization problem and algorithm for reconstructing spectral-depth images from compressed measurements.
- To build/ensemble an optical-computational system prototype to acquire spectral and depth information of a scene using the compressive sensing theory.

1.2.1. Organization of the thesis This doctoral thesis proposes to design, simulate, develop, and implement an efficient and robust CoSDI system capable of recording the scenes' spectral-depth information from a single sensor. Specifically, the doctoral thesis is divided into four main stages (see Fig. 1): theoretical background, compressive spectral imaging, compressive spectral-depth imaging, and multimodal compressive spectral depth imaging.

- **The first stage** briefly introduces the fundamentals of spectral imaging, depth imaging, compressive sensing theory, compressive spectral imaging, and compressive depth imaging. In this stage, the theoretical background is validated by implementing five optical systems –two scan-based spectral imagers and three scan-based depth imagers– and their corresponding decoding algorithms. With the development of these five scan-based spectral and depth imagers, this thesis fulfills **specific objectives 1, 2** and **4**.
- **In the second stage,** the coded aperture snapshot spectral imaging (CASSI) sensing geometry is studied as a baseline to transfer the advantageous properties embedded in its mathematical models into the depth optical systems. This first study results in the side-information-based CASSI system that relies on a single image sensor termed CASSI virtual side information (CASSI-VSI). This new imager is inspired by a double-dispersive spectral imager, where the sample beam path corresponds to the coded-and-dispersed (CaD) wavefront (referred to as CASSI measurement) and the reference path to the spatially

coded wavefront (referred to as encoded grayscale measurement). CASSI-VSI employs a self-customized double Amici prism, as an essential diffraction element for the CoSDI systems design. With the development of the CASSI-VSI, this stage reinforces the fulfillment of the **specific objective 3**.

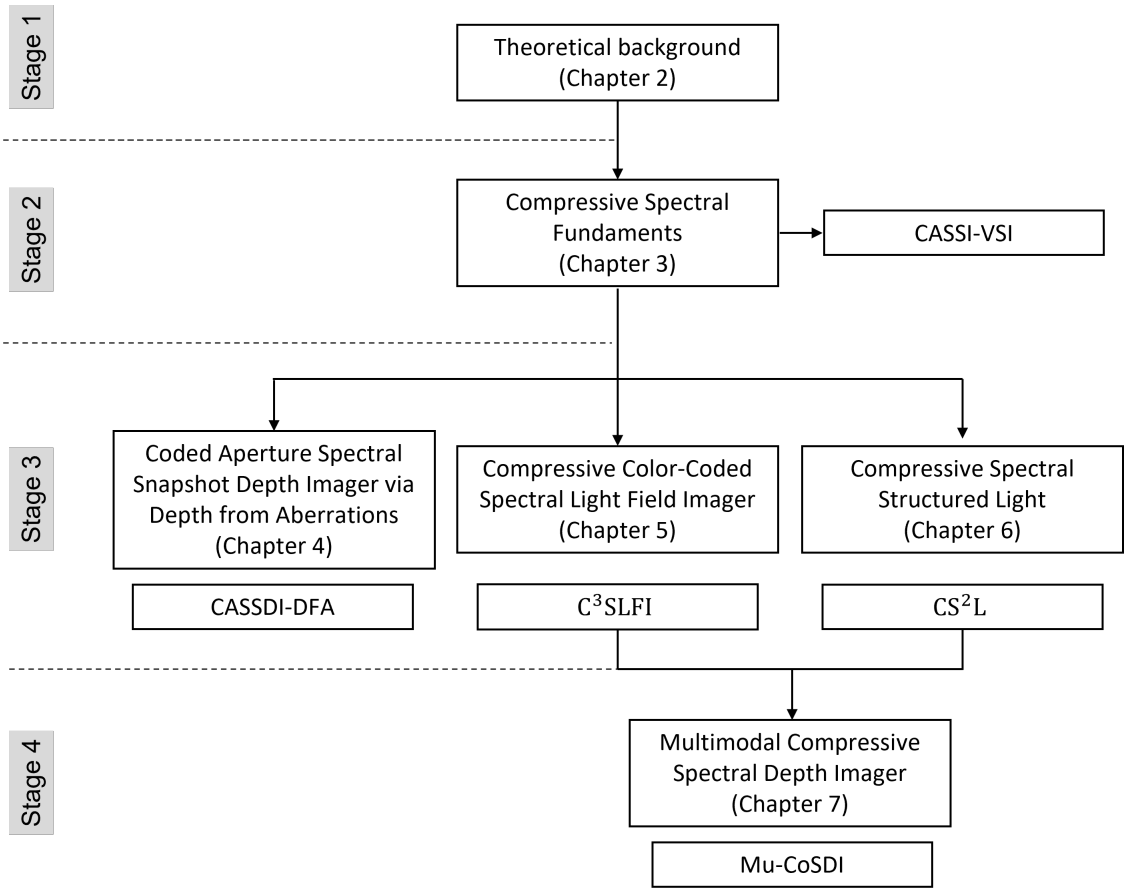


Figura 1. Doctoral thesis sketch.

- **In the third stage,** and inspired by the CASSI sensing geometry, we design three single-sensor CaD-based CoSDI systems, named: compressive spectral depth from aberrations imager (referred to as CASSDI-DFA), compressive color-coded spectral light field imager (referred to as C^3SFLI), and compressive spectral structured light imager (referred to as CS^2LI). Specifically, CASSDI-DFA relies on a 4f system consisting of two lenses with a deformable mirror

placed at its Fourier plane –for phase modulation via Zernike-modeled optical aberrations– followed by a conventional CASSI system. C^{3}SFLI builds on a lenslet array coupled to a color-coded-based CASSI system to capture multiplexed representations of the four-dimensional datacube. The CS^{2}LI emulate a double-dispersive CASSI system using a single double Amici prism along with a coded mirror. With the development and implementation of these three CoSDI systems, this stage contributes to fulfilling the **specific objectives 5, 6,** and **7**.

- **In the fourth stage,** and based on the CS^{2}LI sensing geometry, we introduce a Multimodal CoSDI (referred to as Mu-CoSDI) system capable of recovering SDI information relying on either one of the three depth sensing paradigms: SL, LF, or DFD. Specifically, the Mu-CoSDI system is constructed following a modular optical distribution (e.g., adding a lenslet array or light projectors) with the ability to reconstruct SDI in outdoor and indoor environments. Introducing the multimodal CoSDI system, this thesis fulfills **specific objectives 5, 6,** and **7**.

Moreover, as an indispensable constituent in CoSDI, four advanced computational image reconstruction methods, inspired by the alternating direction method of multipliers (ADMM) paradigm, are proposed. These algorithms are so-called: Woodbury matrix identity theorem-based ADMM (used by the CASSDI-DFA system in Chapter 4), tensor-based ADMM (used by the C^{3}SFLI system in Chapter 5), plug-and-play ADMM (used by the CS^{2}LI system in Chapter 6), and ADMM-based CNN (used by the Mu-CoSDI system in Chapter 8). With the development of these four algorithms, this thesis reinforces the fulfillment of the **specific objective 5**.

2. Fundamentals of Spectral and Depth Imaging

This chapter introduces the traditional depth and spectral imagers used as the building blocks in this dissertation. Specifically, we address the study of scanning-based acquisition systems. The forward model of these systems will be used to develop a hybrid proof-of-concept spectral-depth imager to create a spectral depth image dataset. This dataset is intended to be used to validate the proposed mathematical model, algorithms, and optical systems by simulations.

In Subsection 2.1 and 2.2, we will provide a detailed theoretical description of two traditional scanning-based spectral imagers and three depth imagers, respectively. Moreover, to better comprehend their operation principle, we will implement a tesbed of each optical system in the HDSP laboratory. Finally, Subsection 2.3 will introduce a brief theoretical description of compressive sensing (CS) theory as well as CS-based spectral and depth imagers.

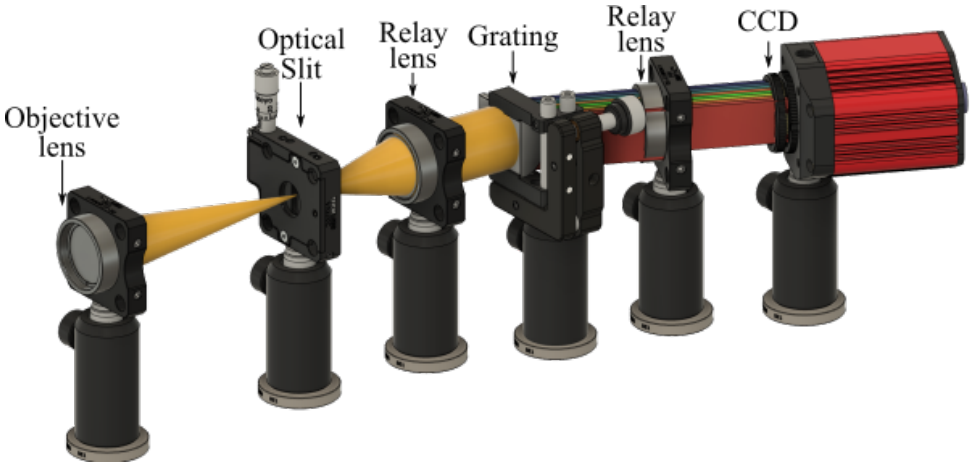
2.1. Spectral Imaging

Spectral imaging (SI) combines two-dimensional imagers and spectroscopy approaches to acquire spatial information across a wide range of wavelengths. SI can be represented as a 3D data cube with two spatial dimensions and a spectral one. Since SIs contains the scene's spectral features, referred to as spectral signatures, which are distinctive and unique for elements found in nature, they have been extensively used in remote sensing frameworks, e.g., classification and detection. Traditional SI sensing approaches rely on scanning protocols that broom an arbitrary fixed dimension to disambiguate the remaining two. These time-consuming sensing protocols were introduced as a solution to the commercial detector limitations, i.e., 1D or 2D pixel arrays to sense a 3D datacube. The most popular scanning imagers are the

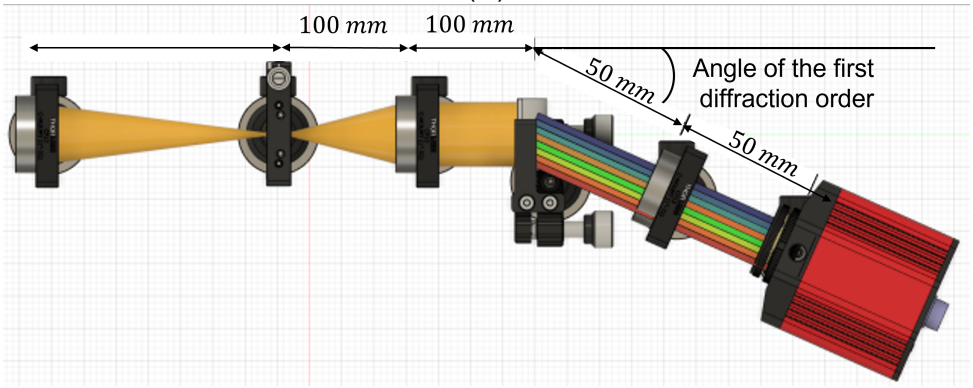
Whiskbroom, Pushbroom, and Filter-based.

2.1.1. Line scan/Pushbroom For the last three decades, the Pushbroom hyperspectral technique has been the most popular and widely used in remote sensing applications. Pushbroom imagers consist of an optical lens that focuses the light coming from a portion of the scene onto the focal plane where the sensor is placed Gupta y Hartley, “Linear pushbroom cameras”; Gómez-Chova y col., “Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images”. The system relies on a slit that limits the area being imaged to a stripe aligned with one of the sensorâ€™s axes, while a diffractive element (e.g., grating) forms a spectrum of the line along the orthogonal axis. The detector is a two-dimensional array whose columns separate wavelengths and whose rows separate resolved spatial points. Conventionally, the linear pushbroom sensor is mounted on a moving platform to sense the orthographic axis spatial/spectral information.

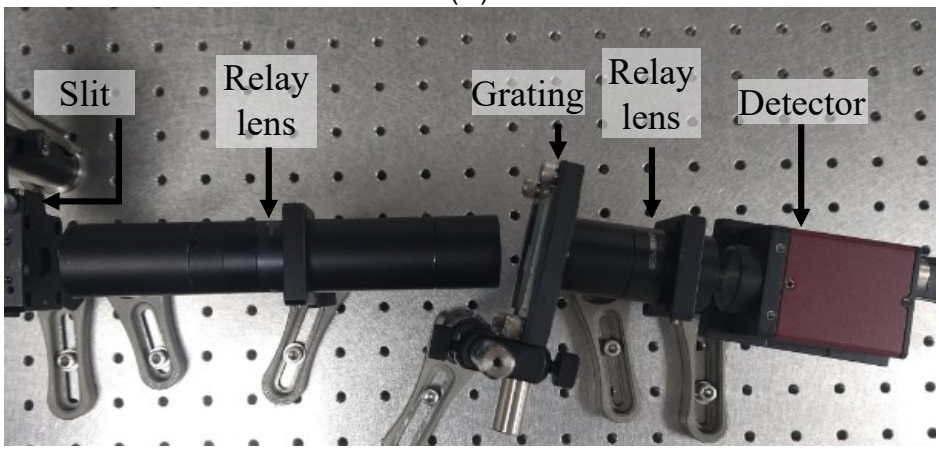
A testbed of the linear Pushbroom imager was assembled in the HDSP laboratory to study their main pros and cons. Moreover, this system was used to acquire a hyperspectral database. The Pushbroom system is composed of an objective lens that focuses the scene light rays to the entrance of an adjustable mechanical slit (Thorlabs VA100C-30 mm, 8-32 Tap), as shown in Fig. 2. The entrance slit has a width of 13.59 mm and a height of 20 μm. Then, a relay lens located at 100 mm from the slit forms a parallel beam that reaches the 600 grooves/mm transmission grating that diffracts the rays. Lastly, the second relay lens located at 50 mm from the grating focuses the diffracted light rays onto the sensor. With the Pushbroom testbed assembled, we proceed to characterize it using a spectrally tunable light source. The tunable source was configured to generate spectral light with a bandwidth of $3.72nm$. Inset (A) in Fig. 3 illustrates the characterized Pushbroom imager spectrogram obtained by illuminating the slit with 281 different wavelengths as shown inset (B) in Fig. 3 in the spectral range $420 - 700$ nm with steps of 1 nm. Figure 4 illustrates a



(A)

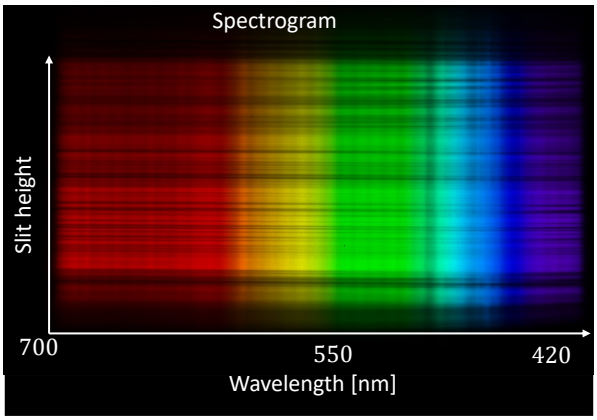


(B)

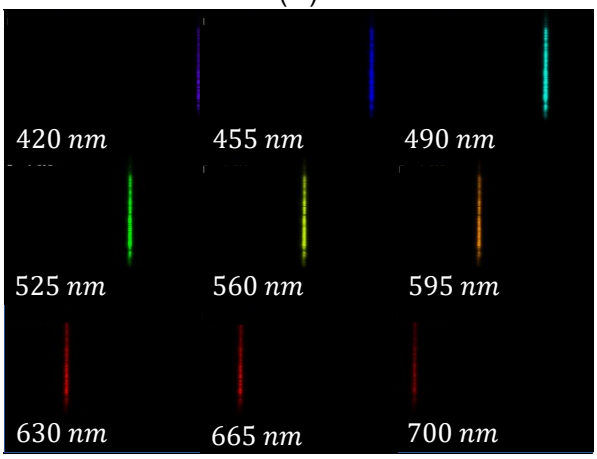


(C)

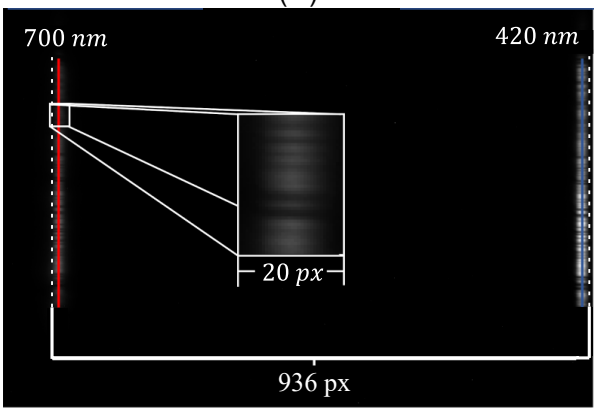
Figura 2. Testbed implementation of the linear Pushbroom imager. (A) Pushbroom imager 3D sketch illustration, where is specify the optical elements. (B) Pushbroom imager 3D sketch illustration, where is specified the optical element distances. (C) Real testbed implementation. *These experiments were carried up in the HDSP laboratory as part of the presented Ph.D.-dissertation.*



(A)



(B)



(C)

Figura 3. Pushbroom spectral characterization. (A) Spectrogram in the wavelength range $420 - 700$ nm. (B) Illustration of 9 out of the 281 spectral bands characterized via spectrally tunable light source. (C) Pushbroom spectral resolution characterization in terms of a camera pixel, where one nanometer corresponds to $\sim$5 pixels.

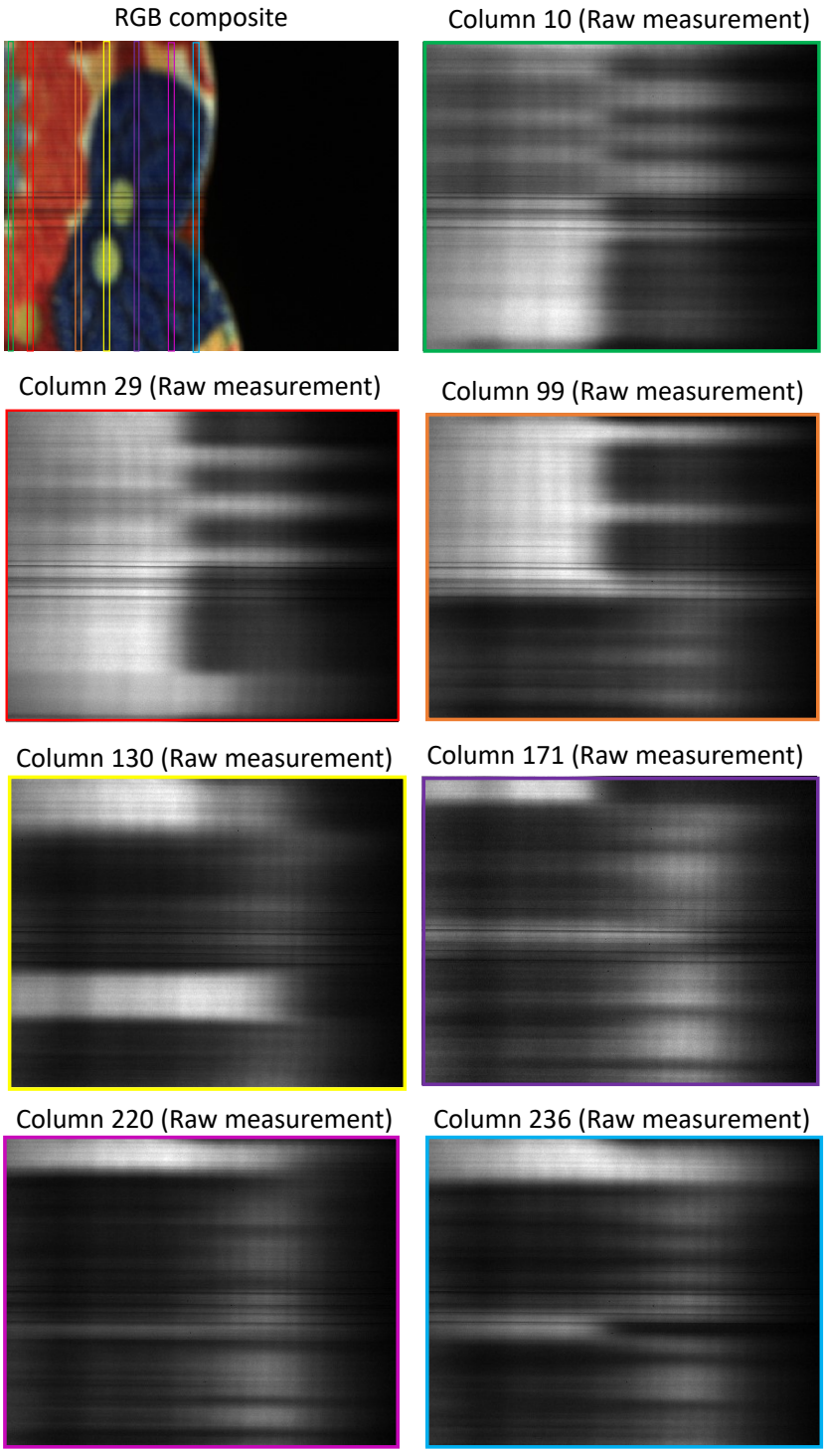
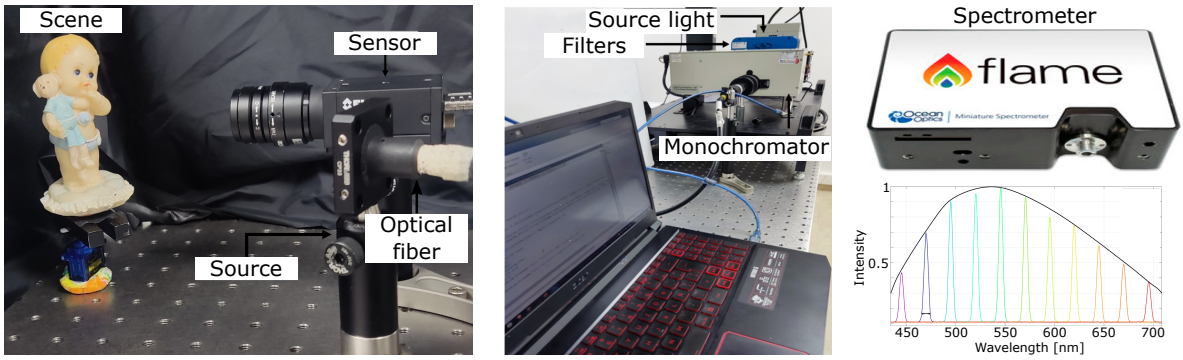


Figura 4. Pushbroom spectral acquisition. (Upper-left image) RGB composition of the acquired spectral datacube. Seven raw measurements for the 10, 29, 99, 130, 171, 220 and 236 acquisitions. Note that the RGB composite shows some spatial quality degradation associated with misalignments generated by the positioner.

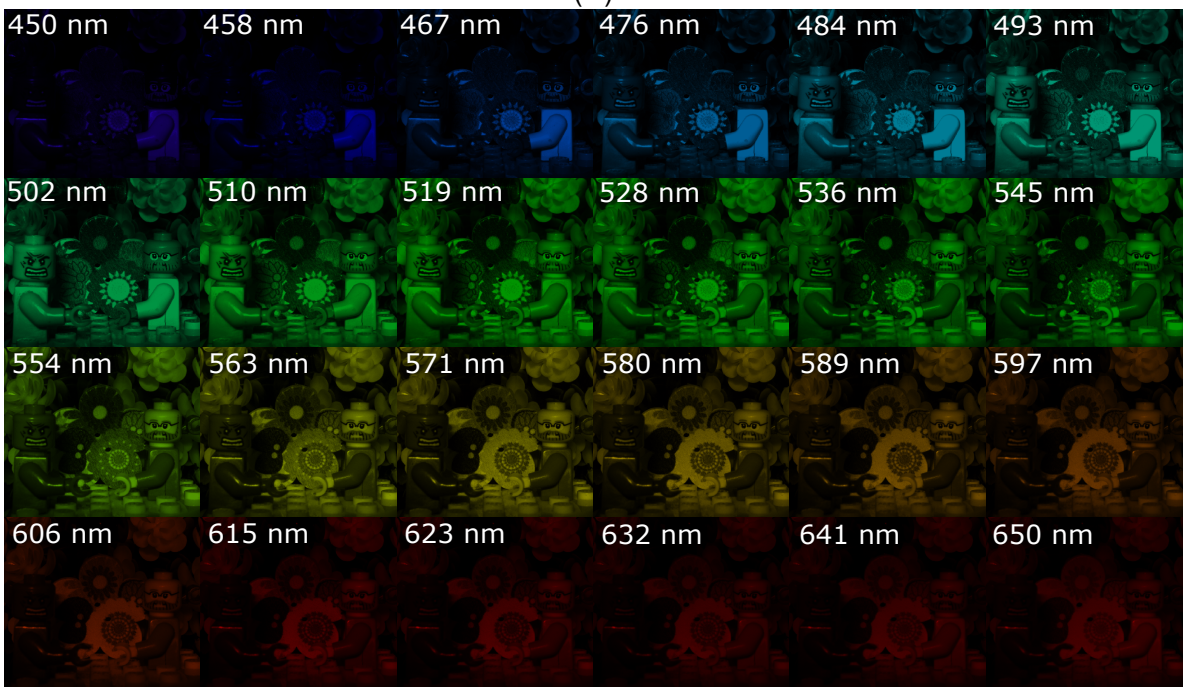
spectral datacube acquisition via the testbed implemented in our HDSP laboratory. For this acquisition, 512 horizontal steps were acquired with a resolution of 48 μm. Note that the horizontal steps represent the total amount of columns and the camera detector size corresponds to the row and spectral bands, respectively. The resulting datacube has a spatial dimension of 776×512 pixels and 936 spectral bands , where one nanometer corresponds to 5 pixels. To better comprehend the Pushbroom data acquisition, Fig. 4 shows seven raw measurements corresponding to steps 10, 29, 99, 130, 171, 220 and 236.

2.1.2. Filter-based approach The Filter-based scan imagers record the whole spatial dimension and scan the spectral domain. Traditionally, these imagers rely on using a set of narrow band-pass filters mounted on a spinning disk (filter wheel). As the disk rotates in sync with the frame rate of the CCD detector, different filters perpendicularly cross the optical path, leading to wavelength selection. It is worth noting that the spectral scanning could be implemented on either illumination or detection sides. In this path, one of the most traditional illumination approaches is based on optical bandpass filters that allow the record of multiple images at different wavelengths according to the number of filters. Another approach is the tunable light source wavelength imagers implemented by exploiting a thin-film tunable filter Favreau y col., “Excitation-scanning hyperspectral imaging microscope”, acoustic optic tunable filter (AOTF) Sheoran y col., “Swept-source digital holography to reconstruct tomographic images”, or monochromator Castellanos-Gomez y col., “Spatially resolved optical absorption spectroscopy of single-and few-layer MoS2 by hyperspectral imaging”; Torabzadeh y col., “Hyperspectral imaging in the spatial frequency domain with a supercontinuum source”. For the detection sides, the optical bandpass filters, liquidâ€“crystal tunable filters, and AOTF can be located in front of the sensor to acquire a specific wavelength of the filtered light.

The spectral scanning method can be transformed from a conventional optical sys-



(A)



(B)

Figura 5. (A) Filter-based scan testbed implementation (left column), TLS- 300XR monochromator optical element (middle column), and light-source spectral characterization via spectrometer Flame-S (right column). (B) A representative acquired spectral datacube with a dimension size of $512 \times 512 \times 24$. *These experiments were carried up in the HDSP laboratory as part of the presented Ph.D. dissertation.*

tem into a hyperspectral imager by developing minimal modifications (e.g., adding a spectral scanning unit on the illumination or detection sides). Another advantage of this approach is its simplicity and relatively low cost. However, such systems are not well suited for spectral analysis that contains complex samples because their spectral resolution is determined by the number of filters and the bandwidth. A testbed of the filter-based imager was built in our laboratory to study their main pros and cons. Moreover, this system was used to acquire a multispectral database. The filter-based system is composed of a tunable light source (Newport TLS130B-300X, 300-1800 nm, 300 W Ozone Free Xe) coupled to an optical fiber to illuminate the scene, and a simple camera-lens system, as shown in Fig. 5(A). With the filter-based testbed assembled, we proceed to characterize it using a spectrometer Flame-S (wavelength range: 350 nm - 1000 nm; optical resolution: 1.34 nm FWHM). For this characterization, the tunable source was configurated to obtain a spectral resolution of 5nm. Figure 5(A) (upper row, right column) illustrates the characterized filter-based imager. Finally, Fig. 5(B) illustrates a spectral datacube acquisition via the testbed implemented in our HDSP laboratoy. Note that obtained multispectral image is composed of 24 spectral bands.

2.2. Depth imaging

Depth estimation or extraction refers to the set of techniques and algorithms that seek to obtain a representation of the spatial structure of a scene. Depth images (DI) can be interpreted as a 2D matrix that contains the scene's (x,y,z) location. In other words, the depth images allow obtaining a measure of the object-sensor distance. Traditionally, DIs are acquired using stereoscopic cameras, structured light, photon counters, microlens arrays, and liquid lenses. These systems embed the depth information into disparity maps, light pattern deformations, multiple angular views, depth-from-defocus maps, and time of flight maps.

2.2.1. Depth from defocus (DFD) The family of methods requiring the acquisition of only one or two images, each with different focal point values, is commonly known as *Depth-From-Defocus (DFD)*. The DFD problem aims to analyze the relative blur of each point of the scene in the two images acquired with different focal points and calculate the depth of the scene's elements. Although the basic theory behind this technique is well known, DFD has found limited use in practice because it is broadly understood to require static scenes, dense surface texture, and images with significant defocus blur. Figure 6 illustrates the defocus effect for an ideal punctual source in a simple camera lens. This optical response is commonly known as the point-spread function (PSF), and it is used to describe the imaging system's response. The DFD-based approaches model light transport in the camera using wave optics to model the depth-dependent scene information. This modeling works for systems constructed with either refractive or diffractive optical elements, where the light emitted by a point –located at the infinity– is represented as a spherical wave. Then, the resulting complex-valued wave field immediately before the lens is given by:

$$W(u,v,z) = \exp\left(ik\sqrt{u^2+v^2+z^2}\right), \tag{1}$$

where (u, v) represents the input plane, z is the object-lens distance, and $k = 2\pi/\lambda$ is the wavenumber. The refractive optical element first delays the phase of this incident wavefront by an amount proportional to the deformable mirror (DM) ϕ of the optical element at each point (u, v). Equivalently, this phase transformation can be mathematically represented as

$$t_{\phi}(u, v) = \exp(ik(n(\lambda) - 1)\phi(u, v)), \tag{2}$$

where $n(\lambda)$ is the wavelength-dependent refractive index of the optical element material and $\phi(u, v)$ represents the lens's phase. The light wave continues to propagate to the camera lens, which induces the following phase transformation J.W. Goodman. "Introduction to Fourier Optics". En: McGraw-Hill physical and quantum electronics series (2005), págs. 1-491

$$t(u, v) = \exp\left(-i\frac{k}{2f}(u^2 + v^2)\right), \tag{3}$$

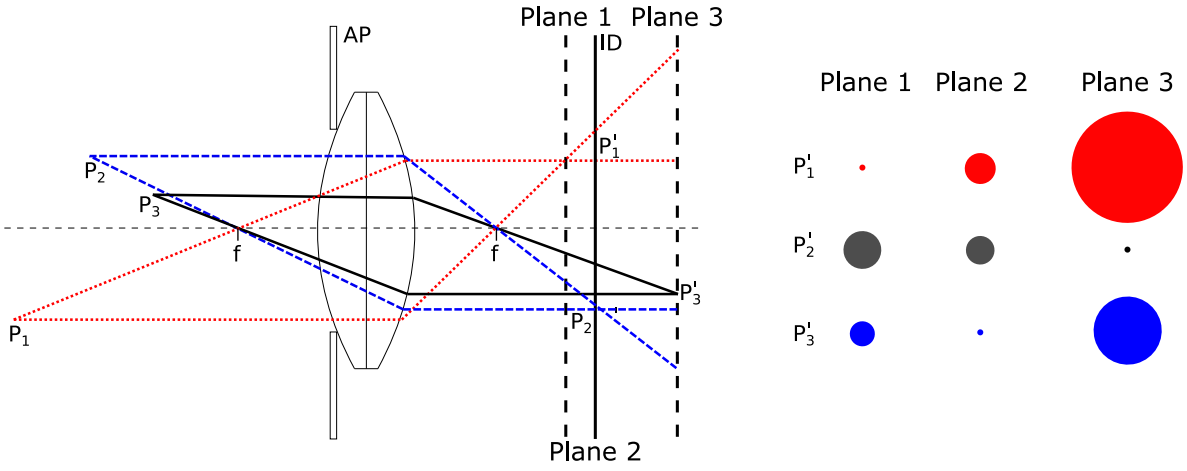


Figura 6. Blur effect for three different depth planes. This figure analyzes the aberrations associated with different depth planes. Here, the colors are used to differentiate the three planes.

where f represents the lens' focal length. A binary circular mask $A(u, v)$ with diameter D was used to model the aperture and block light in regions outside the open aperture. To find the electric field immediately after the lens, the amplitude and phase modulations of the refractive optical element and lens are multiplied with the input electric field:

$$\tilde{W}(u, v) = A(u, v)t_{\phi}(u, v)t(u, v)W(u, v). \tag{4}$$

Finally, the field propagates a distance d_2 to the sensor with the transfer function Goodman, "Introduction to Fourier Optics":

$$T(f_u, f_v) = \exp\left[ikd_2\sqrt{1 - (\lambda f_u)^2 - (\lambda f_v)^2}\right], \tag{5}$$

where (f_u, f_v) are spatial frequencies. This transfer function is applied on the Fourier domain as:

$$\bar{W}(u', v') = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{\tilde{W}(u, v)\right\}_{\frac{f_u}{\lambda z}, \frac{f_v}{\lambda z}} \cdot T(f_u, f_v)\right\}_{\frac{u'}{\lambda f}, \frac{v'}{\lambda f}}, \tag{6}$$

where $\mathcal{F}$ denotes the 2D Fourier transform. Since the sensor measures light intensity, the magnitude-squared is taken to find the values of the PSF $\mathbf{H}$ at each position (u, v) as:

$$H(u', v') = |\bar{W}(u', v')|^2. \tag{7}$$

Inspired by the DFD-based depth systems' ability to encode (via Eq. (7)) and infer the depth information from a simple Tang y col., "Depth from Defocus in the Wild" or complex Chang y Wetzstein, "Deep Optics for Monocular Depth Estimation and 3D Object Detection" geometrical/chromatic aberrations, several works have proposed to design the lens structure to promote low-correlated PSF response per depth plane. This optical element is commonly known as diffractive optical element (DOE).

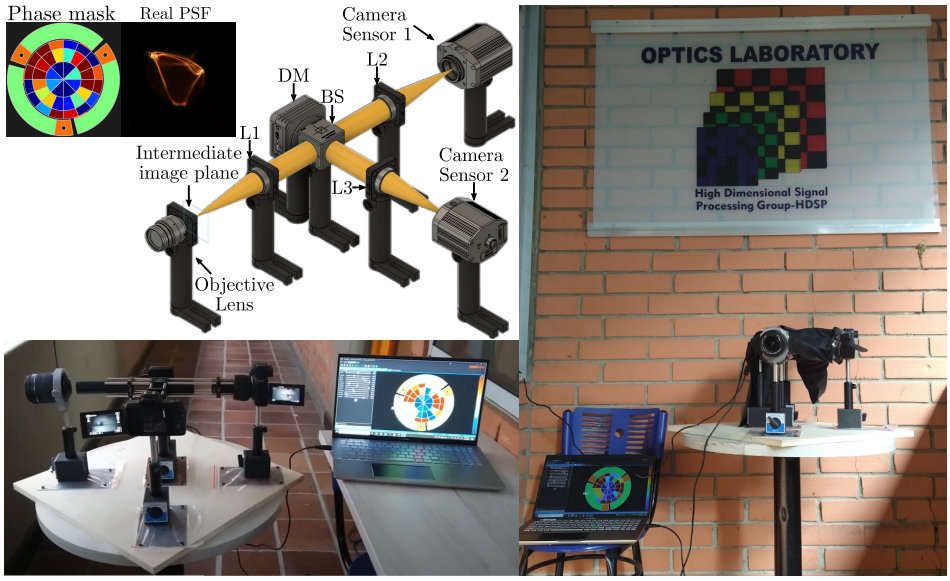


Figura 7. Deformable mirror-based depth-from-defocus system. The focus swept is emulated by deforming the mirror surface using a fourth-order Zernike polynomial shape. *These experiments were carried up in the HDSP laboratory as part of the presented Ph.D. dissertation.*

On the other hand, recent works have exploited the deep-learning capabilities to train algorithms capable of obtaining rough depth estimation. This pure-deep-based approach relies on learning either an arbitrary camera-lens system's geometrical or chromatic aberrations.

PSF-based imagers configurations Based on the depth-dependent PSF principle, depth acquisition was tackled into three fundamental categories: DFD, DOE-based imagers, and pure-deep-based approach.

- **Depth from defocus** Tang y col., "Depth from Defocus in the Wild": This DFD-based approach relies on enforcing a tiny blur condition –by controlling camera focus– to promote the preservation of the high spatial frequencies, allowing near-optimal depth discrimination. Specifically, the blur condition dictates that the radius of the defocus blur kernel is at most three pixels for the scene points.

Following the blur condition, Tang y col., “Depth from Defocus in the Wild” relies on the acquisition of two images taken under controlled focus, where the focus setting of the first image is chosen freely, and the focus of the second is determined to satisfy the tiny blur condition. To emulate and study the controlled focus acquisition sensing process, a deformable mirror (DM)-based testbed system was implemented in the HDSP lab, as shown in Fig. 7. Specifically, the DM is an optoelectronic element that works as a phase mask with the ability to carry out a sweeping sensing protocol by introducing a spherical phase mask, i.e., emulating the lens phase.

- **Deep optics for monocular depth estimation** Chang y Wetzstein, “Deep Optics for Monocular Depth Estimation and 3D Object Detection”: The paradigm of deep optics –deep learning-based end-to-end optics design– inspires this monocular DOE-based depth imager which aims to use the coded defocus blur as an additional depth cue to be decoded by a neural network. Specifically, Chang y Wetzstein, “Deep Optics for Monocular Depth Estimation and 3D Object Detection” proposes interpreting the depth estimation problem with coded defocus blur as an optical-encoder, electronic-decoder system that can be trained end-to-end. To emulate and study the PSF encoding paradigm via phase modulations, a deformable mirror (DM)-based testbed system was implemented in the HDSP lab, as shown in Fig. 7. This testbed is inspired by the Zernike polynomials theory and the adaptive optical systems that relied on DM optoelectronic devices.

- **Towards robust monocular depth estimation**Ranftl y col., “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer”: This approach introduced a pure-deep-based depth algorithm that allows inferring the depth information from a single RGB image acquired by a commercial camera or a smartphone. This approach relies on mixing multiple labeled

datasets during training to learn the depth-dependent geometrical and chromatic aberrations. This training approach results in a depth algorithm robust to changes in depth range and scale changes.

2.2.2. Structured light (SL) Structured light (SL) is one of the most popular 3D surface imaging approaches based on active illumination. SL methodologies rely on illuminating the scene with a specially designed 2D spatially varying pattern M. Servin, J.A. Quiroga y M. Padilla. *Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications*. Wiley, 2014. The structured illumination is generated by a special projector or a light source modulated by a spatial light modulator. Specifically, the SL-based depth imagers are composed of a light projector and a video camera that acquire a 2D image of the scene under the structured-light illumination. The SL paradigm aims to encode and extract the 3D surface shape based on the light patterns' deformations. Accurate 3D surface profiles of objects in the scene can be computed by using various structured light principles and algorithms.

Phase-shifting interferometry theory An ideal stationary fringe pattern is usually modeled by

$$I(x,y) = a(x,y) + b(x,y)\cos\left(\varphi(x,y)\right), \tag{8}$$

where (x,y) index the spatial coordinates, $\varphi(x,y)$ is the unknown phase information, $b(x,y)$ represents the intensity object, and $a(x,y)$ is the background. The main objective of SL is to demodulate the phase function $\varphi(x,y)$ from the interferometric data $I(x,y)$ Servin, Quiroga y Padilla, *Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications*. The estimation of $\varphi(x,y)$ represents an ill-posed inverse problem since $a(x,y)$ and $b(x,y)$ are also unknown functions. Moreover, this inverse problem has two critical ambiguities: $\varphi(x,y)$ is bounded $0 \leq \varphi(x,y) \leq 2\pi$, and at least two measurements are requested to extract the sign ambiguity. Equation 8 is rewritten using the analytic cosine function representation to

eliminate the stationary fringe ambiguities, resulting in

$$I(x,y) = a(x,y) + \frac{1}{2}b(x,y)\left[e^{i\varphi(x,y)} + e^{-i\varphi(x,y)}\right]. \tag{9}$$

Here, it is highlight that all the necessary information to obtain a wrapped approximation of $\varphi(x,y)$ is contained in the term $e^{i\varphi(x,y)}$ of Eq. 9. Specifically, this wrapped version can be obtained by

$$\hat{\varphi}(x,y) = \arctan\left[\frac{\mathsf{Im}\{(1/2)b(x,y)e^{i\varphi(x,y)}\}}{\mathsf{Re}\{(1/2)b(x,y)e^{i\varphi(x,y)}\}}\right], \tag{10}$$

where $\hat{\varphi}(x,y)$ is a wrapped modulo 2π projection of the phase $\varphi(x,y)$, $\mathsf{Re}\{\cdot\}$ and $\mathsf{Im}\{\cdot\}$ are two operator that returns the real numbers of a complex number. However, researchers found that reducing Eq. 9 to Eq. 10 is impossible in its current form because all components of the interferogram overlap both in space and in the frequency domain.

Temporal Carrier Interferometric Signal To resolve the phase estimation ambiguities, the scientific vision community resolved to introduce controlled changes in the sinusoidal signal argument. The acquisition of several samples with different controlled changes along the time can be expressed as Servin, Quiroga y Padilla, *Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications*:

$$I(x,y,t) = a(x,y) + b(x,y)\cos\left(\varphi(x,y) + t\omega_0\right), \tag{11}$$

where $t\omega_0$ is the temporal-carrier phase modulation. Taking the temporal Fourier transform of Eq. 11, the temporal phase-shifted interferogram $I(x,y,t)$ can descri-

bed as

$$I(x,y,\omega) = (1/2)b(x,y)\left[e^{i\varphi(x,y)}\delta(\omega-\omega_0) + e^{-i\varphi(x,y)}\delta(\omega+\omega_0)\right] + a(x,y)\delta(\omega) \tag{12}$$

where $\delta(\cdot)$ represents the delta Dirac and the phase step ω_0 is assumed to be known. To use the $\arctan(\cdot)$ approach established in Eq. 10, it is needed to remove the negative term ($e^{-\varphi(x,y)}$) from Eq. 12 by selecting a proper linear filter that rejects the signals at $\omega = 0$ and $\omega = -\omega_0$ Servin, Quiroga y Padilla, *Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications*. In this direction, linear filters that satisfy these rejecting condition are those frequency transfer function fulffil the so-called quadrature conditions $H(0) = H(-\omega_0) = 0$ and $H(\omega_0) \neq 0$, where $H(\omega) = \mathcal{F}\{h(t)\}$ represents the transfer function of the quadrature linear filter (QLF) $h(t)$. Specifically, the QLF can be defined as $h(t) = \sum_{k=0}^{K-1} c_k\delta(t-k)$ (i.e., a Dirac comb with constant step) with $H(\omega) = \sum_{k=0}^{K-1} c_k e^{-i\omega k}$, where $c_k \in \mathbb{C}$. Then, applying the quadrature linear filter $h(t)$ to the temporal phase-shifted interferogram in Eq. 11, it is obtained by

$$I(t) * h(t) = I(t) * \sum_{k=0}^{K-1} c_k\delta(t-k) = \sum_{k=0}^{K-1} c_k I(t-k), \tag{13}$$

where $*$ denotes the convolution operator. The Fourier transform in the left part of Eq. 13 resulting in

$$I(\omega)H(\omega) = H(0)a(x,y)\delta(\omega) + \frac{b(x,y)}{2}[H(\omega_0)e^{i\varphi(x,y)}\delta(\omega-\omega_0) + H(-\omega_0)e^{-i\varphi(x,y)}\delta(\omega+\omega_0)] \tag{14}$$

for $[\delta(\omega-\omega_0)H(\omega)] = \delta(\omega-\omega_0)H(\omega_0)$, $[\delta(\omega+\omega_0)H(\omega)] = \delta(\omega+\omega_0)H(-\omega_0)$, and $H(\omega)\delta(\omega) = H(0)\delta(\omega)$, where appliying the QLF condition Servin, Quiroga y Padilla,

Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications, Eq. 14 can be rewritten as

$$I(\omega)H(\omega) = \frac{1}{2}H(\omega_0)b(x,y)e^{i\varphi(x,y)}\delta(\omega-\omega_0). \tag{15}$$

Then, returning Eq. (15) to the temporal domain and combining with Eq. (13), results in

$$I(t) * h(t) = \sum_{k=0}^{K-1} c_k I(t-k) = \frac{1}{2}b(x,y)H(\omega_0)e^{i[\varphi(x,y)+\omega_0 t]} \tag{16}$$

where $I(t)*h(t)$ has a support of $2K-1$ samples. In this direction, the computer vision community have been established that the temporal convolution has to evaluate at $t = K-1$ in order to obtain the most robust estimation of the analytic signal for the given number of phase steps

$$\frac{1}{2}b(x,y)H(\omega_0)e^{i\varphi(x,y)} = \sum_{k=0}^{K-1} c_k I_{K-1+k}. \tag{17}$$

Equation (17) is commonly known as the general formula for a linear phase shifting analysis (PSA). Taking the angle part of Eq. (17), and applying some straightforward manipulations, a wrapped modulo 2π projection of $\varphi(x,y)$ can be calculated by

$$\hat{\varphi}(x,y) = \arctan\left(\frac{\sum_{k=0}^{K-1} b_k(x,y)I_{K-1+k}(x,y)}{\sum_{k=0}^{K-1} a_k(x,y)I_{K-1+k}(x,y)}\right) \tag{18}$$

where $c_k = a_k(x,y) + ib_k(x,y)$. The last equation is the arc-tan formulation of a linear PSA.

N-step least-square PSA One of the main limitations of Eq. (18) is that c_k is an unknown function that varies depending on the number of used projection steps K.

A more detailed explanation about the c_k calculation can be found in Servin, Quiroga y Padilla, *Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications*. To overcome this limitation in J Herriot Bruning y col. “Digital wavefront measuring interferometer for testing optical surfaces and lenses”. En: *Applied Optics* 13.11 (1974), págs. 2693-2703 was introduced the N-step least-square PSA (LS-PSA) formula and in CJ Morgan. “Least-squares estimation in phase-measurement interferometry”. En: *Optics Letters* 7.8 (1982), págs. 368-370 was demonstrated that the LS-PSA provides a better phase estimation against the nonuniform phase steps. Finally, in Yeou-Yen Cheng y James C Wyant. “Phase shifter calibration in phase-shifting interferometry”. En: *Applied Optics* 24.18 (1985), págs. 3049-3052 was established the general N-step LS-PSA formula, which is given by

$$\hat{\varphi}(x,y) = \arctan\left(\frac{\sum_{k=0}^{K-1}\sin(k\omega_0)I_{K-1+k}(x,y)}{\sum_{k=0}^{K-1}\cos(k\omega_0)I_{K-1+k}(x,y)}\right) \tag{19}$$

where $\omega_0 = \frac{2\pi}{K}$ and $K \geq 3$. Due to the truncation effect of the arctangent function in (19), the obtained phase $\varphi(x,y)$ is wrapped within the range of $(0, 2\pi)$, and its relationship with $\varphi(x,y)$ is

$$\varphi(x,y) = \hat{\varphi}(x,y) + 2\pi\xi(x,y), \tag{20}$$

where $\xi(x,y) \in \mathbb{Z}$ represents the fringe order of $\varphi(x,y)$.

Multifrequency Heterodyne Principle To remove the modulo 2π ambiguity inherent in Eq. 19, several works have been proposed spatial phase unwrapping (SPU) methodologies Jonathan M Huntley y Henrik O Saldner. “Shape measurement by temporal phase unwrapping: comparison of unwrapping algorithms”. En: *Measurement Science and Technology* 8.9 (1997), págs. 986-992. In this path, one widely used method is the multifrequency time phase unwrapping (MF-TPU) method ba-

sed on the heterodyne principle. MF-TPU methods determine fringe orders using information from different coded patterns or phase maps with different frequencies generated from additional sets of phase-shifted images captured over time. Specifically, based on the N-step PSA approach, the projected fringe patterns intensities can be written as

$$I_k^p(x,y) = 0.5 + 0.5\cos\left(2\pi\nu x + \frac{2\pi k}{K}\right), \tag{21}$$

where $p = \{0, ..., P-1\}$ with $P \in \mathbb{N}$ as the total number of used frequencies, $a(x,y) = b(x,y) = \frac{1}{2}$, and $\nu \in \mathbb{R}$ is the fringe frequency. Having defined the fringe patterns structure to being projected, we proceed to acquire and estimate two wrapped phases $\hat{\varphi}_h(x,y)$ and $\hat{\varphi}_l(x,y)$ via Eq. (21) using a high and low fringe frequency pattern, respectively. Inspirated on Wei Yin y col. "Temporal phase unwrapping using deep learning". En: *Scientific reports* 9.1 (2019), págs. 1-12 and Based on Eq. (20), the two absolute phases can be established as $\varphi_h(x,y)$ and $\varphi_l(x,y)$ corresponding to $\hat{\varphi}_h(x,y)$ and $\hat{\varphi}_l(x,y)$, respectively, have the following relationship:

$$\varphi_h(x,y) = \hat{\varphi}_h(x,y) + 2\pi\xi_h(x,y), \tag{22}$$

and

$$\varphi_l(x,y) = \hat{\varphi}_l(x,y) + 2\pi\xi_l(x,y), \tag{23}$$

where $\xi_h(x,y) \in \mathbb{Z}$ and $\xi_l(x,y) \in \mathbb{Z}$ are the high and low fringe order, respectively. Based on the principle of MF-TPU, $\xi_h(x,y)$ can be calculated by the following formula:

$$\xi_h(x,y) = \text{round}\left[\frac{(\nu_l/(\nu_h - \nu_l))|\varphi_h(x,y) - \varphi_l(x,y)| - \hat{\varphi}_h(x,y)}{2\pi}\right], \tag{24}$$

where ν_h and ν_l are the frequency of high-frequency fringes and low-frequency fringes, and round$(\cdot)$ is the rounding operation. Here it is worth noting that if ν_l is equal to 1, there will be no phase ambiguity so that $\varphi_l(x, y)$ is inherently an unwrapped phase, i.e., $\varphi_l(x, y) = \hat{\varphi}_l(x, y)$ in Eq. (23).

Testbet implementation A testbed of the SL-based imager was built in the optical laboratory to study their main pros and cons. Moreover, this system was used to acquire a multispectral database. The SL-based system is composed of a commercial video projector, and a simple camera-lens system, as shown in Fig. 8. For the depth recovery, eight fringe light patterns were projected –four with a high frequency and four with a low frequency– constructed following the multifrequency heterodyne principle. Note that the illuminated object has a maximum height of ~ 1 cm. The structured light fringes projected by the video-beam are generated by the equation

$$I^j_{i_1,i_2} = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi\nu(i_1 - i_2)}{L} - \frac{2\pi j}{K}\right), \tag{25}$$

where ν is the frequency of fringe patterns, $K \in \mathbb{N}$ is the total of pattern projections per frequency, $L \in \mathbb{N}$ is the total number of pixels in the transverse direction of the fringe pattern, and $j = \{0, ..., K - 1\}$. In this regard, the parameters selected for the experiments are $K = 4$, $L = 1388$, and $\nu = \{14, 2\}$.

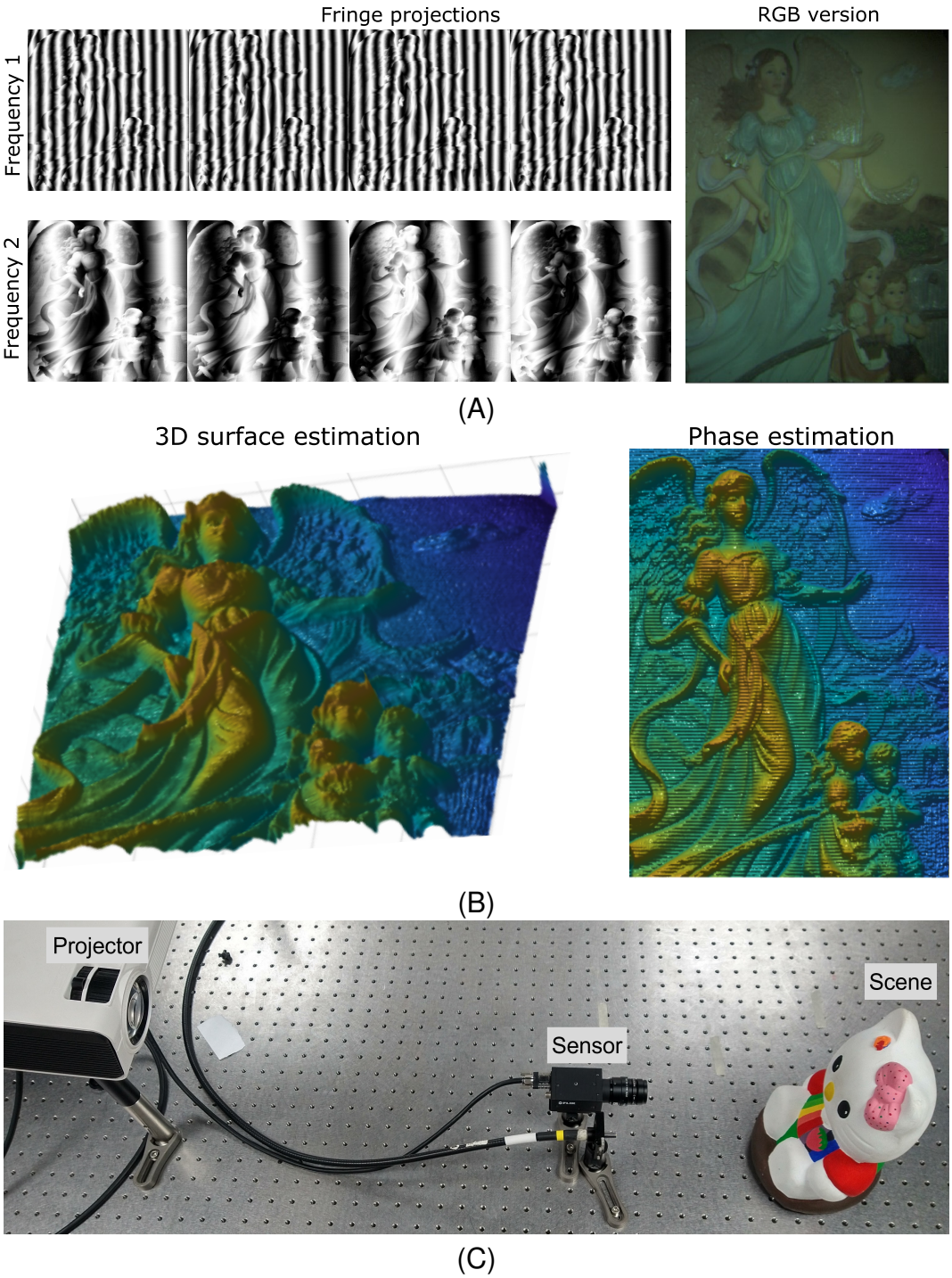


Figura 8. A structured light optical system composed of a video beam as a projector and a camera. *These experiments were carried up in the HDSP laboratory as part of the presented Ph.D. dissertation.*

2.2.3. Light field (LF) In the last two decades, the interpretation of light flow as a high dimensional space has become increasingly popular to describe the information associated with the radiance of a scene Levoy, "Light fields and computational imaging"; Adelson, Bergen y col., *The plenoptic function and the elements of early vision*. In particular, the 7-dimensional plenoptic function $f(x, y, z, \gamma, \theta, \lambda, t)$ has been introduced as a function of space (x, y, z), orientation (γ, θ), wavelength λ and time t Christopher Gilliam, Pier-Luigi Dragotti y Mike Brookes. "On the spectrum of the plenoptic function". En: *IEEE Transactions on Image Processing* 23.2 (2014), págs. 502-516. Plenoptic imaging refers to the framework of acquiring and processing different dimensions of the plenoptic function Adelson, Bergen y col., *The plenoptic function and the elements of early vision*; Leonard McMillan y Gary Bishop. "Plenoptic modeling: An image-based rendering system". En: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*. ACM. 1995, págs. 39-46; Steven J Gortler y col. "The lumigraph". En: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM. 1996, págs. 43-54. Conventional photography, for instance, captures a projection of the plenoptic function by integrating the angular and spectral domain all over a 2D space on the (x, y) plane Ren Ng y col. "Light field photography with a hand-held plenoptic camera". En: *Computer Science Technical Report CSTR* 2.11 (2005), págs. 1-11. Similarly, spectral imaging aims to capture 2D projections of the plenoptic function at different wavelengths (λ), therefore integrating the plenoptic function over the 3D space (x, y, λ) Gary A Shaw y Hsiaohua K Burke. "Spectral imaging for remote sensing". En: *Lincoln laboratory journal* 14.1 (2003), págs. 3-28.

In a similar end, 3D imaging records the distance from the camera sensor to the target surface, thus sensing the 3D space (x, y, z) Yan Cui y col. "3D shape scanning with a time-of-flight camera". En: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. 2010, págs. 1173-1180. Efforts to

measure more than three-dimensions of the plenoptic function simultaneously, revived the well known light field imaging framework. This framework aims to capture the five dimensions $(x, y, z, \gamma, \theta)$ via a 4D light field approximation (x, y, u, v) that parameterizes the light rays by the coordinates of their intersections with two planes placed at arbitrary positions Marc Levoy y Pat Hanrahan. “Light field rendering”. En: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, págs. 31-42; Kshitij Marwah y col. “Compressive light field photography using overcomplete dictionaries and optimized projections”. En: *ACM Transactions on Graphics (TOG)* 32.4 (2013), pág. 46. The 4D light field is parameterized through the incoming light ray intersection coordinates with the camera and the image planes placed at $z = 0$ and $z = 1$, respectively. The coordinate systems for the first and second planes are denoted by $(u; v)$ and $(s; w)$, respectively, as shown in Fig 9 (A). Specifically, the arbitrary incoming light ray goes through the uv plane at coordinate $(u; v)$ and then intersects the sw plane at coordinate $(s; w)$, and is thus denoted by $\mathcal{L}(u; v; s; w)$. To visualize the positional changes in (u, v) caused by the changing of camera position along a linear path (by a linearly translating camera), a horizontal line is drawn out in the image and the camera plane, each one with a constant coordinate v^* and w^*, respectively. This projection results in the well-known Epipolar plane image (EPI) Yongbing Zhang y col. “Light-field depth estimation via epipolar plane image analysis and locally linear embedding”. En: *IEEE Transactions on Circuits and Systems for Video Technology* 27.4 (2016), págs. 739-747, depicted in Fig. 9 (B). Here, we notice that the EPI projection is composed of simple linear structures even though the photographs contain complex shapes and intensity changes.

Based on the LF geometry illustrated in Fig. 9, the depth value z can be calculated

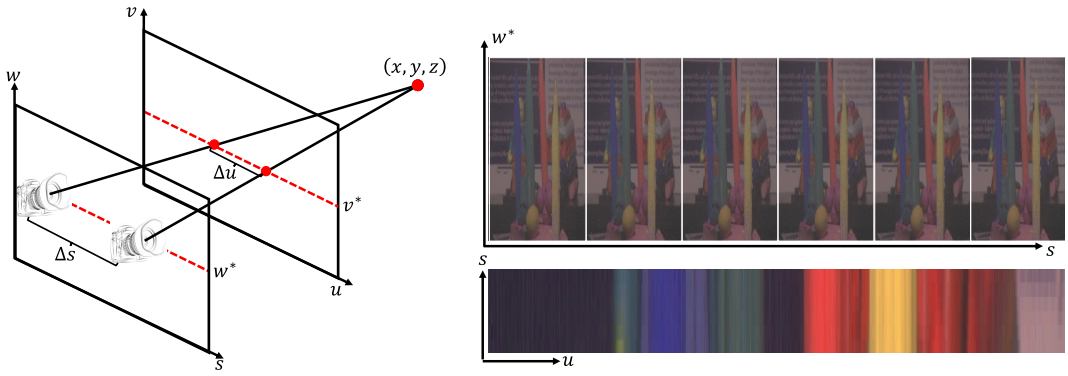


Figura 9. Definition of an EPI. The depicted EPI is related to the horizontal lines in the corresponding images. Inspired by Yongbing Zhang y col. “Light-field depth estimation via epipolar plane image analysis and locally linear embedding”. En: *IEEE Transactions on Circuits and Systems for Video Technology* 27.4 (2016), págs. 739-747.

from the following inverse proportion

$$z = -\frac{\Delta s}{\Delta u}, \tag{26}$$

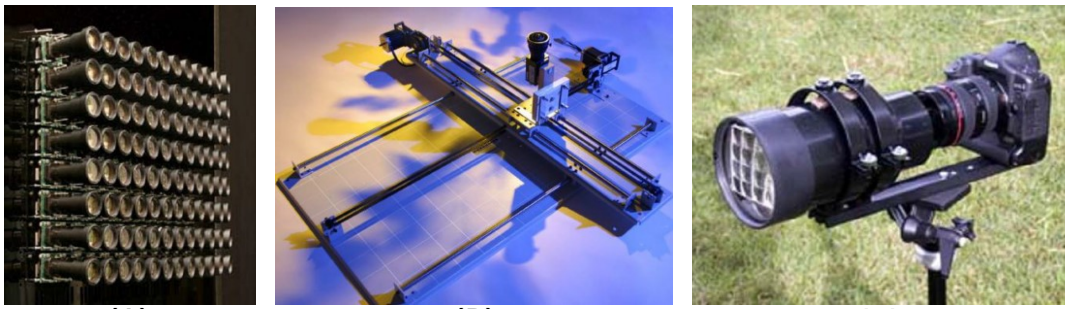
where Δs and Δu are the geometrical distance between two cameras along the line and the distance between the scene points moved in the image plane, respectively. In other words, the orientation of the linear structures corresponds to the depth of the corresponding scene point. This concept is defined as the depth-slope relationship. Based on the EPI orientation principle, several works have developed depth estimation methodologies Zhao y col., “Heterogeneous camera array for multispectral light field imaging”; Kang Zhu y col. “Hyperspectral Light Field Stereo Matching”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018); R. Ng y Stanford University. Computer Science Department. *Digital Light Field Photography*. Stanford University, 2006.

Light field imagers configurations Based on the EPI principle, the 4D light field acquisition was tackled into three fundamental categories: multi-sensor capture, time-sequential capture, and multiplexed imaging.

- **The multi-sensor** approach relies on an array of image sensors distributed on a surface to capture light field samples simultaneously from different viewpoints. The spatial and angular dimensions are determined by the sensors and the number of cameras (and their distribution), respectively. Therefore, the 4D light field is recorded by combining the captured images, as is shown in Fig. 10 (A).

- **The time-sequential** approach relies on a single image sensor mounted on a mechanical gantry to measure the light field at different positions, as is shown in Fig. 10 (B). Against the multi-sensor systems, the costs of the time-sequential capture systems are lower. In addition, these systems can capture light fields at dense angular resolutions at the expense of the temporal resolution, i.e., these systems are suitable just for motionless scenes. A testbed implementation developed in the HDSP laboratory is illustrated in Fig. 11 (A). Insets (B) and (C) in Fig. 11 shows the resulting RGB composite light field measurement and the resulting depht map estimation.

- **The multiplexing** imaging aims to encode the 4D light field into a 2D sensor plane by multiplexing the angular into the spatial domain. This approach enables acquiring dynamic light field information with a single image sensor but imposes a tradeoff between the spatial and angular resolutions. Multiplexed imaging can be further divided into spatial multiplexing and frequency multiplexing Gaochang Wu y col. “Light field image processing: An overview”. En: *IEEE Journal of Selected Topics in Signal Processing* 11.7 (2017), págs. 926-954. The frequency multiplexing approach relies on modulation masks to achieve a



(A) (B) (C)

Figura 10. Light Field imagers. (A) Camera array system composed of 8×12 video cameras Bennett Wilburn y col. “High performance imaging using large camera arrays”. En: 24.3 (2005), págs. 765-776. (B) A motorized linear stage for mounting a single camera to capture 4D light fields Jonas Unger y col. “Capturing and rendering with incident light fields”. En: *Proc. Eurograph. Workshop Rendering Techn.* 2003, 141â€“-149. (C) An integral camera that uses a 4×5 array of lenses and prisms to capture a low angular resolution light field Todor G Georgiev y col. “Spatio-Angular Resolution Tradeoffs in Integral Photography.” En: *Proceedings of the 17th Eurographics Conference on Rendering Techniques*. Goslar, DEU: Eurographics Association, 2006, 263â€“272. DOI: 10.2312/EGWR/EGSR06/263-272.

particular property in the Fourier domain to encode different 2D slices of the light field into different frequency bands. Spatial multiplexing relies on using an interlaced lens, called lenslet array, to acquire a set of elemental images that represent samples from different 2D slices of the light field are captured by the sensor, as is shown in Fig. 10 (C). It enables capturing a light field in a single image with a single exposure. However, their main issue is the trade-off between angular and spatial resolution at the image sensor. Although their resolution issues, this approach has been the most extensively used to capture light field images.

Lenslet array light field cameras Light field cameras based on multiplexing theory are generally divided into unfocused light field (ULF) Adelson y Wang, “Single lens stereo with a plenoptic camera”; Ng y Department, *Digital Light Field Photography* and focused light field (FLF) Andrew Lumsdaine y Todor Georgiev. “The fo-

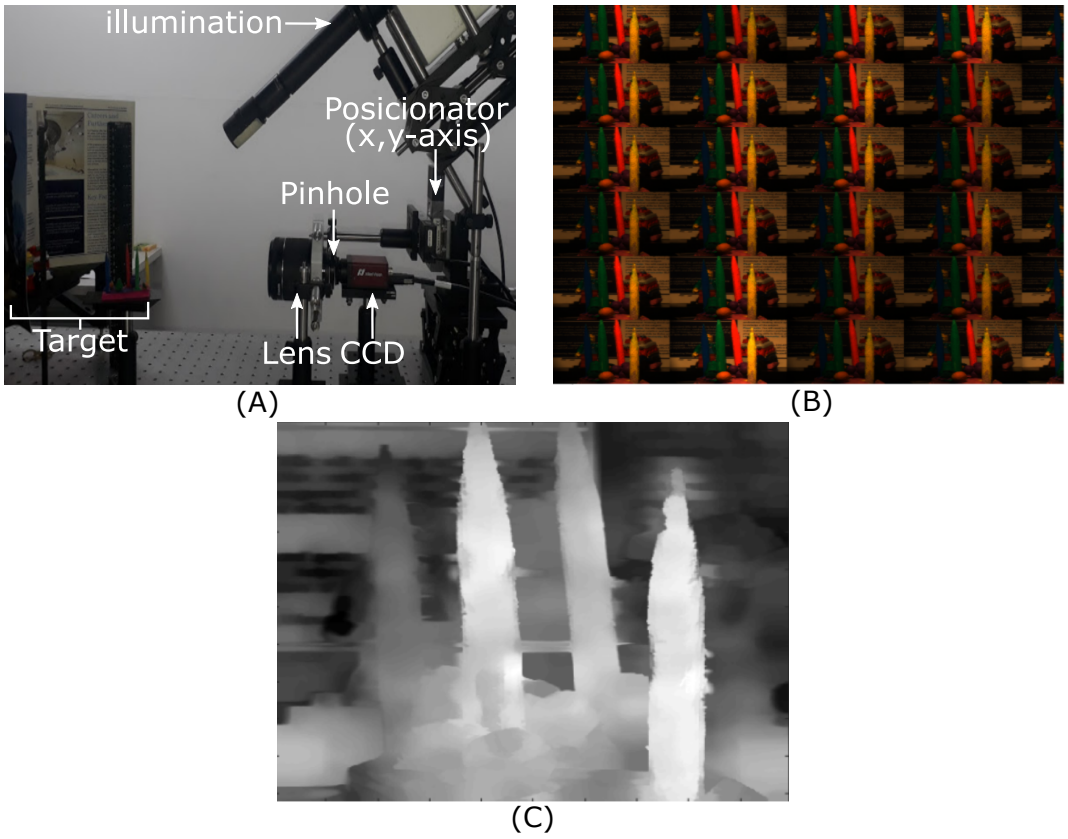


Figura 11. Time-sequential light field imagers. (A) Testbed implementation. (B) Light field acquisition with 6×4 views angles. (C) Depth estimation from the LF data in (B). *These experiments were carried up in the HDSP laboratory as part of the presented Ph.D. dissertation.*

cused plenoptic camera". En: *2009 IEEE International Conference on Computational Photography (ICCP)*. 2009, págs. 1-8. DOI: 10.1109/ICCPHOT.2009.5559008 camera. These systems are called plenoptic 1.0 and 2.0 cameras, respectively. Specifically, these systems can be described as Shuaishuai Zhu y col. "On the fundamental comparison between unfocused and focused light field cameras". En: *Applied Optics* 57.1 (2018), A1-A11:

- In a ULF camera, the lenslet array images the exit pupil onto the detector array because the focal length of the MLA is much shorter than the distance between the exit pupil of the objective lens and the microlens array (MLA). Figure 12

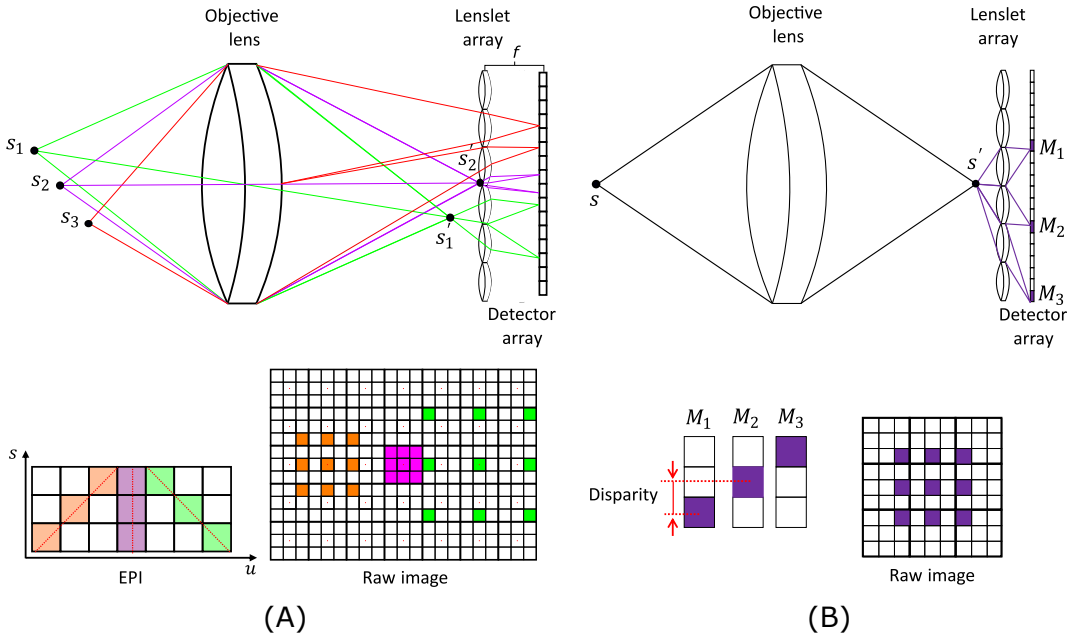


Figura 12. Image formation pipeline in an (A) ULF and (B) FLF camera. Here, the image formation is illustrated just for the depth variable, i.e., the chromatic aberrations were not considered.

(A) shows an example of the image acquisition in a ULF system, where three points, S1 (green), S2 (purple), and S3 (orange), are used as targets. A two-dimensional (2D) s-u slice, referred to as an epipolar plane image, is shown in inset (A) of Figs. 9 and Fig. 12. Here, the EPI is composed of three lines associated with the depths of the objects S1, S2, and S3, respectively. The object's depth can be rendered by measuring the corresponding line slope in the EPI.

- In an FLF optical system, the objective lens forms an intermediate image, S', behind the lenslet array, as shown in Fig. 12(B). Then the MLA acts as a camera array and relays the intermediate image onto the detector array, as shown in Fig. 12(B). The depth information can be derived from the disparities between different subimages or by the rearrangement these into an EPI projection.

2.3. Compressive imaging

Compressive sensing (CS) theory establishes that it is possible to recover signals from a reduced set of measurements. CS-based systems simultaneously allow the sensing process to encode and compress high-dimensional scenes. CS theory is supported by two fundamental principles Emmanuel J Candès y Michael B Wakin. “An introduction to compressive sampling”. En: *IEEE Signal Processing Magazine* 25.2 (2008), págs. 21-30: sparsity and incoherence. The sparsity principle states that the signal of interest must have a sparse representation when expressed in an appropriate representation basis. The incoherence principle extends the duality between time and frequency, and expresses the idea that objects having a sparse representation must be spread out in the domain in which they are acquired.

- **Sparsity**: asserts that the signal of interest $\mathbf{f}$ must be sparse when expressed in an appropriate representation basis $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$. The signal $\mathbf{f}$ is expanded over

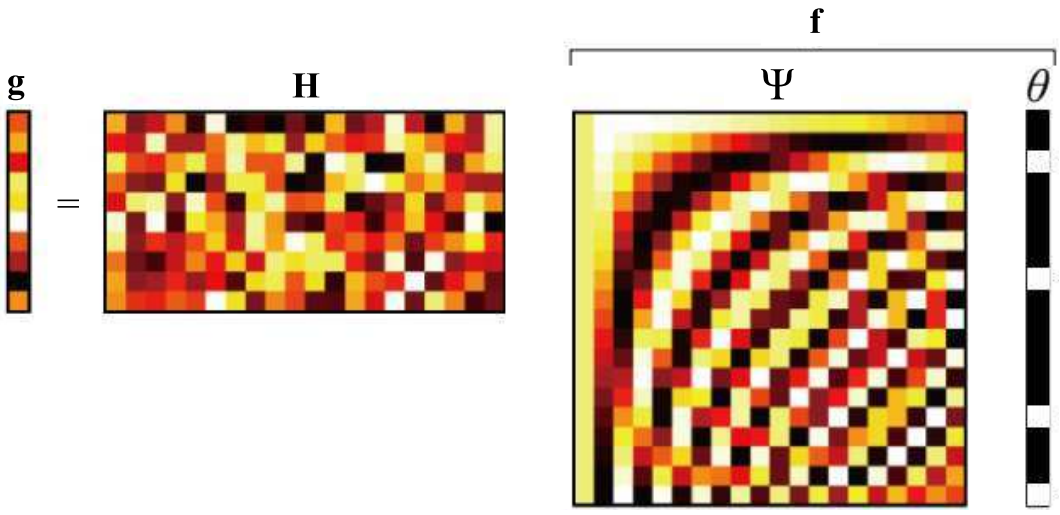


Figura 13. Compressive sensing model illustration for an arbitrary sensing matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ with $m \ll n$, $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$ is the orthogonal representation basis, $\mathbf{g} \in \mathbb{R}^m$ represents the compressed measurement, and $\boldsymbol{\theta} \in \mathbb{R}^n$ as the sparse projection of $\mathbf{f} \in \mathbb{R}^n$ on $\mathbf{\Psi}$.

the basis $\mathbf{\Psi}$ as follows, $\mathbf{f} = \mathbf{\Psi}\boldsymbol{\theta}$, where $\boldsymbol{\theta} \in \mathbb{R}^n$ denotes the sparse representation of $\mathbf{f}$ in the $\mathbf{\Psi}$ representation, as shown in Fig. 13. Sparse signals can be represented with as few as possible significant coefficients, i.e., its negligible coefficients can be discarded.

- **Incoherence:** the incoherence principle places a constraint on the base pair $(\mathbf{H}, \mathbf{\Psi})$ used to acquire/sample in $\mathbf{H}$ and represent in $\mathbf{\Psi}$ the object $\mathbf{f}$. This restriction manages to simplify the required treatment. The restriction consists on obtaining a minimum coherence value $\mu(\mathbf{H}, \mathbf{\Psi})$ between this pair of bases which is calculated as $\mu(\mathbf{H}, \mathbf{\Psi}) = \sqrt{n} \cdot max(|\mathbf{H\Psi}|)$, where $max(\cdot)$ returns the maximum value in its argument and $\mu(\mathbf{H}, \mathbf{\Psi}) \in [1, \sqrt{n}]$. Coherence measures the largest correlation between any pair of elements of $\mathbf{H}$ and $\mathbf{\Psi}$. If this base pair contains correlated elements, its coherence value will be large, otherwise its value will be small, satisfying the CS constraints, since the higher the incoherence value, the smaller the measures needed to capture the signal $\mathbf{f}$.

2.3.1. Compressive spectral and depth imaging Two of the most relevant CS-based frameworks are the compressive spectral imaging (CSI) and compressive depth imaging (CDI). These systems first encode a high-dimensional scene with one or more coded apertures, followed by an additional modulating element for the non-spatial dimension (e.g., prism, grating, galvanometer, lenslet array, deformable mirror). Finally, an encoded-and-modulated scene is integrated by a detector array. The resulting recorded intensities by the detector array are commonly known as compressed measurement. Mathematically, the CS process can be modeled as

$$\mathbf{g} = \mathbf{Hf} + \boldsymbol{\epsilon}, \tag{27}$$

where $\mathbf{f} \in \mathbb{R}^n$ represents the scene, $\mathbf{g} \in \mathbb{R}^m$ represents the measurements with $m \ll n$, $\mathbf{H} \in \mathbb{R}^{m \times n}$ represents the sensing matrix, and $\boldsymbol{\epsilon} \in \mathbb{R}^m$ represents the noise. The statement of the inverse optimization problem for the reconstruction of the signal $\mathbf{f}$ can be written as

$$\tilde{\mathbf{f}} = \boldsymbol{\Psi} \left(\arg\min_{\boldsymbol{\theta}} \|\mathbf{g} - \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}\|_2^2 + \sum_{\iota=1} \tau_\iota W_\iota(\mathbf{A}_\iota \boldsymbol{\theta}) \right) \tag{28}$$

where the ι-th regularization function $W_\iota(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is a closed, proper, and convex function, $\mathbf{A}_\iota \in \mathbb{R}^{m_\iota \times n}$ is an arbitrary matrix with $m_\iota \in \mathbb{R}$, and $\tau_\iota \in \mathbb{R}$ is a regularization variable. One of the most used W_ι regularization functions in these reconstruction problems is: $W_\iota(\cdot) = \|\cdot\|_1$ with $\mathbf{A}_\iota = \mathbf{I} \in \mathbb{R}^{n \times n}$ as the identity matrix (i.e., $m_\iota = n$), where $\|\cdot\|_1$ represents the norm ℓ_1. Another of the most used regularization functions is the total variation Leonid I Rudin, Stanley Osher y Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. En: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), págs. 259-268: $W_\iota(\cdot) = \|\cdot\|_{TV}$ with $\mathbf{A}_\iota = \boldsymbol{\Psi}^T$. In this way, both CSI and CDI state that it is possible to design sampling protocols capable of capturing the essential information from spectral and depth images with the sparseness feature and condensing it into a small amount of data.

Compressive spectral imaging Conventionally, spectral images are acquired by spectrometers that scan adjacent zones of the underlying scene and merge the results into a spectral image Jay S Pearlman y col. “Hyperion, a space-based imaging spectrometer”. En: *IEEE Transactions on Geoscience and Remote Sensing* 41.6 (2003), págs. 1160-1173. However, the main disadvantages of these sensing protocols relate to the large amounts of collected data and the inherent relationship between the number of scanned zones and the resulting spatial and spectral resolution. To alleviate these difficulties of SI, researchers have explored compressive imaging (CI) alternatives Marco F Duarte y col. “Single-pixel imaging via compressive

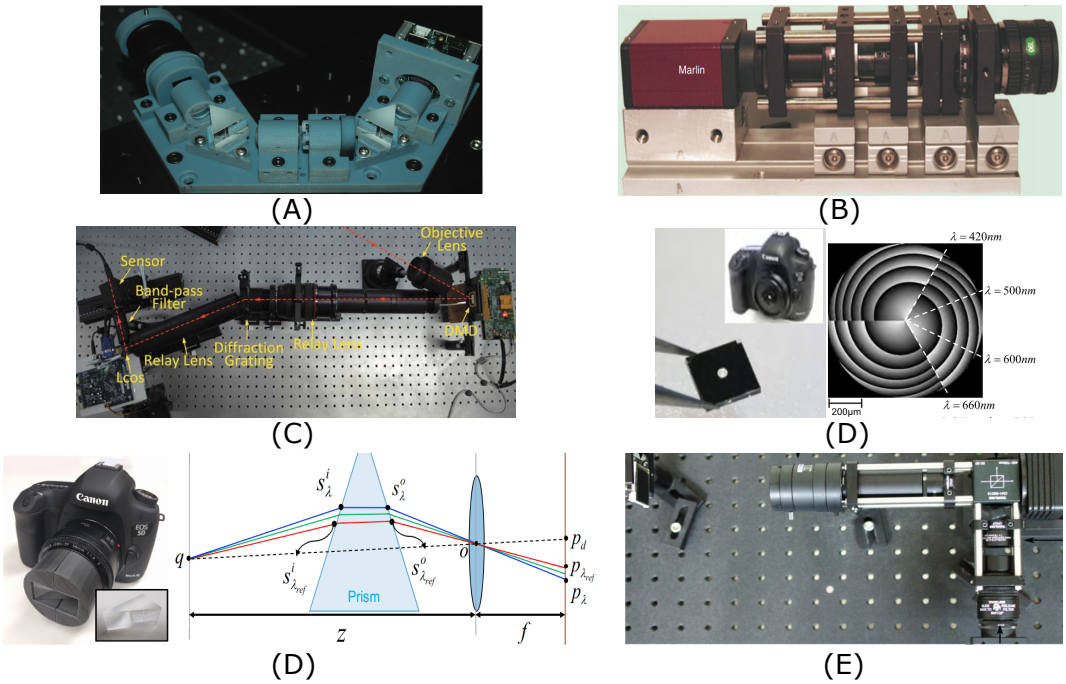


Figura 14. Compressive spectral imagers. (A) DD-CASSI, (B) SD-CASSI, (C) DCSI, (D) CSI with DOE, (E) CSI with prism, and (F) Deformable mirror coding. Adapted from ME Gehm y col. “Single-shot compressive spectral imaging with a dual-disperser architecture”. En: *Optics express* 15.21 (2007), págs. 14013-14027; Ashwin Wagadarikar y col. “Single disperser design for coded aperture snapshot spectral imaging”. En: *Applied optics* 47.10 (2008), B44-B51; Henry Arguello y Gonzalo R Arce. “Colored coded aperture design by concentration of measure in compressive spectral imaging”. En: *IEEE Transactions on Image Processing* 23.4 (2014), págs. 1896-1908; Daniel S Jeon y col. “Compact snapshot hyperspectral imaging with diffracted rotation”. En: *ACM Transactions on Graphics (TOG)* 38.4 (2019), págs. 1-13; Seung-Hwan Baek y col. “Compact single-shot hyperspectral imaging using a prism”. En: *ACM Transactions on Graphics (TOG)* 36.6 (2017), pág. 217; Miguel Marquez y col. “Compressive spectral imaging via deformable mirror and colored-mosaic detector”. En: *Optics Express* 27.13 (2019), págs. 17795-17808.

sampling”. En: *IEEE Signal Processing Magazine* 25.2 (2008), págs. 83-91, coining this area, compressive spectral imaging (CSI) Yitzhak August y col. “Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains”. En: *Applied Optics* 52.10 (2013), págs. D46-D54. CI principles

allow acquiring information about a scene with a reduced amount of coded random projections. Based on the CI theory, researchers have adopted joint sensing and reconstruction paradigms that measure a set of two-dimensional projections of the 3D spectral datacube and obtain an estimation of the underlying image by using computational-based reconstruction algorithms Hoover Rueda, Henry Arguello y Gonzalo R Arce. “DMD-based implementation of patterned optical filter arrays for compressive spectral imaging”. En: *JOSA A* 32.1 (2015), págs. 80-89; Manya V Afonso, José M Bioucas-Dias y Mário AT Figueiredo. “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems”. En: *IEEE Transactions on Image Processing* 20.3 (2010), págs. 681-695. In this direction, several CSI systems have been proposed in the literature by taking advantage of optical elements such as coded apertures (CAs) and dispersive elements (i.e., prisms, gratings) Xun Cao y col. “Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world”. En: *IEEE Signal Processing Magazine* 33.5 (2016), págs. 95-108. In particular, one of the most popular CSI systems is the coded aperture snapshot spectral imager (CASSI) Wagadarikar y col., “Single disperser design for coded aperture snapshot spectral imaging”, which allows sensing spectral images with high spectral resolution at the expense of low spatial resolution due to the shearing introduced by the dispersive element. State-of-the-art CSI systems are listed below, which are also illustrated in Fig. 14:

- **Double prism system** (see Fig. 14 (A) Gehm y col., “Single-shot compressive spectral imaging with a dual-disperser architecture”): DD-CASSI allows preserving spatial features with high accuracy –by performing a dispersive-encoding-dispersive sensing process– at the cost of the spectral resolution. Specifically, the main design features of the system are two dispersive elements arranged in change and surrounding a coded aperture with binary values. Unlike thin-film approaches to spectral filtering, this structure is easily controllable and spatially

variable spectral filter functions with narrow features.

- **Single prism system** (see Fig. 14 (B) Wagadarikar y col., “Single disperser design for coded aperture snapshot spectral imaging”): the SD-CASSI system is composed of a single prism and a coded aperture that works by first encoding a scene with one or multiple coded apertures, followed by a shearing operation that couples the spectral dimension of the spatially encoded scene with one of its spatial dimensions, and finally integration of the encoded and sheared scene on a focal plane array.

- **Dual coded hyperspectral imaging** (see Fig. 14 (C) Henry Arguello y Gonzalo R Arce. “Colored coded aperture design by concentration of measure in compressive spectral imaging”. En: *IEEE Transactions on Image Processing* 23.4 (2014), págs. 1896-1908): DCSI separately codes spatial and spectral dimensions within a single exposure, achieving an independent spectral code for each sensor pixel, i.e., as a color-coding geometry. Specifically, the DCSI comprises two modulation stages: spatial, via a digital micromirror Device (DMD), and spectral (via a diffraction grating coupled with a liquid crystal on silicon (LCOS)).

- **CSI based on diffractive optical elements** (see Fig. 14(D)) Jeon y col., “Compact snapshot hyperspectral imaging with diffracted rotation”: a compact optical system for acquiring hyperspectral images is proposed based on the design of an optical diffractive element attached to a conventional sensor. This method replaces the common optical elements in CSI (prism, coded aperture, lens, diffraction grating) with a single optical element.

- **CSI based on a dispersive element** (see Fig. 14 (E) Seung-Hwan Baek y col. “Compact single-shot hyperspectral imaging using a prism”. En: *ACM Transactions on Graphics (TOG)* 36.6 (2017), pág. 217) the dispersive element-based

imager comprises a simple prism-camera system that allows spectral image recovery from a single exposure. However, this CSI's reconstruction accuracy is mainly related to the deep-learning-based algorithms' inference capabilities.

- **CSI based on deformable mirrors as coding element** (see Fig. 14 (F) Miguel Marquez y col. "Compressive spectral imaging via deformable mirror and colored-mosaic detector". En: *Optics Express* 27.13 (2019), págs. 17795-17808): This is a compact CSI architecture based on the use of deformable mirrors and a color filter detector to produce compressive spatio-spectral projections without the need for a grid or prism. Unlike the Jeon y col., "Compact snapshot hyperspectral imaging with diffracted rotation" system, which designs a DOE based on the deformation that the PSF acquires with the astigmatism-like aberration, the DMC allows multiple captures and exploits the use of the first fifteen Zernike polynomials from the use of a reconstruction algorithm based on tensor notation.

Compressive depth imaging (CDI) Opposite to the CSI framework, the development of CDI systems build upon a wide variety of sensing protocols (detailed in Subsection 2.2). These systems employ a wide variety of optical elements, such as light modulators (LM), deformable mirrors (DM), piezo-electric elements, lenslet arrays, diffractive elements, and tunable focal lens, among others. The more relevant state-of-the-art CDI systems are listed (see Fig. 15) as follows.

- **Masked-lens-based depth systems** (see Fig. 15(A)-(B)) Anat Levin y col. "Image and depth from a conventional camera with a coded aperture". En: *ACM Transactions On Graphics (TOG)* 26.3 (2007), págs. 1-9: Masked-lens systems rely on a simple modification of a conventional lens, by inserting a patterned cardboard disc into the aperture, which permits the recovery of both an all-focus image and depth from a single image. Specifically, the introduced

pattern produces a characteristic distribution of image frequencies that is very sensitive to the exact scale of defocus blur.

- **Structured light single-pixel imager** (see Fig. 15(C)) Ming-Jie Sun y col. “Single-pixel three-dimensional imaging with time-based depth resolution”. En: *Nature Communications* 7 (2016), págs. 1-6. DOI: https://doi.org/10.1038/ncomms12010: The system based on the single-pixel camera relies on short-pulse structured illumination and a high-speed photodiode. The 3D single-pixel imager consists of a pulsed laser and a DMD to provide time-varying structured illumination. A high-speed photodiode is used with a Fresnel lens condenser system to measure each pattern’s resulting backscatter intensity.

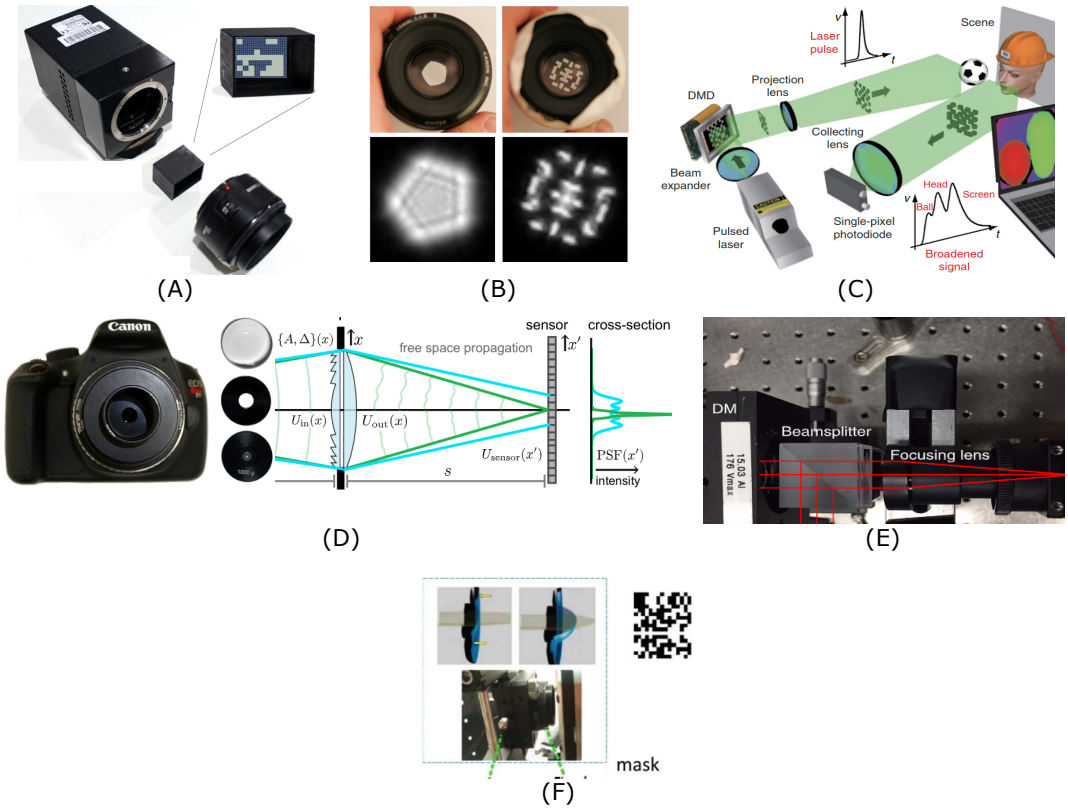


Figura 15. Compressive depth imagers. (A)-(B) Mask-based systems, (C) structured light based on single pixel paradigm, (D) diffractive optical elements-based depth imager, (E) deformable mirror-based system, and (F) tunable focal liquid lens.

- **Diffractive optical element (DOE)-based depth imager** (see Fig. 15(D)) Chang y Wetzstein, “Deep Optics for Monocular Depth Estimation and 3D Object Detection”: This work introduces the design of a diffractive optical element for the monocular depth estimation problem, modeling the PSF response as a trainable neural network parameter. In particular, DOE-based DI systems leverage the point-spread-function properties to customize DOEs with a depth-varying PSF’s anisotropic shape.

- **Deformable mirror-based system** (see Fig. 15(E)) Patrick Llull y col. “Image translation for single-shot focal tomography”. En: *Optica* 2.9 (2015), págs. 822-825: This approach aims to dynamically control the lens’s transverse and longitudinal position as an encoding element, thereby extending the depth of field without degrading the depth resolution. For this purpose, a deformable mirror is used to carry out a focal and transversal variation throughout the exposure time, with which a method of accumulating coded images is obtained, having the PSF as the coding element.

- **Tunable focal liquid lens** (see Fig. 15(F)) Xin Yuan y col. “Efficient patch-based approach for compressive depth imaging”. En: *Applied Optics* 55.27 (2016), págs. 7556-7564: This system presents an optical architecture that moves its focal plane through a liquid lens and modulates the scene in different focal planes using a DMD. Note that each focal plane is spatially modulated using a different coded aperture. Then, the sensor integrates the resulting spatial modulated stack of projections into a single measurement.

3. CASSI-VSI: Coded Aperture Snapshot Spectral Imaging via Virtual Side Information

Several CSI systems have been proposed in the literature by taking advantage of optical elements such as coded apertures (CAs) and dispersive elements (i.e. prisms, gratings) Cao y col., “Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world”. In particular, one of the most popular CSI systems is the coded aperture snapshot spectral imager (CASSI) Wagadarikar y col., “Single disperser design for coded aperture snapshot spectral imaging”, which allows sensing spectral images with high spectral resolution at the expense of low spatial resolution, due to the shearing introduced by the dispersive element. To overcome the spatial resolution limitation of CASSI, newer CSI schemes have incorporated prior information about the scene to further reduce the required number of measurements and to help the reconstruction algorithm to attain better reconstructions Zhiwei Xiong y col. “Snapshot hyperspectral light field imaging”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 3270-3278; Laura Galvis y col. “Coded aperture design in compressive spectral imaging based on side information”. En: *Applied Optics* 56.22 (2017), págs. 6332-6340; Xin Yuan y col. “Compressive hyperspectral imaging with side information”. En: *IEEE Journal of Selected Topics in Signal Processing* 9.6 (2015), págs. 964-976; Hoover Rueda y col. “Single aperture spectral+ ToF compressive camera: toward hyperspectral+ depth imagery”. En: *IEEE Journal of Selected Topics in Signal Processing* 11.7 (2017), págs. 992-1003. One of the most popular approaches relies on co-located imaging systems that observe the same scene, but each system is set to acquire a specific projection of the scene. The latter has been referred to as side-information Yuan y col., “Compressive hyperspectral imaging with side information”. Examples of these systems include the CASSI system co-located with a grayscale or RGB color sensor,

where the spatial information of the co-located sensors are exploited to improve the poor spatial features recovered from the CASSI measurement.

The positive impact of side-information in CSI algorithms and the fact that the sensing geometry of CSI systems can be designed to accommodate fewer and cheaper optical elements, lead naturally to the question: ***Is it possible to develop a CSI system that exploits the advantages of side information without relying on the use of additional co-located sensors?***

To answer this question, this chapter introduces a compressive snapshot spectral imaging architecture that builds on the ideas of side information but relies on a single image sensor. Specifically, we propose to split the incoming light into two paths with one path dispersing the coded light, while the other just relaying it, so as to multiplex a traditional coded-and-dispersed measurement with a modulated grayscale version of the 3D cube in a single image sensor. We have coined the latter, coded aperture snapshot spectral imaging virtual side information – ***CASSI-VSI****.*

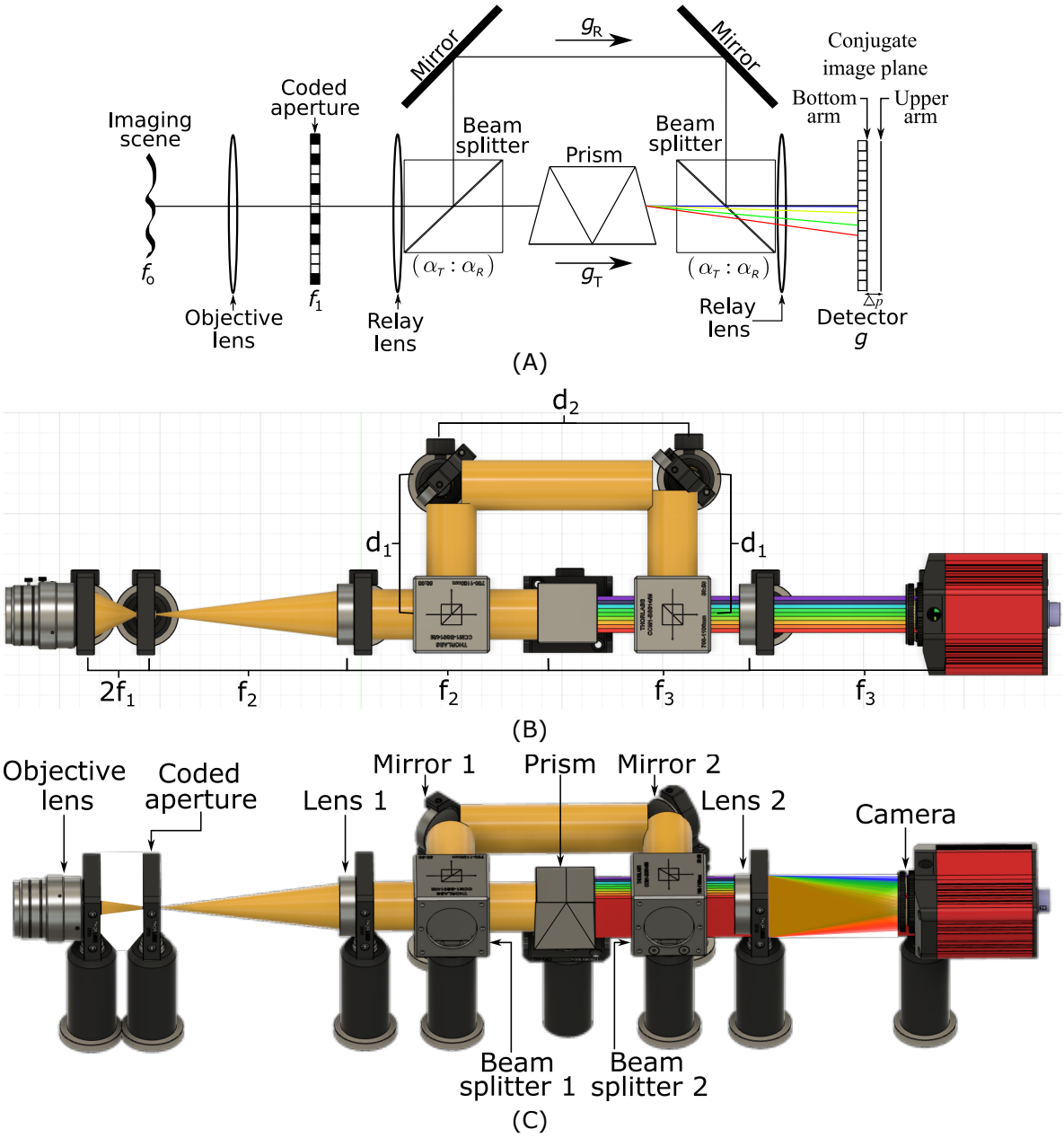


Figura 16. (A) Sketch of the proposed CASSI-VSI architecture. (B)-(C) Upper and lateral view of the designed system where the distance configuration and the optical element labels are specified, respectively. (D) Examples of simulated CASSI-VSI compressed measurements for $(\alpha_T : \alpha_R) = (0 : 1)$ (left image), $(\alpha_T : \alpha_R) = (0.5 : 0.5)$ (center image), and $(\alpha_T : \alpha_R) = (1 : 0)$ (right image).

3.1. Continuous and discrete sensing model

Recent advances in optical instrumentation have largely expanded the design space of snapshot compressive imaging systems, which results in many innovative techniques that open up new applications in diverse computer vision areas of study, including remote sensing, autonomous driving, security, and precision agriculture. In this regard, the techniques based on side information configurations have become increasingly popular because of their high reconstruction performance. The performance of the reconstruction algorithms based on side information directly depends on the quality of such information. To guarantee high quality, these approaches rely on a secondary high-resolution co-registered sensor, referred as to dual-arm imagers. These systems attain high spatially resolved spectral images, but their implementation relies on SLM, demanding high implementation costs and high calibration complexity. In general, side information entails optical systems with dual sensors, which trade-offs reconstruction quality, compression ratio, implementation complexity, and costs. Further, dual-camera systems become prohibitive in the infrared spectrum due to costs or physical constraints, e.g., a Bayer sensor in the infrared range. An alternative to overcome the dual-sensor requirement is to develop a reconstruction methodology that estimates the side-information directly from the compressed measurement. Therefore, the dual-camera systems become unfeasible to be applied directly in state-of-the-art CSI systems such as the CASSI system since its compressed measurement does not preserve the scene's spatial features with high quality.

To alleviate the implementation costs, implementation complexity, and compression ratios, the proposed optical system is capable of acquiring a single coded-and-dispersed measurement that preserves the spatial and spectral features of the scene. Specifically, we propose to split the incoming light into two paths, with one dispersing it, while the other just relaying it, to then multiplex them in a single image sensor.

The latter methodology is here referred to as virtual side-information. Further, we propose an alternating direction method of multipliers (ADMM)-based reconstruction methodology, which exploits the designed compressed measurement nature to simultaneously estimate the grayscale and the spectral image. This approach uses the iterative estimation of the grayscale image as prior information into the spectral image estimation subproblem. Note that since the side information is estimated from the compressed measurements, the proposed methodology has been coined, coded aperture snapshot spectral imaging via virtual side-information (CASSI-VSI).

CASSI-VSI is composed of an objective lens, a pair of relay lenses, a coded aperture (CA), a pair of beam-splitters, a pair of mirrors, a prism and a grayscale sensor, accommodated as detailed in Fig. 16(A). This architecture works as follows: first, the scene is focused by the objective lens onto the CA, where it is spatially modulated; second, a relay lens propagates the modulated spectral image through the first beam-splitter, which splits the incoming wave-front into two paths, one path propagates the wave-front through a dispersing element (prism) and to the second beam-splitter, while the other just relays it to the second beam-splitter by using two mirrors. Finally, the two incoming wave-fronts are combined by the second beam-splitter and focused onto the sensor by a relay lens. Here it is remarkable to highlight that CASSI-VSI is an optical system envisioned to be used in computer vision applications in outdoor environments, i.e., for incoherent illumination. Additionally, although the CASSI-VSI system shows certain aparent similarities with interferometry systems (i.e., with coherent illumination), its operation is neither related nor inspired by these optical systems.

3.1.1. Continuous Sensing Model The proposed testbed in Fig. CASSI-VSI has two propagation paths, separated by the beam splitter, but a single objective lens and sensor are common to both. The first step is to define the CASSI-VSI's propagation model in a continuous notation to develop the reconstruction methodology. Particu-

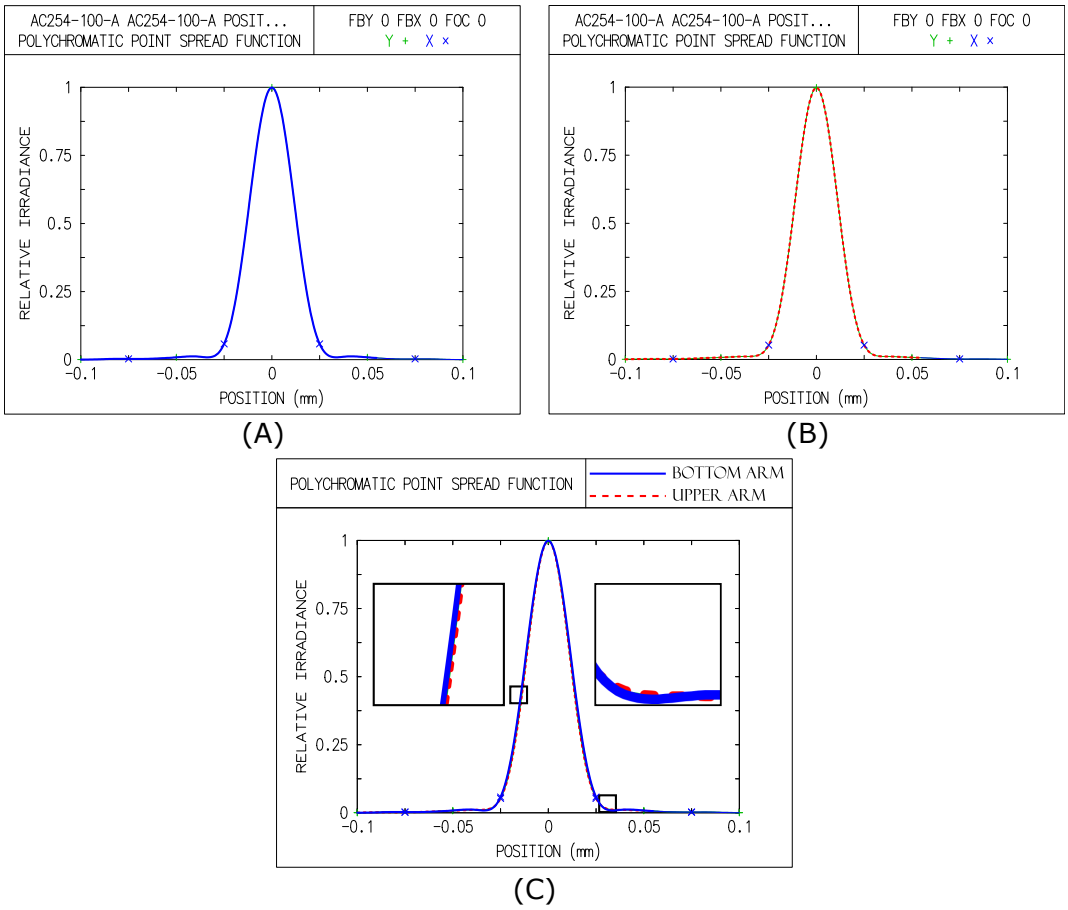


Figura 17. Intensity PSF analysis for the (A) bottom arm, (B) upper arm, and (C) comparison between (A) and (B). The CASSI-VSI system was emulated using two achromatic doublet lenses (Thorlabs, AC254-100-A-ML - f=100 mm) with $f_2 = f_3 = 100$ mm, $\Delta p \approx 10\mu m$, and $d_1 = d_2 = 50$ mm (see Fig. 16).

larly, let $f_o(x, y, \lambda)$ be the spatio-spectral scene in the intensity representation, where x and y index the spatial coordinates and λ indexes the wavelength. The objective lens forms an image in the plane where the coded aperture is located, such that it modulates the spectral density over all wavelengths with the code $k(x, y)$, resulting in the coded field

$$f_1(u, v, \lambda) = k(u, v) \iint \gamma(\lambda) f_o(x, y, \lambda) h_1(x - u, y - v, \lambda) dx dy \tag{29}$$

with

$$k(u,v) = \frac{1}{2}\sum_{j_1,j_2} C_{j_1,j_2}\text{rect}\left(\frac{u}{\Delta_c} - j_1, \frac{v}{\Delta_c} - j_2\right), \tag{30}$$

where $C_{j_1,j_2} \in \{0,1\}$ is the binary coding applied to the $(i_1,i_2)^{th}$ spatial location with a 1 representing a transmissive code element a 0 representing an opaque code element; $h_1(\cdot)$ represents a shift-invariant optical impulse response of the relay optics; $\gamma(\lambda)$ represent the spectral source density; $i_1 = \{0,\ldots,N_x-1\}$, and $i_2 = 0,\ldots,N_y-1$ indexing the rows and columns, respectively, and Δ_c denoting the pixel side size. The coded density then passes through a beam-splitter that splits f_1 into two paths according to the intensity ratio (reflectivity/transmissivity) of the beam splitter α_T/α_R. Particularly, the intensity splitting ratios are defined as $\alpha_T = |\overrightarrow{E}_t/\overrightarrow{E}_i|^2$ and $\alpha_R = |\overrightarrow{E}_r/\overrightarrow{E}_i|^2$, where $\overrightarrow{E}_i$ represents the incident electric field amplitude, $\overrightarrow{E}_t$ represents the transmitted electric field amplitude, and $\overrightarrow{E}_r$ represents the reflected electric field amplitude. In the transmission branch, the coded density is propagated through a dispersive element, which spectrally shears it following a wavelength-dependent function $S(\lambda)$, (refer to supplementary material Eq. A.20). The spectral density immediately before the detector plane (resulting from the bottom arm) can be expressed as

$$f_2(x',y',\lambda) = \iint \alpha_T f_1(u,v,\lambda) h_T(u-x'-S(\lambda), v-y')dudv, \tag{31}$$

where $h_T(u-S(\lambda)-x', v-y', \lambda)$ represents the intensity impulse response of the transmission optical system for the CASSI arm. Finally, the detector array measures the intensity of the incident light rather than the spectral density by integrating the datacube along the wavelength axis over the spectral range Λ. Then, the measurement at the $(i_1, i_2')^{th}$ pixel is represented by

$$(g_T)_{i_1,i_2'} = \int_\Lambda \rho(\lambda) \iint f_2(x',y',\lambda) p(i_1,i_2';x',y')dx'dy'd\lambda, \tag{32}$$

with

$$p(i_1, i_2'; x', y') = \text{rect}\left(\frac{x'}{\Delta_d} - i_1, \frac{y'}{\Delta_d} - i_2'\right), \tag{33}$$

where Δ_d represents the pixel's side length, $p(\cdot)$ is the pixel sampling function, and $\rho(\lambda)$ represents the detector's normalized quantum efficiency. Meanwhile, in the second branch (reflection), light travels unmodified towards the detector plane, which results in the power spectral density (immediately before converging the detector)

$$f_3(x', y', \lambda) = \iint \alpha_R f_1(u, v, \lambda) h_R(x' - u, y' - v) du dv, \tag{34}$$

where $h_R(x - x', y - y', \lambda)$ represents the impulse response of the reflection optical system. Similarly to Eq. (32), the measurement at the $(i_1, i_2')^{th}$ pixel for the upper arm can be represented as

$$(g_R)_{i_1, i_2'} = \int_\Lambda \rho(\lambda) \iint f_3(x', y', \lambda) p(i_1, i_2'; x', y') dx' dy' d\lambda. \tag{35}$$

Here, it is worth to note that light from the two paths meet at the same detector. Then, based on Eqs. (32) and (35), the multiplexing process in the sensor can be mathematically represented as

$$\begin{aligned} g_{i_1, i_2'} &= \int_\Lambda \rho(\lambda) \iint \left[f_3(x', y', \lambda) + f_2(x', y', \lambda)\right] p(i_1, i_2'; x', y') dx' dy' d\lambda \\ &= (g_T)_{i_1, i_2'} + (g_R)_{i_1, i_2'} \end{aligned} \tag{36}$$

where **g** represents the compressed measurement acquired by the CASSI-VSI system.

3.1.2. Discrete Sensing Model In order to establish a linear system to describe the acquisition of the compressive measurement, **F** is defined as a

$$F_{i_1,i_2,i_3} = \int_{\Lambda}\iint\iint\left[\iint f_o(x,y,\lambda)h_1(x-u,y-v,\lambda)dxdy\right]h_T(u-x',v-y') \times \text{rect}\left(\frac{x'}{\Delta_d}-i_1,\frac{y'}{\Delta_d}-i_2,\frac{S(\lambda)+|S(\lambda_{min})|}{\Delta_d}-i_3\right)dudvdx'dy'd\lambda, \quad (37)$$

where $\mathbf{F}\in\mathbb{R}^{N_x\times N_y\times N_\lambda}$ represents the discrete spectral image cube and $i_3=\{0,...,N_\lambda-1\}$. Based on Eq. (37), a grayscale version (panchromatic image) of the source can be expressed as

$$Z_{i_1,i_2} = \int_{\Lambda}\iint\iint\left[\iint f_o(x,y,\lambda)h_1(x-u,y-v,\lambda)dxdy\right]h_T(u-x',v-y') \times \text{rect}\left(\frac{x'}{\Delta_d}-i_1,\frac{y'}{\Delta_d}-i_2\right)dudvdx'dy'd\lambda, \quad (38)$$

where h_T is the matrix version of the intensity PSF of the bottom arm (without the prism dispersion effect) and $\mathbf{Z}\in\mathbb{R}^{N_x\times N_y}$. Notice that the path lengths for the bottom and the upper arms are different, mathematically implying a mismatch in the conjugate image planes, i.e., one of the two projections is off-focus. For this work, the CASSI-VSI system is aligned so that the bottom arm's conjugate image plane and the sensor plane match. Therefore, the grayscale version generated by the upper arm can be expressed as

$$\tilde{Z}_{i_1,i_2} = \int_{\Lambda}\iint\iint\left[\iint f_o(x,y,\lambda)h_1(x-u,y-v,\lambda)dxdy\right]h_R(u-x',v-y') \times \text{rect}\left(\frac{x'}{\Delta_d}-i_1,\frac{y'}{\Delta_d}-i_2\right)dudvdx'dy'd\lambda, \quad (39)$$

where $\mathbf{h}_R$ is the matrix version of the intensity PSF of the upper arm and $\tilde{\mathbf{Z}}\in\mathbb{R}^{N_x\times N_y}$ represents the grayscale version generated by the upper arm. In this regard, the CASSI-VSI reconstruction performance is limited by the quality of $\tilde{\mathbf{Z}}$ concerning **Z**,

i.e., $\|\tilde{\mathbf{Z}} - \mathbf{Z}\| < Error$. Therefore, in the current CASSI-VSI configuration (see Fig. 16), its reconstruction performance is guaranteed for cases where the diameter of the intensity PSF of the upper arm is smaller than the camera pixel width. To support this claim, Fig. 17 illustrated an intensity PSF analysis developed in the professional Optics Software for Layout and Optimization (OSLO) in which the CASSI-VSI system was emulated using two achromatic doublet lenses (Thorlabs, AC254-100-A-ML - f=100 mm) with $f_2 = f_3 = 100$ mm, $d_1 = d_2 = 50$ mm, and $\Delta p \approx 10\mu m$ (see Fig. 16), i.e., a path length difference of 100 mm. Figure 17 shows that the intensity PSF's diameters for the upper and bottom arm are smaller than 25. Finally, a comprehensive evaluation study was conducted by calculating the normalized mean square error (NMSE) between the grayscale image version generated by the upper and bottom arms, i.e., $\|\mathbf{Z} - \tilde{\mathbf{Z}}\|^2 / \|\mathbf{Z}\|^2$. For this analysis, 4500 images were used where the average NMSE was 0.0041 with a standard deviation of $9.7362e^{-4}$. Additionally, the metrics peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) metrics were also calculated, obtaining the values $55.10 \pm 3.06[dB]$ and 0.9983 ± 0.0021, respectively. Thus, if and only if the degradation generated by the mismatch between the conjugate image planes is minimal, the following assumption holds $\mathbf{Z} \approx \tilde{\mathbf{Z}}$ with $\mathbf{Z} = \sum_{i_3=0}^{N_\lambda - 1} \mathbf{F}_{:,:,i_3}$. Finally, in a subsequent analysis of the development of this work it was found that an alternative to circumvent the path length mismatch is to construct the CASSI-VSI by mimicking the Machâ€“Zehnder interferometer elements distribution (Disclaimer: here, we do refer uniquely to the location of the optical elements, and not to the physical phenomenon associated).

Then, following Eqs. (32)-(38), assuming that the intensity PSF h_T-h_R are ideal, and $\mathbf{Z} \approx \tilde{\mathbf{Z}}$, the discrete version of the measurement can be expressed as,

$$G_{i_1,i_2'} = \sum_{i_3=0}^{N_\lambda - 1} [\alpha_T C_{i_1,(i_2'-i_3)} F_{i_1,(i_2'-i_3),i_3} + \alpha_R C_{i_1,i_2'-\beta} Z_{i_1,i_2'-\beta}] + \epsilon_{i_1,i_2'}, \tag{40}$$

for $i_2' = \{0, ..., (N_y + N_\lambda - 1) - 1\}$, where $G_{i_1,i_2'}$ is the 2D sensor intensity map, $\beta \in \{0, ..., N_\lambda - 1\}$ is a fixed value that indicates the detector column where the first column of the modulated grayscale projection impinges, and $\epsilon_{i_1,i_2'}$ represents the noise. In computer vision for the design of recovery algorithms, a widespread approximation is to neglect the lens aberrations by assuming their intensity PSF as ideal, i.e., a delta Dirac. Although this assumption is physically unrealistic, this simple model suffices for several computer vision and commercial applications. Notice that, modeling the noise variable $\boldsymbol{\epsilon}$ as a colored Gaussian noise distribution has been widely accepted in high-level computer vision tasks. Though this simple model is physically unrealistic, it suffices for some computer vision tasks, e.g., reconstruction, recognition, classification, detection, and understanding. Because the scope of this thesis is limited to reconstruction tasks and based on the plethora of state-of-the-art computational optics reconstruction algorithms, in this thesis, the noise is approximated by a colored noise Gaussian distribution.

In Eq. (40) and below, it is assumed that $C_{i_1,i_2'-\beta} = Z_{i_1,i_2'-\beta} = 0$ when $(i_2' - \beta) < 0$. Note that the output of the system, $\mathbf{G}$ in Eq. (40), is an $N_x \times (N_y + N_\lambda - 1)$ grayscale image, that yields a compression ratio of $\frac{(N_y+N_\lambda-1)}{N_y N_\lambda}$. Now, note that the proposed optical system can be represented as a linear system of the form,

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}, \tag{41}$$

where $\mathbf{f} \in \mathbb{R}^n$ is the vector representation of the spectral image, with $n = N_x N_y N_\lambda$, and $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the proposed sensing matrix, with $m = N_x(N_y + N_\lambda - 1)$. The structure of $\mathbf{H}$ depends on the accommodation of the optical elements inside the

CASSI-VSI system. In particular, its entries are given by

$$H_{i,j} = \begin{cases} \alpha_T c_u, & \text{if } i = u + N_x \lfloor \frac{j}{N_x N_y} \rfloor,\ \beta \neq a \\ \alpha_R c_u, & \text{if } i = u + \beta N_x\ ,\ i < v,\ \beta \neq a \\ c_u & \text{if } i = u + \beta N_x\ ,\ i < v,\ \beta = a \\ 0, & \text{otherwise} \end{cases}, \tag{42}$$

for $i = 0, ..., m-1$, $j = 0, ..., n-1$, $u = mod(j, N_x N_y)$, $v = N_x(\beta + N_y)$, $a = \lfloor \frac{j}{N_x N_y} \rfloor$, where c_u are the entries of $\mathbf{c} \in \mathbb{R}^{N_x N_y}$, the vectorized version of the CA. Further, the matrix $\mathbf{H}$ can be decoupled as $\mathbf{H} = \mathbf{H}_T + \mathbf{H}_R$ with

$$(H_T)_{i,j} = \begin{cases} \alpha_T c_u, & \text{if } i = u + N_x \lfloor \frac{j}{N_x N_y} \rfloor \\ 0, & \text{otherwise} \end{cases}, \tag{43}$$

$$(H_R)_{i,j} = \begin{cases} \alpha_R c_u, & \text{if } i = u + \beta N_x\ ,\ j < v \\ 0, & \text{otherwise} \end{cases}, \tag{44}$$

where $\mathbf{H}_T \in \mathbb{R}^{m \times n}$ is a traditional CASSI matrix (see Fig. 18(A)), and $\mathbf{H}_R \in \mathbb{R}^{m \times n}$ is a multiplexing matrix with a binary modulation (see Fig. 18(B)), given by $\mathbf{H}_R = [\mathbf{B}^T, \mathbf{0}^T_{N_x(N_\lambda - 1) \times n}]^T$ with $\mathbf{B} = [\mathbf{1}^T_{N_\lambda} \otimes \text{diag}(\mathbf{c})]$. Here it is worth noting that $\mathbf{H} = \mathbf{H}_T$ when $\alpha_T = 1$ and $\mathbf{H} = \mathbf{H}_R$ when $\alpha_T = 0$.

3.2. Reconstruction algorithm

Conventional reconstruction algorithms based on side information use a panchromatic or RGB representation of the scene as data prior to improve their perfor-

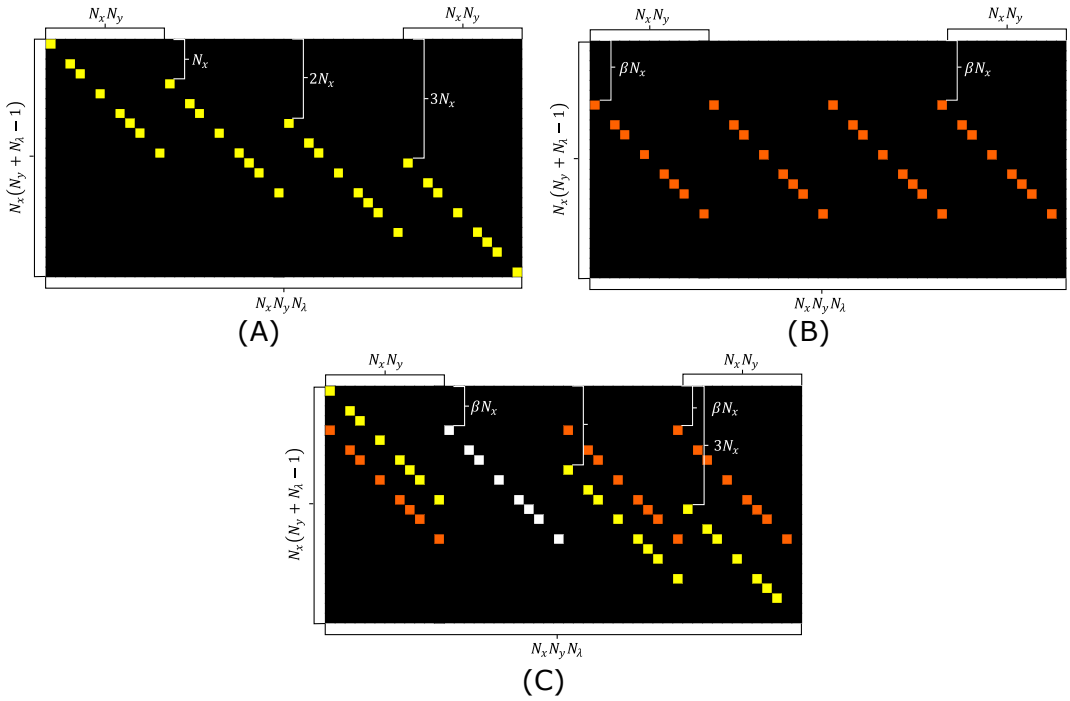


Figura 18. Illustrative example of matrices (A) $\mathbf{H}_T$, (B) $\mathbf{H}_R$ and (C) $\mathbf{H} = \mathbf{H}_T + \mathbf{H}_R$. The yellow, orange and white squares represent the locations of the α_T, α_R and $\alpha_T + \alpha_R = 1$.

mance. Building on this concept and the distributed compressive sensing (DCS) Hamid Palangi, Rabab Ward y Li Deng. “Distributed Compressive Sensing: A Deep Learning Approach”. En: *IEEE Transactions on Signal Processing* 64.17 (2016), págs. 4504-4518. DOI: 10.1109/TSP.2016.2557301, to propose a virtual side-information reconstruction methodology that alternatively estimates the spectral data cube and its panchromatic version simultaneously. This approach exploits both intra- and inter- signal correlations to obtain improvements in the panchromatic and spectral reconstructions.

Recall that $\mathbf{H}$ can be decoupled as the sum of the CSI matrix $\mathbf{H}_T$ and the panchromatic modulated sensing matrix $\mathbf{H}_R$, which allows rewriting the forward model in Eq. (41) as

$$\mathbf{g} = [\alpha_T \mathbf{H}_T + \alpha_R \mathbf{H}_R]\, \mathbf{f} + \boldsymbol{\epsilon}. \tag{45}$$

In particular, this decoupling can be expressed mathematically as $\mathbf{g} = \mathbf{g}_T + \mathbf{g}_R$, where $\mathbf{g}_T = \alpha_T \mathbf{H}_T \mathbf{f} + \boldsymbol{\epsilon}_T$ with $\mathbf{g}_T \in \mathbb{R}^m$, $\mathbf{g}_R = \alpha_R \mathbf{H}_R \mathbf{f} + \boldsymbol{\epsilon}_R$ with $\mathbf{g}_R \in \mathbb{R}^m$. Here it is worth noting that physically the detector doesn't split the noise (i.e., it is unique), in fact, splitting the noise is a mathematical assumption done to formulate a hierarchical optimization problem based on the ADMM approach. Based on Eqs. (43)-(44), the alternating virtual side-information reconstruction problem is given by

$$\begin{aligned} \underset{\mathbf{f},\mathbf{z},\boldsymbol{\theta}_f,\boldsymbol{\theta}_z,\mathbf{g}_T,\mathbf{g}_R}{\arg\min} \quad & \tfrac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \tfrac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{P}\mathbf{z}\|_2^2 + \tfrac{1}{2}\|\mathbf{g} - (\mathbf{g}_R + \alpha_T \mathbf{H}_T \mathbf{f})\|_2^2 \\ & + \tfrac{1}{2}\|\mathbf{g} - (\mathbf{g}_T + \alpha_R \mathbf{H}_R \mathbf{f})\|_2^2 + \tau_f \|\boldsymbol{\theta}_f\|_1 + \tau_z \|\boldsymbol{\theta}_z\|_1 \\ \text{subject to} \quad & \boldsymbol{\theta}_f = \boldsymbol{\Psi}_f \mathbf{f};\ \boldsymbol{\theta}_z = \boldsymbol{\Psi}_z \mathbf{z};\ \mathbf{z} = \mathbf{D}\mathbf{f}, \end{aligned} \tag{46}$$

where $\mathbf{D} \in \mathbb{R}^{\ell \times n}$ represents a decimation matrix defined as $\mathbf{D} = \mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{\ell \times \ell}$, $\mathbf{z} \in \mathbb{R}^\ell$ is the grayscale projection of $\mathbf{f}$ with $\mathbf{z} = \mathbf{D}\mathbf{f}$ and $\ell = N_x N_y$, $\boldsymbol{\Psi}_z \in \mathbb{R}^{\ell \times \ell}$ is an orthonormal representation basis with $\boldsymbol{\theta}_z = \boldsymbol{\Psi}_z \mathbf{z}$, $\boldsymbol{\theta}_f \in \mathbb{R}^n$ is a sparse representation of $\mathbf{f}$ in the basis $\boldsymbol{\Psi}_f \in \mathbb{R}^{n \times n}$, with $\boldsymbol{\Psi}_f^T \boldsymbol{\Psi}_f = \boldsymbol{\Psi}_f \boldsymbol{\Psi}_f^T = \mathbf{I}_{n \times n}$, and $\{\tau_f, \tau_z\} \in \mathbb{R}_+$ are regularization parameters. In Eq. (46), we use the $\ell_2 - \ell_1$ optimization functions to promote sparse reconstructions that fulfill the CS condition for ill-posed sensing matrices. Moreover, to limit the solution space and exploit the inter-image correlation, we establish two ℓ_2 fidelity functions based on the decoupling of the matrix $\mathbf{H}$. These ℓ_2 functions describe a direct correlation between $\mathbf{f}$ and the projections $\mathbf{g}$, $\mathbf{g}_T$, and $\mathbf{g}_R$. Finally, the equality constraint $\mathbf{z} = \mathbf{D}\mathbf{f}$ is included as a direct relationship between $\mathbf{f}$ and $\mathbf{z}$.

To solve Eq. (46) we follow and ADMM methodology. For this, we calculate the augmented Lagrangian as

$$\begin{aligned} \mathcal{L}(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_z, \boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \boldsymbol{\chi}_3, \mathbf{g}_T, \mathbf{g}_R) = \quad & \gamma_1(\mathbf{f}, \mathbf{g}_T, \mathbf{g}_R) + \tfrac{\sigma_3}{2}\|\mathbf{z} - \mathbf{D}\mathbf{f} - \boldsymbol{\chi}_3\|_2^2 \\ & + \gamma_3(\boldsymbol{\theta}_f) + \gamma_4(\boldsymbol{\theta}_z) + \tfrac{\sigma_1}{2}\|\boldsymbol{\theta}_f - \boldsymbol{\Psi}_f \mathbf{f} - \boldsymbol{\chi}_1\|_2^2 \\ & + \tfrac{\sigma_2}{2}\|\boldsymbol{\theta}_z - \boldsymbol{\Psi}_z \mathbf{z} - \boldsymbol{\chi}_2\|_2^2 + \gamma_2(\mathbf{z}), \end{aligned} \tag{47}$$

where $\gamma_1(\mathbf{f}, \mathbf{g}_T, \mathbf{g}_R) = \frac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \frac{1}{2}\|\mathbf{g} - (\mathbf{g}_R + \alpha_T \mathbf{H}_T \mathbf{f})\|_2^2 + \frac{1}{2}\|\mathbf{g} - (\mathbf{g}_T + \alpha_R \mathbf{H}_R \mathbf{f})\|_2^2$, $\gamma_2(\mathbf{z}) = \frac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{P}\mathbf{z}\|_2^2$, $\gamma_3(\boldsymbol{\theta}_f) = \tau_f \|\boldsymbol{\theta}_f\|_1$, and $\gamma_4(\boldsymbol{\theta}_z) = \tau_z \|\boldsymbol{\theta}_z\|_1$. Based on Eq. (47), the alternative form of Eq. (46) can be expressed as

$$\begin{aligned} \underset{\mathbf{f},\mathbf{z},\boldsymbol{\theta}_f,\boldsymbol{\theta}_z,\mathbf{g}_R,\mathbf{g}_T}{\arg\min} \quad & \tfrac{\sigma_1}{2}\|\boldsymbol{\theta}_f - \boldsymbol{\Psi}_f \mathbf{f} - \boldsymbol{\chi}_1^{\iota}\|_2^2 + \gamma_1(\mathbf{f}, \mathbf{g}_T, \mathbf{g}_R) + \gamma_2(\mathbf{z}) + \gamma_3(\boldsymbol{\theta}_f) \\ & + \gamma_4(\boldsymbol{\theta}_z) + \tfrac{\sigma_2}{2}\|\boldsymbol{\theta}_z - \boldsymbol{\Psi}_z \mathbf{z} - \boldsymbol{\chi}_2^{\iota}\|_2^2 + \tfrac{\sigma_3}{2}\|\mathbf{z} - \mathbf{D}\mathbf{f} - \boldsymbol{\chi}_3^{\iota}\|_2^2, \end{aligned} \tag{48}$$

where $\boldsymbol{\chi}_1^{\iota+1} = \boldsymbol{\chi}_1^{\iota} - (\boldsymbol{\theta}_f - \boldsymbol{\Psi}_f \mathbf{f})$, $\boldsymbol{\chi}_2^{\iota+1} = \boldsymbol{\chi}_2^{\iota} - (\boldsymbol{\theta}_z - \boldsymbol{\Psi}_z \mathbf{z})$, and $\boldsymbol{\chi}_3^{\iota+1} = \boldsymbol{\chi}_3^{\iota} - (\mathbf{z} - \mathbf{D}\mathbf{f})$. Regarding the convergence of the alternated virtual side information (A-VSI) algorithm, the optimization problem established in Eq. (48) is composed of three proper convex and closed functions, $\gamma_1(\cdot)$, $\gamma_2(\cdot)$ and $\|\cdot\|_2^2$. Therefore, if $\{\sigma_1, \sigma_2, \sigma_3\} > 0$, we can invoke the Theorem 8 in Jonathan Eckstein y Dimitri P. Bertsekas. “On the Douglasâ€”Rachford splitting method and the proximal point algorithm for maximal monotone operators.” En: *Mathematical Programming* 55.1-3 (1992), págs. 293-318, by Eckstein and Bertsekas, which, guarantees the convergence of **Algorithm 1**. Here, to estimate $\boldsymbol{\theta}_z$ and $\boldsymbol{\theta}_f$, the closed-form solutions are obtained via soft-thresholding–referred to as sof$(\cdot)$ in **Algorithm 1**.

3.3. Simulation results

The performance of the proposed optical systems and reconstruction methodologies is first evaluated via simulated compressed measurements, using the model described in Eqs. (41), and reconstructing them using the A-VSI algorithm. For these simulations, fifty spectral images from four public datasets were used: 15 from HINS David H Foster y col. “Frequency of metamerism in natural scenes”. En: *JOSA A* 23.10 (2006), págs. 2359-2372, 7 from CAVE Fumihito Yasuma y col. “Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum”. En: *IEEE Transactions on Image Processing* 19.9 (2010), págs. 2241-2253,

Algorithm 1 A-VSI algorithm

Input: $\mathbf{g} \in \mathbb{R}^m$, $\{\mathbf{H}, \mathbf{H}_T, \mathbf{H}_R\} \in \mathbb{R}^{m \times n}$, $\mathbf{D} \in \mathbb{R}^{\ell \times n}$, $\mathbf{\Psi}_f \in \mathbb{R}^{n \times n}$, $\mathbf{\Psi}_z \in \mathbb{R}^{\ell \times \ell}$, $\{\alpha_T, \alpha_R\} \in [0, 1]$, $Iter$

Result: $\mathbf{f}$, $\boldsymbol{\theta}_f$

1. $\{\mathbf{g}_T, \mathbf{g}_R\} \leftarrow \mathbf{0} \in \mathbb{R}^m$, $\{\boldsymbol{\chi}_1, \boldsymbol{\theta}\} \leftarrow \mathbf{0} \in \mathbb{R}^n$, $\{\mathbf{z}, \boldsymbol{\chi}_2, \boldsymbol{\chi}_3\} \leftarrow \mathbf{0} \in \mathbb{R}^\ell$

for $\iota \leftarrow 1$, *Iter* **do**

2. $\boldsymbol{\eta}_1 = \mathbf{H}^T\mathbf{g} + \alpha_T\mathbf{H}_T^T[\mathbf{g} - \mathbf{g}_R] + \alpha_R\mathbf{H}_R^T[\mathbf{g} - \mathbf{g}_T] + \sigma_1\mathbf{\Psi}^T[\boldsymbol{\theta}_f - \boldsymbol{\chi}_1] + \sigma_3\mathbf{D}^T[\mathbf{z} - \boldsymbol{\chi}_3]$

3. $\boldsymbol{\eta}_2 = \mathbf{P}^T\mathbf{H}^T\mathbf{g} + \sigma_2\mathbf{\Psi}_z^T(\boldsymbol{\theta}_z - \boldsymbol{\chi}_2) + \sigma_3(\mathbf{D}\mathbf{f} - \boldsymbol{\chi}_3)$

4. $\mathbf{f}^{\iota+1} = \left[\alpha_T^2\mathbf{H}_T^T\mathbf{H}_T + \alpha_R^2\mathbf{H}_R^T\mathbf{H}_R + \sigma_1\mathbf{I} + \mathbf{H}^T\mathbf{H} + \sigma_3\mathbf{D}^T\mathbf{D}\right]^{-1}\boldsymbol{\eta}_1$

5. $\mathbf{z}^{\iota+1} = \left[(\sigma_2 + \sigma_3)\mathbf{I} + \mathbf{P}^T\mathbf{H}^T\mathbf{H}\mathbf{P}\right]^{-1}\boldsymbol{\eta}_2$

6. $\boldsymbol{\theta}_z^{\iota+1} = \text{soft}(\mathbf{\Psi}_z\mathbf{z}^{(\iota+1)} + \boldsymbol{\chi}_2^{(\iota)}, \tau_z/\sigma_2)$

7. $\boldsymbol{\theta}_f^{\iota+1} = \text{soft}(\mathbf{\Psi}_f\mathbf{f}^{(\iota+1)} + \boldsymbol{\chi}_1^{(\iota)}, \tau_f/\sigma_1)$

8. $\boldsymbol{\chi}_1^{\iota+1} = \boldsymbol{\chi}_1^{\iota} - (\boldsymbol{\theta}_f - \mathbf{\Psi}_f\mathbf{f})$

9. $\boldsymbol{\chi}_2^{\iota+1} = \boldsymbol{\chi}_2^{\iota} - (\boldsymbol{\theta}_z - \mathbf{\Psi}_z\mathbf{z})$

10. $\boldsymbol{\chi}_3^{\iota+1} = \boldsymbol{\chi}_3^{\iota} - (\mathbf{z} - \mathbf{D}\mathbf{f})$

11. $\mathbf{g}_R^{\iota+1} = \mathbf{g}_R^{\iota} - (\mathbf{g} - \alpha_T\mathbf{H}_T\mathbf{f})$

12. $\mathbf{g}_T^{\iota+1} = \mathbf{g}_T^{\iota} - (\mathbf{g} - \alpha_R\mathbf{H}_R\mathbf{f})$

end

25 from ICVL Boaz Arad y Ohad Ben-Shahar. “Sparse Recovery of Hyperspectral Signal from Natural RGB Images”. En: *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, págs. 19-34 and 3 from C^{3}SLFI Miguel Marquez, Hoover Rueda-Chacon y Henry Arguello. “Compressive Spectral Light Field Image Reconstruction via Online Tensor Representation”. En: *IEEE Transactions on Image Processing* 29 (2020), págs. 3558-3568. These spectral images were resized to have equal spatial and spectral dimensions $N_x \times N_y = 512 \times 512$ and $N_\lambda = 24$, respectively. In the simulations, we compare the performance of the proposed system and

methodology on three fronts. First, we determine the optimal intensity splitting ratio as a function of the sensing matrix conditioning metrics (i.e., condition number and Gershgorin circles) and in terms of the compressed measurement distortion metrics (i.e., signal-to-noise-ratio and intensity ratios between $\mathbf{g}_T$ and $\mathbf{g}_R$). Second, we test the impact of the beam-splitter splitting ratio in the reconstruction quality. Third, we compare the proposed methodology against the CASSI and against the CASSI with grayscale side information Lizhi Wang y col. "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging". En: *IEEE transactions on pattern analysis and machine intelligence* 39.10 (2016), págs. 2104-2111.

The CASSI measurements are simulated using a fix random black-and-white coded aperture with $50\,\%$ transmittance. The sparse promoting bases are set to be $\mathbf{\Psi}_z = \mathbf{\Psi}_{1D-W} \otimes \mathbf{\Psi}_{1D-W}$ and $\mathbf{\Psi}_f = \mathbf{\Psi}_z \otimes \mathbf{\Psi}_{1D-DCT}$, where $\mathbf{\Psi}_{1D-W} \in \mathbb{R}^{N_x \times N_x}$ and $\mathbf{\Psi}_{1D-DCT} \in \mathbb{R}^{N_\lambda \times N_\lambda}$ represent the 1D Wavelet (Symlet 8) basis, and the 1D discrete cosine transform, respectively. All simulations were conducted using an Intel Core i7 3960X 3.30 GHz processor with 32 GB RAM memory. To compare the quality of the reconstructions, we use the peak signal to noise ratio (PSNR), the root mean squared error (RMSE), the spectral angular mapper (SAM) and the structural similarity (SSIM) metrics. Specifically, the PSNR and SSIM metrics are calculated band-per-band and averaged, the RMSE metric is calculated pixel-wise, and the SAM metric is estimated for each spectral signature and averaged.

3.3.1. Beam-splitting-ratio Theoretical Analysis One of the most critical parameters in the design of the proposed sensing matrix is the splitting ratio of the beam-splitter. Thus, to determine the splitting ratio that yields the best trade-off, between the panchromatic and spectral reconstruction quality, the following four metrics were used: conditional number, Gershgorin circles, signal-to-noise-ratio, and dynamic range. The conditional number David J Brady. *Optical imaging and spectroscopy*. John Wiley & Sons, 2009 and the Gershgorin radius Howard E Bell. "Gershgorin's

theorem and the zeros of polynomials". En: *The American Mathematical Monthly* 72.3 (1965), págs. 292-295 metrics provide an approximation of the degree of ill-posedness of the inverse problem, i.e., without demanding high computational resources Yuri Mejia y Henry Arguello. "Binary codification design for compressive imaging by uniform sensing". En: *IEEE Transactions on Image Processing* 27.12 (2018), págs. 5775-5786; Miguel Marquez y Henry Arguello. "Coded aperture optimization for single pixel compressive computed tomography". En: *Journal of Computational and Applied Mathematics* 348 (2019), págs. 58-69; Andrés Jerez, Miguel Márquez y Henry Arguello. "Adaptive coded aperture design for compressive computed tomography". En: *Journal of Computational and Applied Mathematics* 384 (2020), pág. 113174. The signal-to-noise ratio metric quantifies the $\mathbf{g}_T$ and $\mathbf{g}_R$ degradation sensibility for scenarios with high noise levels.

- **Gershgorin radius and conditional number:** These metrics determine the lower bounds of the sensing matrix eigenvalues and establish the sensitivity of the matrix to distortions. Figure 19 summarizes the singular radii and condition number analysis for matrices $(\alpha_T\mathbf{H}_T + \alpha_R\mathbf{H}_R)\mathbf{D}^T$ and $(\alpha_T\mathbf{H}_T + \alpha_R\mathbf{H}_R)$, varying the splitting ratio within $\alpha_T \in \{0, 0.1, 0.2, ..., 0.9, 1\}$. To determine the CASSI-VSI's optimal splitting ratio configuration following a statistical process, we analyze 1000 CA realizations –per splitting ratio– for images with sizes $N_x = N_y = 64$ and $N_\lambda = 12$. Figure 19 shows that the best conditioning performance for $\mathbf{HD}^T$ and $\mathbf{H}$ matrices is achieved at $\alpha_T = 0$ and $\alpha_T = 1$, respectively. However, these exact splitting values also represent the worst-case scenario for $\mathbf{H}$ and $\mathbf{HD}^T$, respectively. Nonetheless, a good trade-off between them is obtained at around $\alpha_T = 0.75$, where they intersect, which suggests that $\alpha_T/\alpha_R > 2$.

- **Signal-to-noise ratio:** To determine the optimal splitting ratio of the beamsplitter from a signal-to-noise ratio (SNR) perspective, we first add to $\mathbf{g}$ –generated

with different configurations of α_T/α_R– a noise of 20 SNR. The SNR metric is defined as $SNR = 10 \cdot \log(\sum_{n=1}^{N} f_n^2 / \sum_{n=1}^{N} [f_n - \hat{f}_n]^2)$, where f represent the noised signal, f is the original signal, and N is the signal length. Then, we calculated the resulting SNR for the two decoupled measurements, $\mathbf{g}_R$ and $\mathbf{g}_T$, which are plotted in Fig. 20(A). Additionally, the average SNR curve between $\mathbf{g}_R$ and $\mathbf{g}_T$ is included for comparison purposes. This analysis was developed using fifty spectral images. The noise is generated as additive colored Gaussian noise from the total intensity, $\mathrm{g} = \mathrm{g}_T + \mathrm{g}_R$. Notice that, in practice, photon noise is often modeled using a Gaussian distribution whose variance depends on the expected photon count. This approximation is typically assumed because, for larger counts, the central limit theorem ensures that the Poisson distribution approaches the Gaussian one. Note that the intersection of both curves, is considered as the optimal SNR trade-off between the measurements. This premise is verified by plotting the averaged SNR (black line), where it can be observed that the maximum SNR is achieved at around $\alpha_T = 0.6$. A detailed analysis of the averaged SNR behavior along the rows of the split measurements, g_T and g_R, is summarized in **Appendix 9.1** in the supplementary material.

- **Dynamic range:** The CASSI-VSI's two arms receive –ideally– as input a similar wavefront –but with complementary wavefront intensities between them– which are attenuated separately in each arm by the elements: lens-prism (CASSI measurement) and relay lens (grayscale measurement). Then, the resulting two wavefronts converging to the sensor array have disparate intensities per spatial location, generating unbalanced measurements. To establish the optimal α_T from a dynamic range perspective, we analyze the ratio between g_T and g_R, where $\frac{(\mathbf{g}_T)_i}{(\mathbf{g}_R)_i} = \frac{(\mathbf{g}_R)_i}{(\mathbf{g}_T)_i} = 1$ is the desired dynamic range performance, which guarantees they are at the same scale. Figure 20(B) illustrates the cumulative absolute error between the intensity ratios, defined as $\sum_i |\frac{(\mathbf{g}_T)_i}{(\mathbf{g}_R)_i} - \frac{(\mathbf{g}_R)_i}{(\mathbf{g}_T)_i}|$, for the

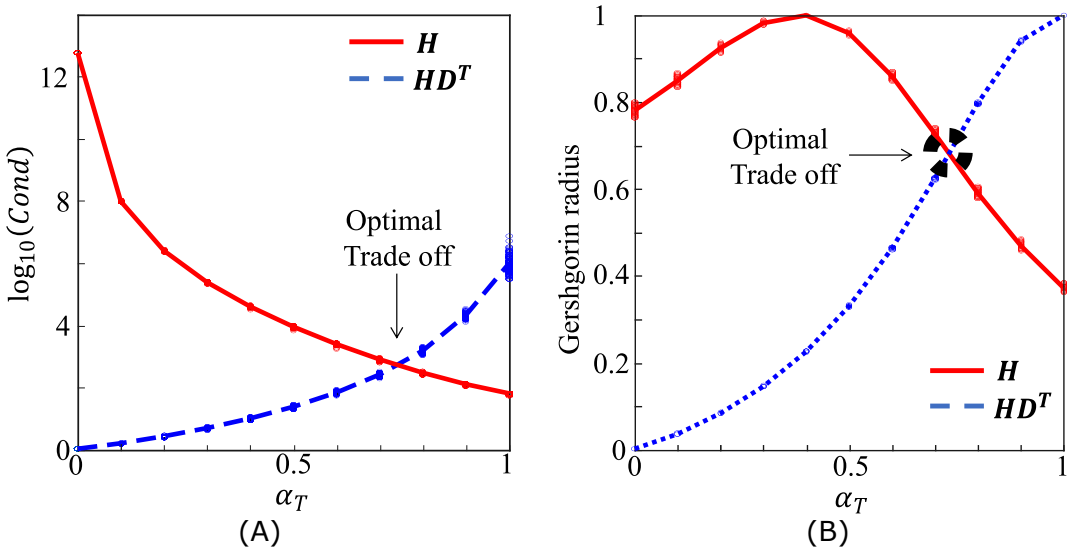


Figura 19. (A) Condition number and (B) normalized Gershgorin radii of the matrices $\mathbf{HD}^T$ (blue line) and $\mathbf{H}$ (red line), where $\mathbf{H} = \alpha_T \mathbf{H}_T + \alpha_R \mathbf{H}_R$. As a reminder, α_R and α_R are the splitting ratio in intensity.

different splitting ratios, averaging over the 50 spectral images as in Fig. 20(A). Based on the cumulative absolute error metric tendency, it can be observed that, from a dynamic range perspective, the optimal splitting ratio lies around $\alpha_T = 0.7$. A detailed analysis of the averaged intensity behavior along the rows of the split measurements, $\mathbf{g}_T$ and $\mathbf{g}_R$, is summarized in **Appendix 9.2** in the supplementary material.

3.3.2. Beam-splitting-ratio Reconstruction Performance To verify the theoretically-found best beam-splitting ratio above, we test the performance of the proposed optical system by reconstructing $\mathbf{f}$ from the compressed measurement $\mathbf{g}$, varying the eleven splitting ratios ($\alpha_T : \alpha_R$) established in Figs. 19 and 20. In particular, the compressed measurements are simulated with three different additive colored Gaussian noise of SNR = $\{15, 20, 30\}$ dB.

Figure 21 summarizes the PSNR, SSIM, RMSE and SAM metrics of image reconstruction using the proposed algorithm on the set of 50 spectral images with a noise

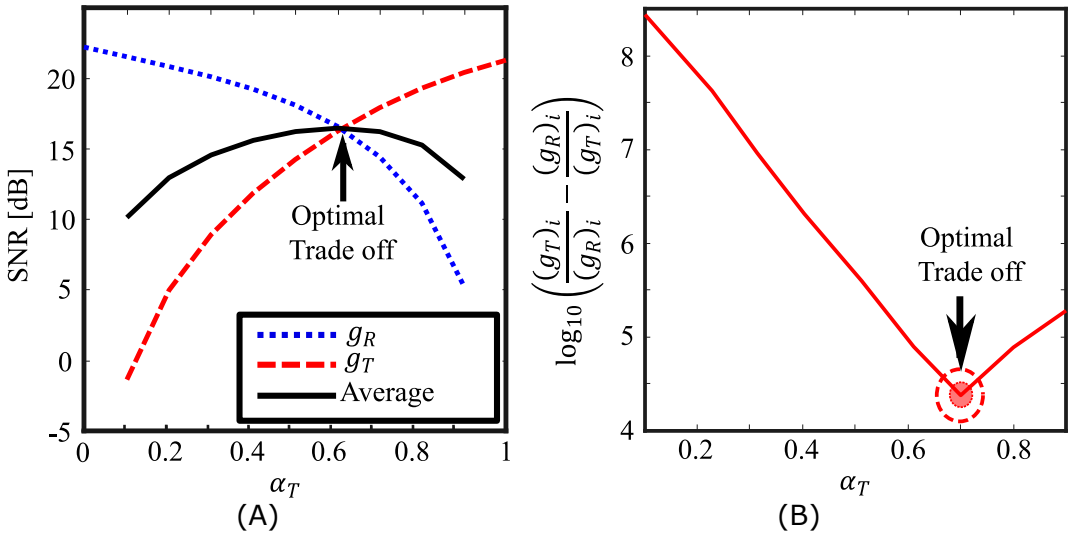


Figura 20. (A) SNR metric between the sensor noise and the compressed measurements, (blue line) g_R, (red line) g_T, (black line) their average. (B) Cumulative absolute error between the intensity ratios of the compressed measurements g_T and g_R.

level of SNR=20 [dB]. A detailed analysis of the A-VSI methodology reconstruction performance for the three different noise levels is summarized in **Appendix 9.C** in the supplementary material. In these figures, the metric for each image is plotted as a green dot, and the average value and standard deviation are included as error bars. There, the best spectral results, in terms of SAM, are obtained for $\alpha_T = 1$. This is expected since for this condition the light throughput is used solely in the spectral arm. On the other hand, the best PSNR, SSIM, and RMSE metrics are attained at around $\alpha_T \in \{0.6\}$. To complement the A-VSI reconstruction performance analysis, Fig. 22(A) presents the results of the panchromatic image estimation in terms of the PSNR. There, the PSNR results tend to improve when $\alpha_T \to 0$ with $\alpha_T = 0$ as the optimal splitting ratio for both algorithms. This result was expected because, when $\alpha_T = 0$, the grayscale's power intensity is the larger possible relative to the background noise power intensity.

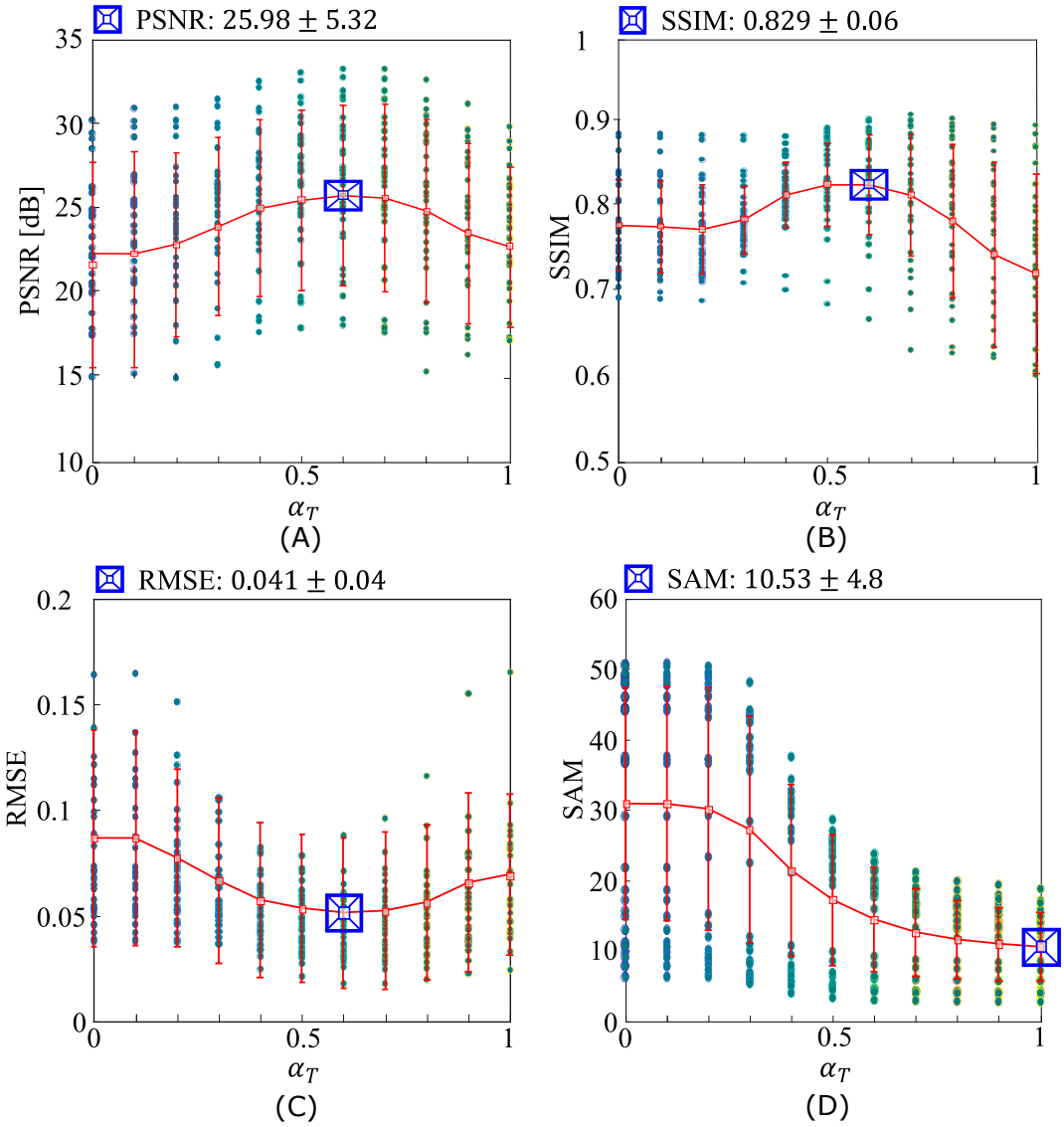


Figura 21. Reconstruction results in terms of (A) PSNR, (B) SSIM, (C) RMSE, and (D) SAM.

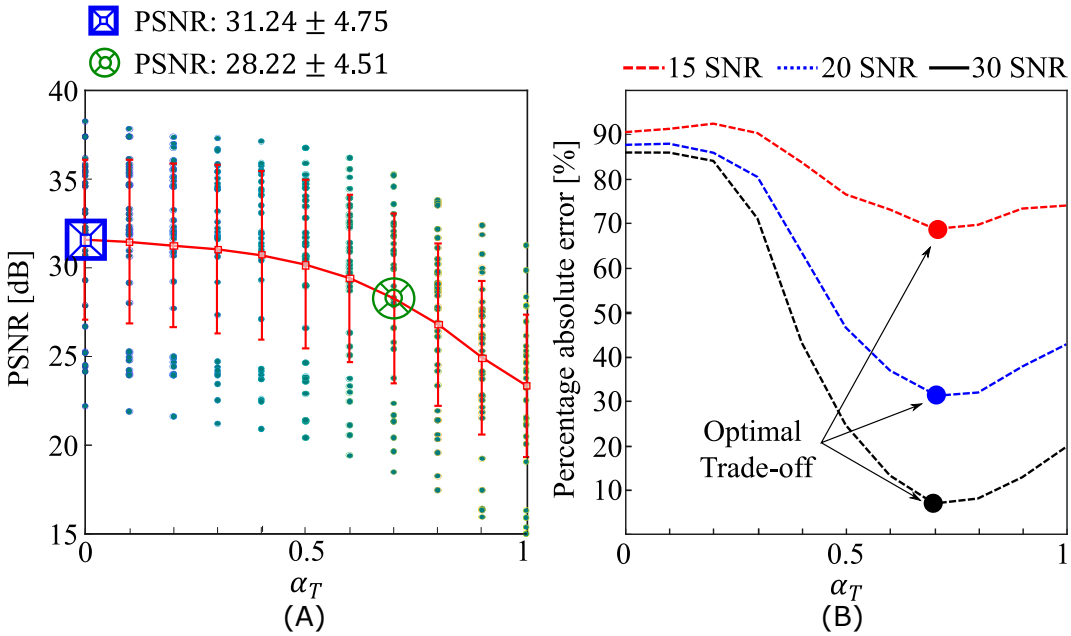


Figura 22. (A) Panchromatic image reconstruction results in terms of PSNR. Overall best trade-off reconstruction result is highlighted in green, whereas the panchromatic best in blue. (B) General performance metric results for different SNR scenarios.

3.3.3. General Performance Metric Although optimal beam-splitting ratios were found for each metric in Fig. 21, for real optical implementations purposes it is desired to establish a unique (optimal) splitting ratio. Therefore, a general performance metric, that fuses all the metric results to determine the optimal splitting, is proposed. First, the metrics are transformed into dimensionless metrics, where the percentage error metric is selected. To develop this metric, we first define four matrices $\mathbf{\Phi}^0$, $\mathbf{\Phi}^1$, $\mathbf{\Phi}^2$, and $\mathbf{\Phi}^3$, that contain the PSNR, SSIM, RMSE, and SAM results, respectively, for all the tested images and the evaluated splitting ratios. In particular, these matrices have dimensions of $L_1 \times L_2$, with $L_1 = 50$ (images) and $L_2 = 11$ (splitting scenarios). Then, a percentage error analysis is developed for each individual row of the matrices, defined as $\Phi^0_{\iota_1,\iota_2} = |\frac{\max(\mathbf{\Phi}^0_{\iota_1,:}) - \Phi^0_{\iota,\iota_2}}{\max(\mathbf{\Phi^0_{\iota_1},:})}|$, $\Phi^1_{\iota_1,\iota_2} = |\frac{\max(\mathbf{\Phi}^1_{\iota_1,:}) - \Phi^1_{\iota,\iota_2}}{\max(\mathbf{\Phi^1_{\iota_1},:})}|$, $\Phi^2_{\iota_1,\iota_2} = |\frac{\min(\mathbf{\Phi}^2_{\iota_1,:}) - \Phi^2_{\iota_1,\iota_2}}{\min(\mathbf{\Phi^2_{\iota_1},:})}|$, and $\Phi^3_{\iota_1,\iota_2} = |\frac{\min(\mathbf{\Phi}^3_{\iota_1,:}) - \Phi^3_{\iota_1,\iota_2}}{\min(\mathbf{\Phi}^3_{\iota_1,:})}|$ with $\iota_1 = \{0, ..., L_1 - 1\}$ and $\iota_2 = \{0, ..., L_2 - 1\}$. Then, for each splitting ratio the average percentage error is

defined as,

$$p_{\iota_2} = \frac{1}{4L_1}\left[\sum_{q=0}^{3}\sum_{\iota_1=0}^{L_1-1}\Phi^{q}_{\iota_1,\iota_2}\right], \tag{49}$$

where $\mathbf{p} \in \mathbb{R}^{L_2}$ is a vector that contains the global cumulative percentage error for each splitting ratio. Based on Eq. (49), inset (B) in Fig. 22 summarizes the general performance metric for the results reported in Fig. 21. In Fig. 22 (B) it can be noticed that the lowest overall percentage error achieved by the A-VSI reconstruction methodology is obtained at around $\alpha_T = 0.7$. This result holds independent of the level of noise, and $\alpha_T \neq 1$, which means that the CASSI-VSI overcomes the CASSI system in overall performance. This result validates the theoretical analysis developed in Fig. 19, which established that the transmission path requires higher light throughput than the reflection path, with a proportion of around $\{\alpha_T, \alpha_R\} = \{0.7 : 0.3\}$.

3.3.4. Comparison with the State-of-the-art Imaging Systems With the optimal beam-splitting ratio established, we now evaluate the performance of the CASSI-VSI with the state-of-the-art imaging systems. For this, we compare against the conventional CASSI Wagadarikar y col., “Single disperser design for coded aperture snapshot spectral imaging” and the dual-arm CASSI Wang y col., “Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging” that uses panchromatic side information. It is worth noting that the CASSI-VSI is fixed with the optimal splitting ratio $\{\alpha_T, \alpha_R\} = \{0.7, 0.3\}$ and all the measurements are distorted with a noise power level of SNR = 30dB. Further, the panchromatic image used in the dual-arm CASSI is simulated without noise and perfect image registration is assumed. Figure 23 summarizes the reconstruction results. Specifically, results are evaluated in four fronts: first, the illustration of RGB composites of the attained reconstructions along with the SAM; second, the illustration of the panchromatic re-

constructions along with the SSIM metric; third, the cummulative absolute errors per spectral band along with the PSNR and RMSE metrics; and fourth, an illustration of 8 out of the 24 reconstructed bands is included. It can be noticed that the visual quality attained by the CASSI-VSI is comparable to that of the dual-arm CASSI, and overcomes that of CASSI. The metric results in terms of PSNR, RMSE and SSIM, validate the visual quality results. Furthermore, the dual-arm CASSI overcomes the results obtained by the CASSI-VSI and CASSI in terms of the SAM metric, respectively. Note that in the scenarios where just a single sensor is available, the proposed optical system is a suitable alternative.

3.3.5. Comparison with the State-of-the-art Algorithms In this subsection, instead of using our proposed A-VSI reconstruction algorithm, we use a deep learning-based reconstruction algorithm, lambda-net Lizhi Wang y col. “Hyperspectral image reconstruction using a deep spatial-spectral prior”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, págs. 8032-8041, to evaluate the reconstruction performance attained with a compressed measurement of the CASSI against our proposed CASSI-VSI. To conduct a fair comparison, the λ-net was trained using the CASSI and the CASSI-VSI forward models independently. Both λ-net instances were trained using the author’s recommended parameters, including, training dataset (150 spectral images), testing dataset (10 images), image size ($256 \times 256 \times 24$), fixed coded aperture realization, and batch size (2). For the CASSI-VSI instance, we fixed the splitting ratio to $\alpha_T = 0.7, \alpha_R = 0.3$. Figure 25 illustrates the evolution of the averaged testing accuracy, measured in terms of the PSNR of image reconstruction, for the ten spectral images used for testing. Note that this calculation was conducted every 200 epochs, during the 33000 epochs used for training. The PSNR tendency in Fig. 25 shows that the λ-net with the CASSI-VSI forward model (red line) entails an improvement of up to 3.49 dB over the results obtained by the λ-net with the CASSI forward model (blue line). To further evaluate the

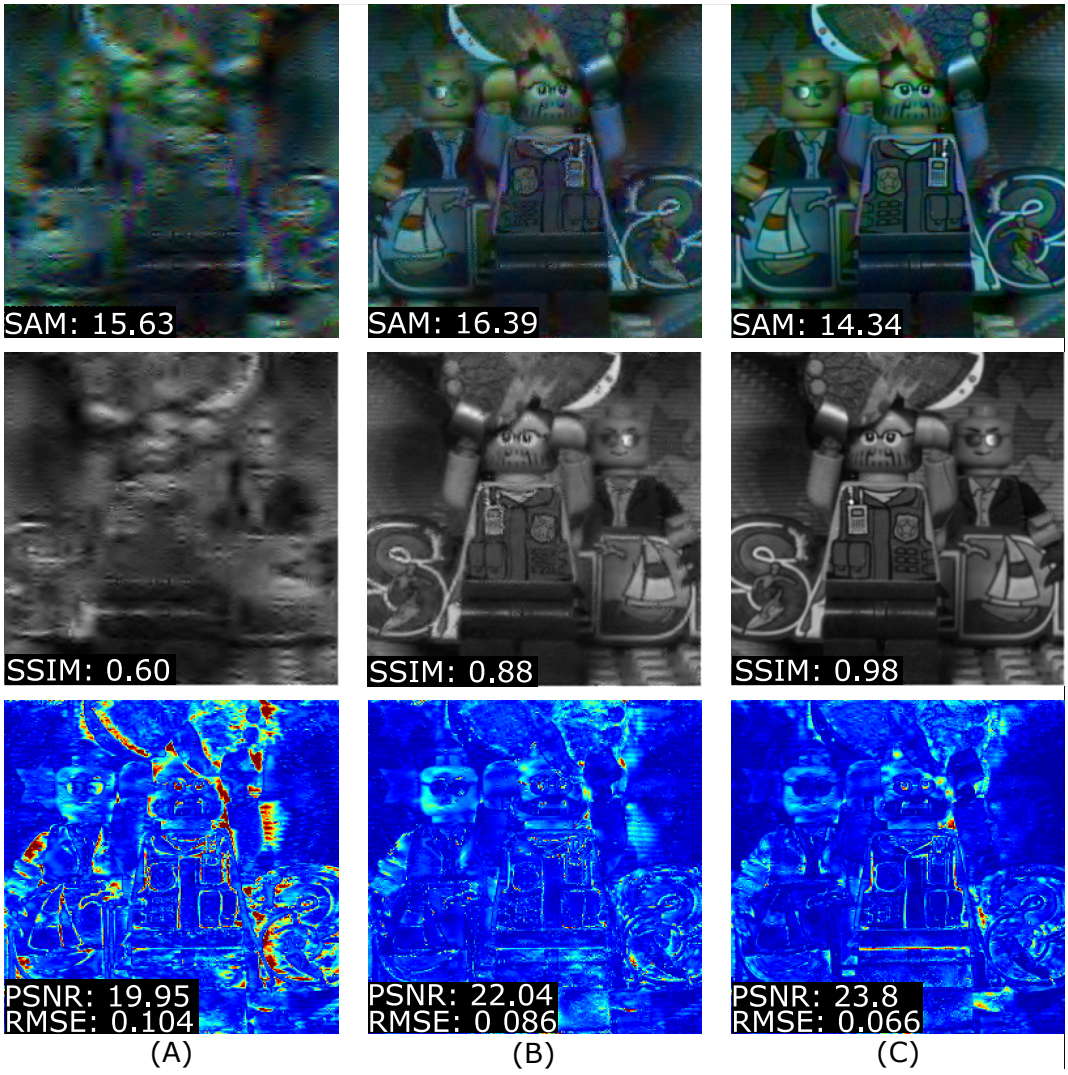


Figura 23. Reconstructed images mapped to RGB, when SNR$= 30$ [dB] for (A) CASSI, (B) CASSI-VSI, and (C) dual-arm CASSI.

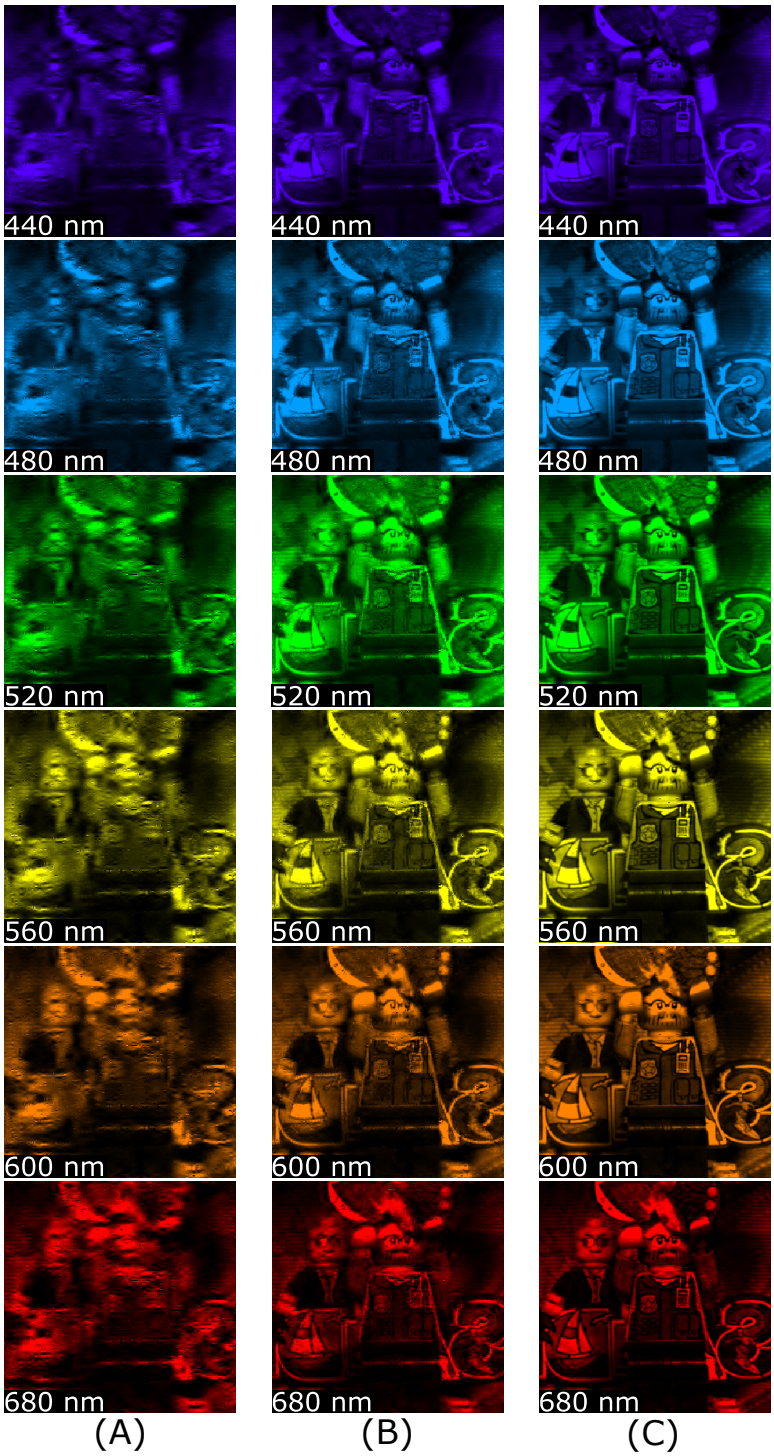


Figura 24. Reconstructed images mapped to RGB, when SNR$= 30$ [dB] for (A) CASSI, (B) CASSI-VSI, and (C) dual-arm CASSI.

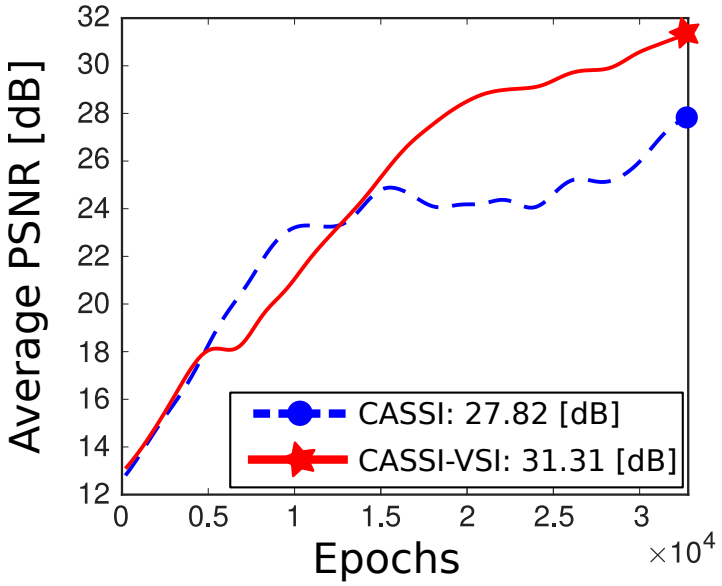


Figura 25. Evolution of the λ-net averaged testing accuracy, in terms of PSNR, for the CASSI (blue line) and CASSI-VSI (red line) forward models. The network is trained during 33000 epochs, and the PSNR is calculated every 200 epochs using the λ-net test set with 10 spectral images.

performance of these networks, Figs. 26-27 shows the result of testing them in 40 out of the 50 spectral images used in te simulations, adding a colored Gaussian noise with SNR of 20 dB and 30 dB, respectively. The networks are not tested in all the 50 images because 10 of them were part of the training dataset. In this figure it can be confirmed that, in general, the PSNR attained with the CASSI-VSI outperforms that of CASSI, for both noisy scenarios.

3.4. Proof of concept experiments

A testbed of the proposed system was built in our laboratory, so as to demonstrate the proposed ideas through a proof-of-concept prototype with a splitting ratio of $\alpha_T = 0.75 : \alpha_R = 0.25$. This prototype, shown in Fig. 28, uses a matched achromatic doublet pair (Thorlabs MAP10100100-A, f1=100.0 mm, f2=100.0 mm) as the objective lens (L1) to image the scene onto a digital micromirror device (DMD, Texas Instruments, D4120), which emulates the coded aperture. Then, a standard relay lens (L2)

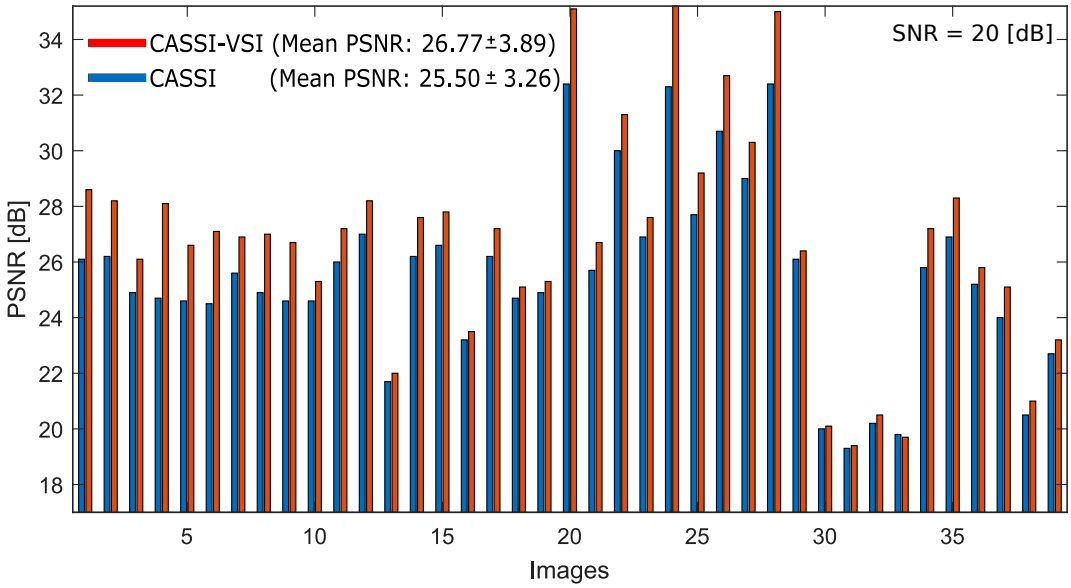


Figura 26. Testing accuracy of the λ-net after 33000 epochs, in terms of PSNR, for the CASSI (blue bars) and CASSI-VSI (red bars). The testing is conducted over 40/50 spectral images used in the simulations, adding colored Gaussian noise to the measurements with SNR=20 dB.

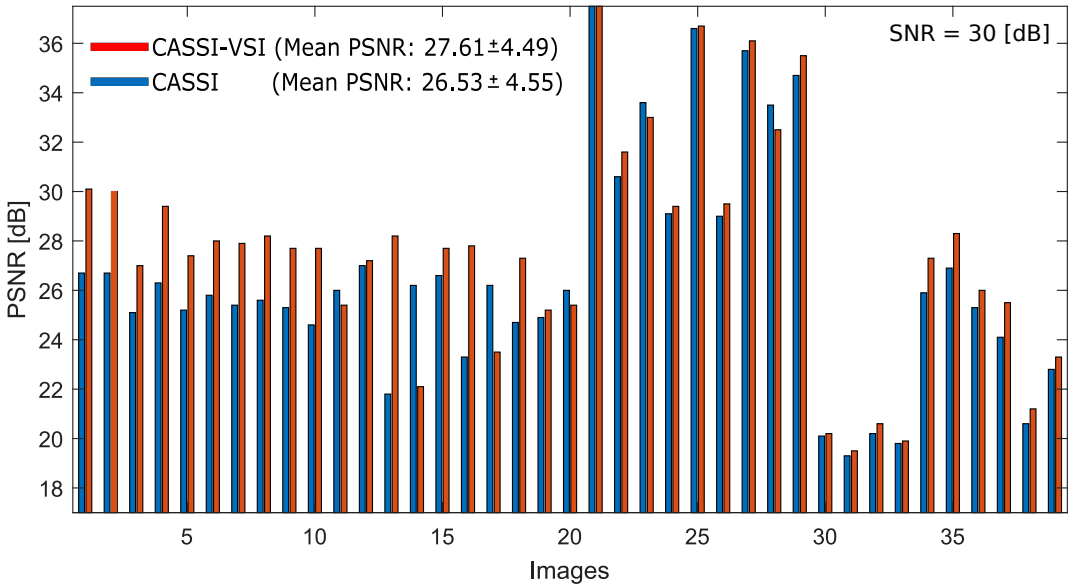


Figura 27. Testing accuracy of the λ-net after 33000 epochs, in terms of PSNR, for the CASSI (blue bars) and CASSI-VSI (red bars). The testing is conducted over 40/50 spectral images used in the simulations, adding white Gaussian noise to the measurements with SNR=30 dB.

(Thorlabs AC254-100-A-ML, f=100 mm) located at 100 mm of the DMD is used to attain two 4f-systems, split by the beam splitter (BS1) (Thorlabs CCM1-BS013, 30 mm non-polarizing, 50:50 beamsplitter), which transfers half the light to the CASSI-arm

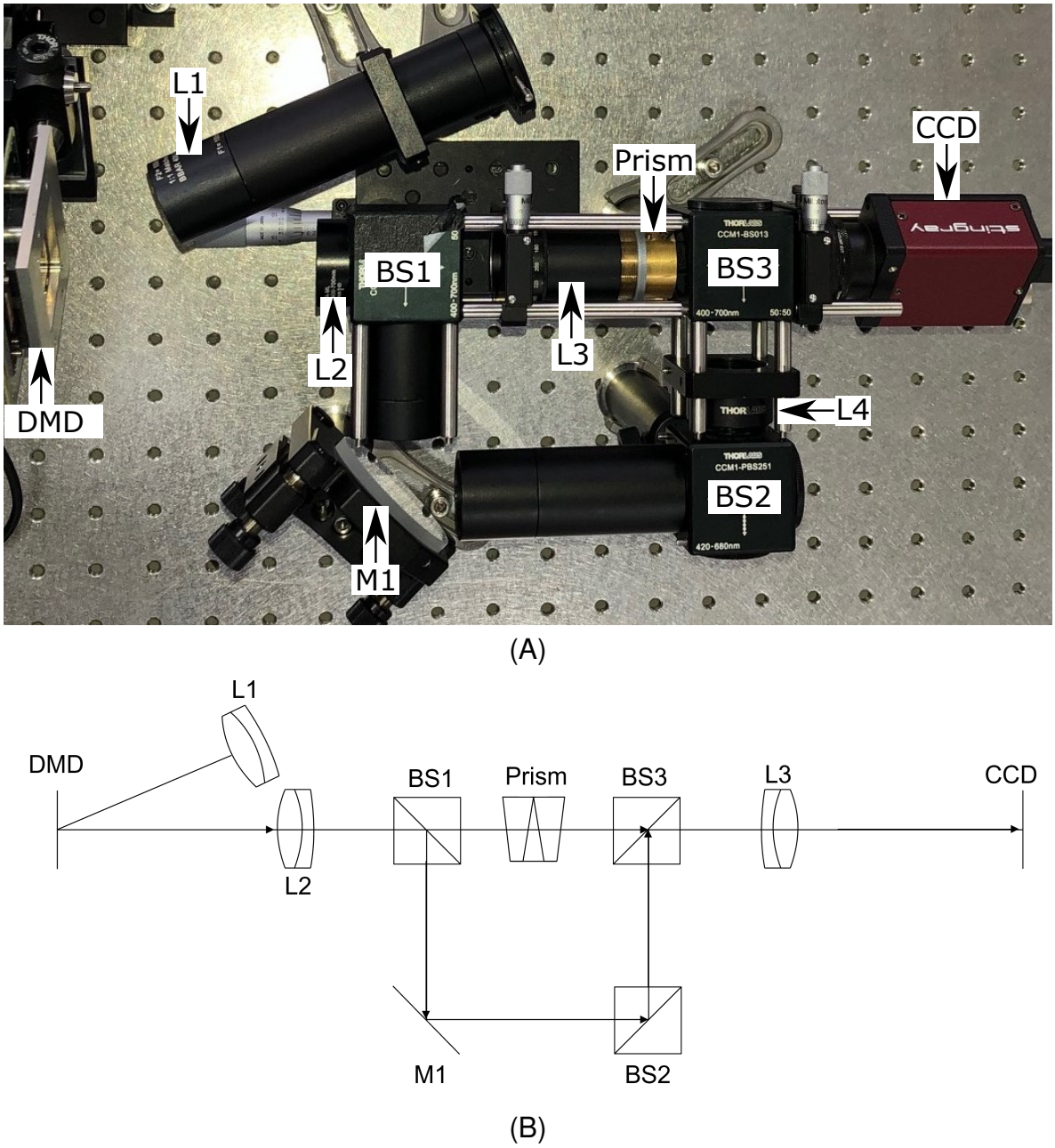


Figura 28. (A) Testbed implementation of the CASSI-VSI. (B) Geometric sketch of the optical system.

(transmissive) and the other half to the side-arm (reflective). In the CASSI-arm, a lens (L3) (Thorlabs AC254-100-A-ML, f=100 mm) and a customized double Amici prism (see Supplementary material 9) are coupled to a rotation mount (Thorlabs CRM1P, 30 mm cage rotation mount) to precisely adjust the dispersion angle horizontally. In the reflective side-arm, a broadband dielectric mirror (M1) (Thorlabs BB4-E03), and a second beam splitter (BS2) (30 mm Cage Cube-Mounted Non-Polarizing Beamsplitter), are used jointly with the lens (L4) (Thorlabs AC254-100-A-ML, f=100 mm) to transfer the coded light to the sensor. Finally, a beam splitter (BS3) is used to mix the light from both arms and onto the sensor (Stingray F-080B, 4.65 μm pixel size), where both light paths are integrated simultaneously.

The CASSI-VSI design is based on two 4F configurations that share the same input lens (L1) but differ in the exit lens (L3, L4). This 4f design avoids magnification and reduces the pixel mismatch. Due to the 4f relay system and the pixel sizes of the DMD (13.6 μm) and image sensor (4.65 um), a 2×2 pixel binning in the DMD is conducted, which results in a 6×6 binning in the image sensor. Note that the BS2 could be replaced by a broadband dielectric mirror if the BS1 exhibited a splitting ratio of $(\alpha_T : \alpha_R) = (0.75, 0.25)$.

The experiments evaluate two target scenes named Church and Boat, for which a compressive projection was acquired with the testbed. To attain the traditional CASSI system, the reflective arm was blocked and the integration time doubled. The raw compressive projections exhibit a spatial resolution of 680×741 pixels. Thus, $L = 741 - 680 + 1 = 62$ spectral bands can be recovered. Figure 29(first row) shows the (A) Boat and (B) Church raw measurements, which are given as input to the A-VSI algorithm. Note that the **Algorithm 1** returns spectral data cubes with their spectral signatures without illumination corrections, i.e., in intensity. Therefore, an additional offline reconstruction step is the spectral reflectance estimation, see Supplemental material 9. Additionally, Fig. 29(A)-(B)(second row) illustrates the RGB

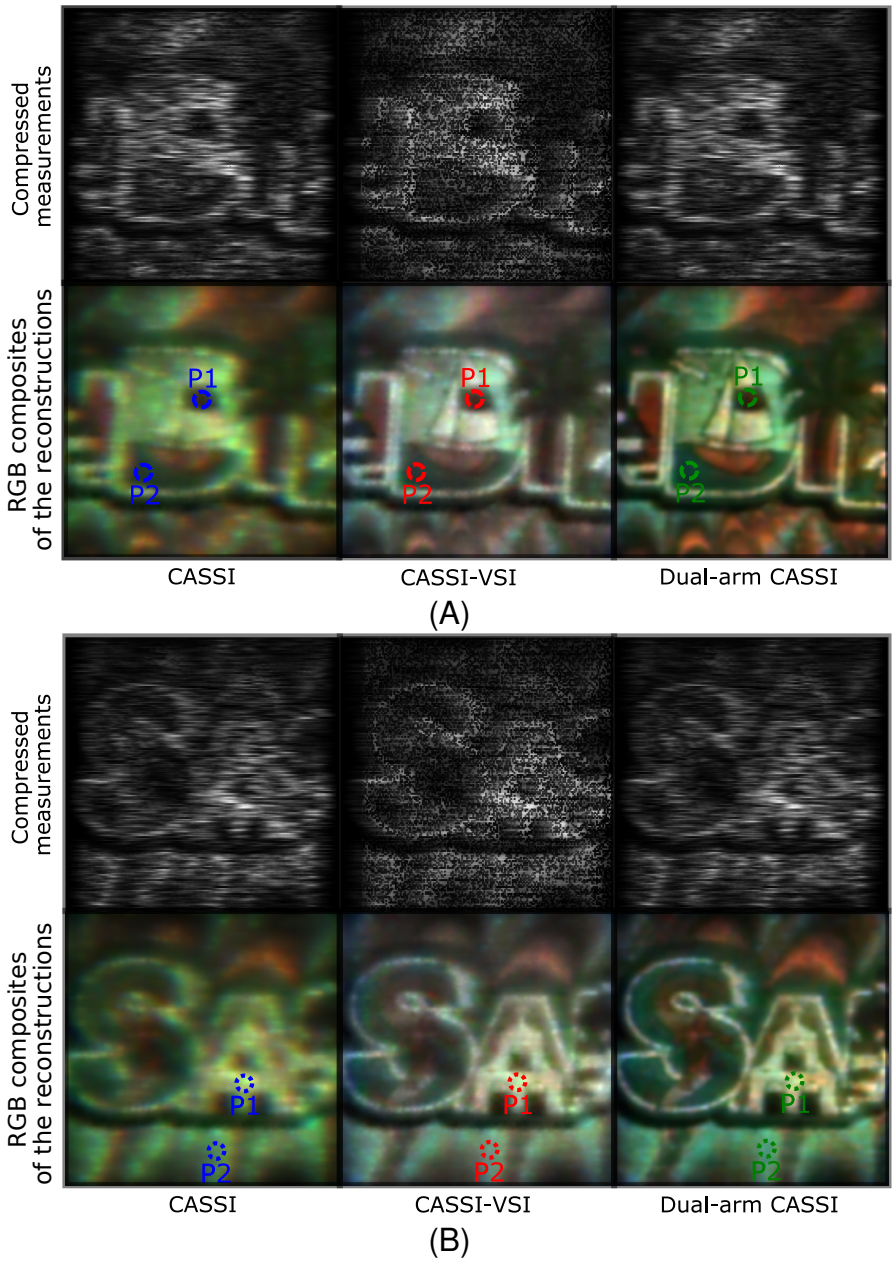


Figura 29. Testbed results. Compressed measurements and RGB composites of the (A) Boat and (B) Church scenes reconstructions.

composites from the reconstructed spectral images via A-VSI. Finally, Figs. 30-31 shows the spatial reconstruction per band to evaluate the accuracy of the spectral reconstruction, 6 out of the 62 spectral bands are depicted. Moreover, to validate the fidelity of the spectral reconstruction, Fig. 32 shows the spectral signature of two

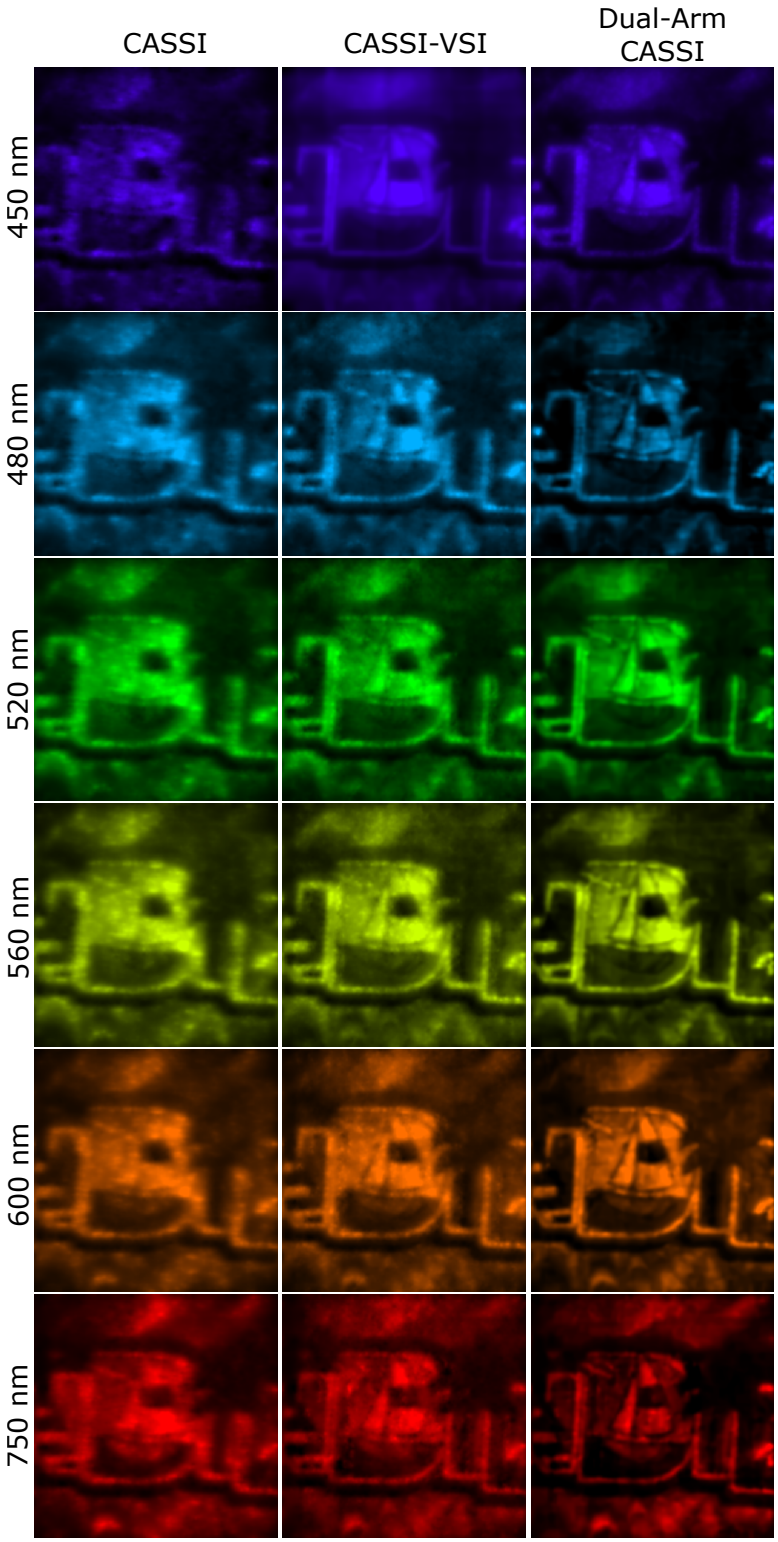


Figura 30. Testbed results. Illustration of 6 out of the 62 reconstructed spectral bands of the Boat scene.

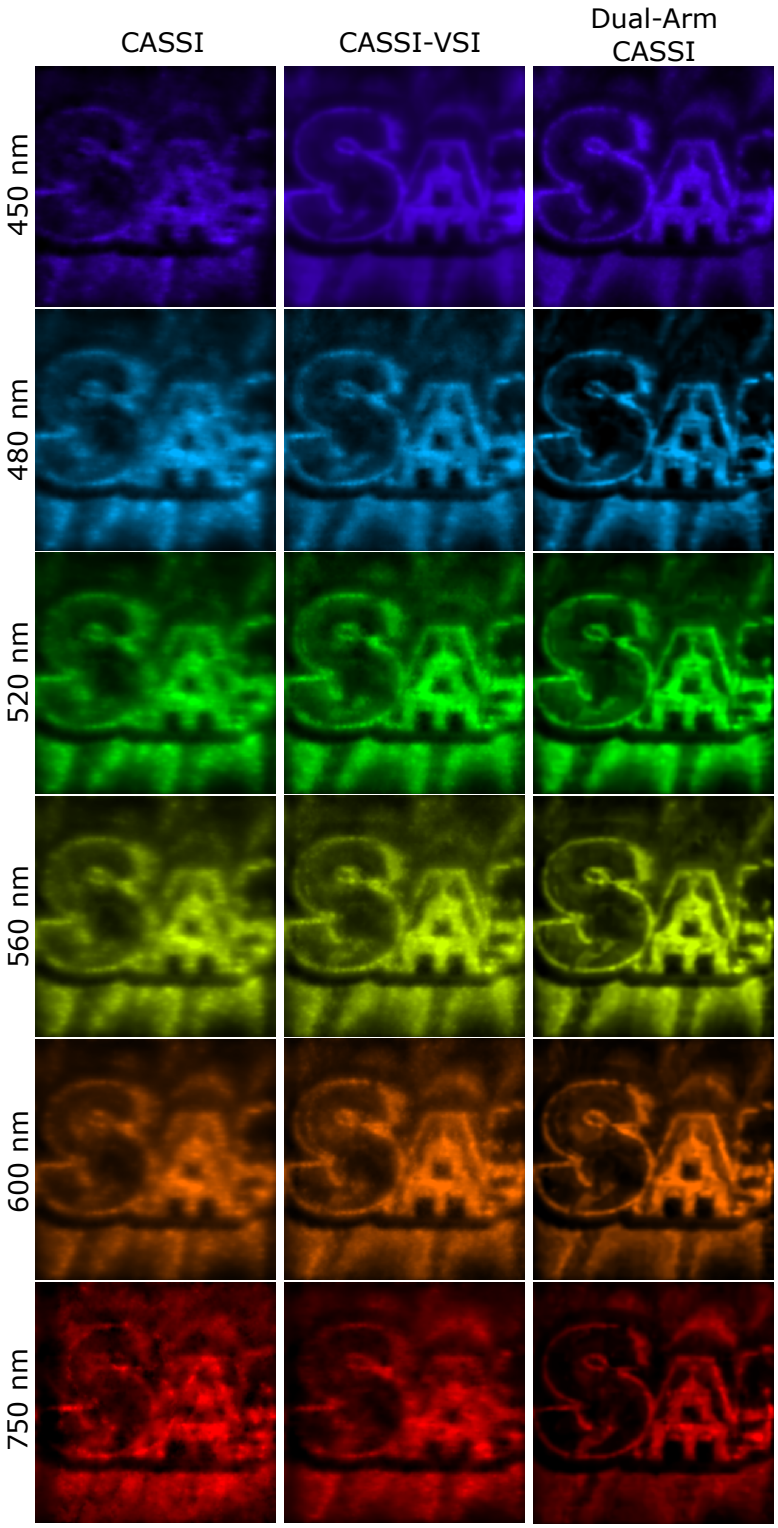


Figura 31. Testbed results. Illustration of 6 out of the 62 reconstructed spectral bands of the Church scene.

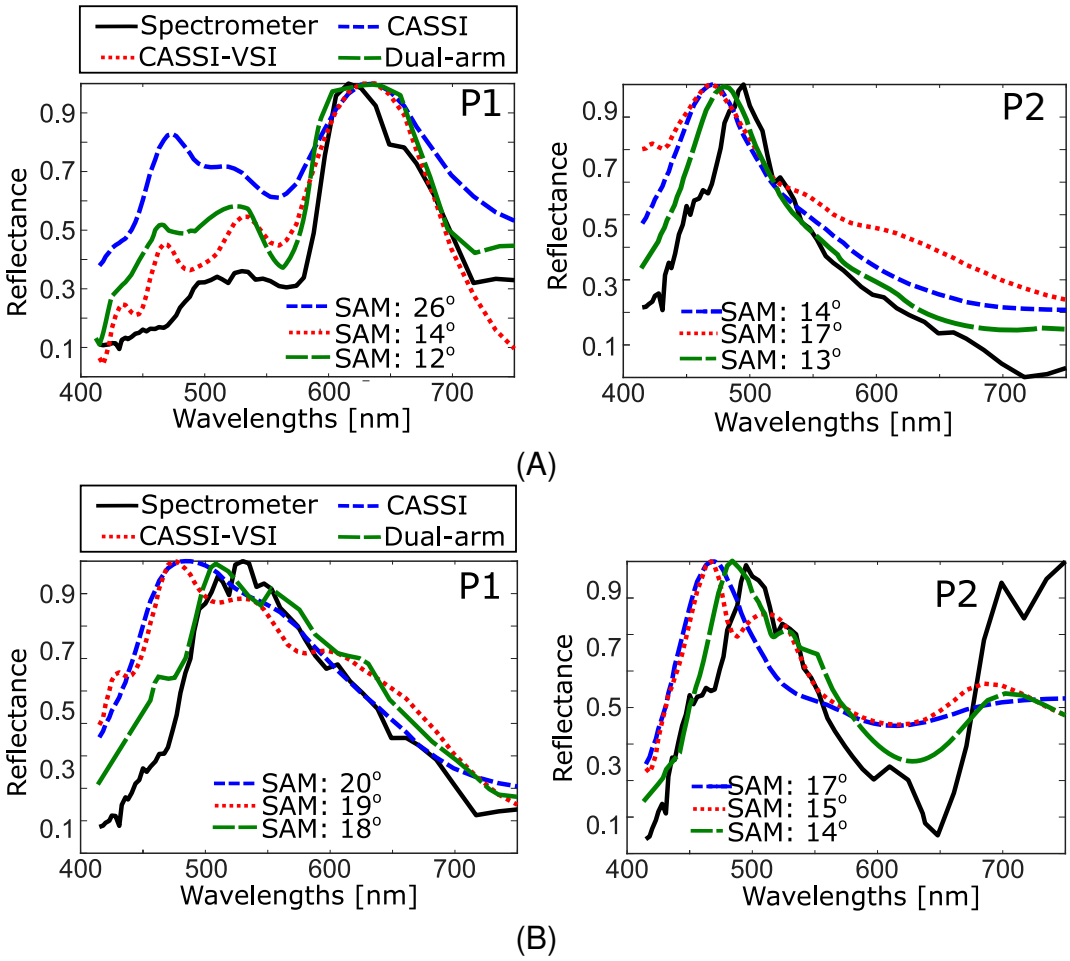


Figura 32. Spectral signatures of two spatial points from the (A) Boat (P1, red sailing boat; P2, blue ocean) and (B) Church scene (P1, white church; P2, blue background).

scene points measured with a spectrometer (Ocean Optics Flame S-VIS-NIR-ES), assumed to be the ground truth (black solid line), and compared with the ones obtained with the CASSI (blue dashed line), the CASSI-VSI (red dotted line), and the dual-arm CASSI (green dashed line). In this figure it can be noted that the reconstructed spectral signatures resemble those obtained with the spectrometer, with SAM values smaller than $18°$ for dual-arm CASSI, $19°$ for CASSI-VSI and $20°$ for CASSI. In summary, it can be observed that the proposed CASSI-VSI testbed overcomes the spatial and spectral performance of the CASSI system, and approximates that of the

dual sensor system.

3.5. Conclusions

This chapter proposed the CASSI-VSI system which exploits the usage of side-information to boost the spectral image quality of a compressive imaging system but employing a single image sensor. The multiplexing of a coded-and-dispersed projection along with a coded projection permitted the estimation of a panchromatic version of the scene which was simultaneously used as prior information for the spectral reconstruction. The proposed system relies on a single sensor, which goes against the state-of-the-art systems that rely on the use of additional sensors. The latter is critical when the system is deployed in other spectral domains, such as infrared, where the cost of additional sensors is prohibited.

The optimal light throughput between each branch of the CASSI-VSI was established theoretically, following the Gershgorin theorem and the condition number of the sensing matrices, and validated through simulations. These results concluded that the transmission path requires higher light throughput than the reflected projection, with a proportion of $\alpha_T, \alpha_R = 0.7 : 0.3$. Further, the performance of the proposed imaging system was demonstrated via simulations against the state-of-the-art alternatives, and through a proof-of-concept laboratory implementation, which confirmed that the proposal represents an efficient alternative to acquire spectral images when just a single sensor is available.

4. CASSDI-DFA: Compressive Spectral Depth Imager via Depth from Aberrations

Compressive spectral imaging (CSI) and Compressive Depth Imaging (CDI) systems take advantage of phase-coding elements such as, diffractive optical elements (DOEs), liquid crystal on silicon (LCoS), deformable mirror (DM), digital micromirror device (DMD) Baek y col., “Single-shot Hyperspectral-Depth Imaging with Learned Diffractive Optics”; Jeon y col., “Compact snapshot hyperspectral imaging with diffracted rotation”; Lai y col., “Compressed ultrafast tomographic imaging by passive spatiotemporal projections”. These approaches build upon a differentiable optical image formation model that accounts for either fixed or optimizable phase mask designs. In particular, DOE-based CDI Baek y col., “Single-shot Hyperspectral-Depth Imaging with Learned Diffractive Optics” and CSI Jeon y col., “Compact snapshot hyperspectral imaging with diffracted rotation” systems leverage the point-spread-function properties to customize DOEs with a depth/spectrally-varying PSF’s anisotropic shape. Inspired by the DOE framework, Baek y col., *End-to-end hyperspectral-depth imaging with learned diffractive optics* introduced a CoSDI system that retrieves a 4D datacube from a single exposure. DOE-based approaches allow compact systems at the expense of depth/spectral reconstruction accuracy, coding flexibility, and light throughput, thus falling short in capabilities for optical science applications Xianglei Liu y col. “Fast wide-field upconversion luminescence lifetime thermometry enabled by single-shot compressed ultrahigh-speed imaging”. En: *Nature Communications* 12.1 (2021), págs. 1-9; Pengpeng Ding y col. “Single-shot spectral-volumetric compressed ultrafast photography”. En: *Advanced Photonics* 3.4 (2021), págs. 1-6; Lai y col., “Compressed ultrafast tomographic imaging by passive spatiotemporal projections”; Yunhua Yao y col. “Single-Shot Real-Time Ultrafast Imaging of Femtosecond Laser Fabrication”. En: *ACS Photonics* 8.3 (2021),

págs. 738-744; Jiali Yao y col. “Total variation and block-matching 3d filtering-based image reconstruction for single-shot compressed ultrafast photography”. En: *Optics and Lasers in Engineering* 139 (2021), pág. 106475; Tatsuki Tahara y col. “Multidimensional digital holographic microscopy based on computational coherent superposition for coherent and incoherent light sensing”. En: *Holography, Diffractive Optics, and Applications X*. Vol. 11551. International Society for Optics y Photonics. 2020, pág. 115510M; WenLiang Zhu y col. “Architectural design of deep metallic sub-wavelength grating for practical holography display”. En: *AOPC 2017: Optoelectronics and Micro/Nano-Optics*. Vol. 10460. International Society for Optics y Photonics. 2017, 104601E; Chun-Wei Tsai y col. “Enhancing performance of LCoS-SLM as adaptive optics by using computer-generated holograms modulation software”. En: *Holography: Advances and Modern Trends V*. Vol. 10233. International Society for Optics y Photonics. 2017, pág. 102331D; Elisabet Pérez-Cabré y María Sagrario Millán. “Liquid crystal spatial light modulator with optimized phase modulation ranges to display multiorder diffractive elements”. En: *Applied Sciences* 9.13 (2019), pág. 2592; Baoqing Sun y col. “3D computational imaging with single-pixel detectors”. En: *Science* 340.6134 (2013), págs. 844-847, e.g., optical thermometry, photoluminescence lifetime, microscopy contrast. Therefore, opto-electronic coding elements (e.g. LCoS, DM, DMD and tunable lenses) have been preferred. A popular phase-based CDI framework is the focal stack system Xing Lin y col. “Coded focal stack photography”. En: *IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2013, págs. 1-9 that relies on a tunable focal length device synced with a DMD. The focal length tuning is modeled as a phase modulation (e.g., defocus aberration) via a 4f system with two lenses and a DM or LCoS located in the focal plane. Then, the phase-modulated wavefront is spatially encoded by the DMD and integrated by the sensor. Inspired by the phase coding CDI systems, recently proposed CSI systems employ either an LCoS Xing Lin y col. “Dual-coded com-

pressive hyperspectral imaging". En: *Optics Letters* 39.7 (2014), págs. 2044-2047 or a DM Marquez y col., "Compressive spectral imaging via deformable mirror and colored-mosaic detector" to better encode the spectral information and thus to improve the image reconstruction.

The positive impact of phase and amplitude coding devices in the CDI/ CSI frameworks and the fact that their propagation models and coding geometries share similarities, lead naturally to the question: ***Is it possible to develop a compressive spectral-depth imager (CoSDI) based on a hybrid amplitude-phase coding approach without relying on multiple sensors?***

To answer this question, this chapter proposes a compressive spectral depth imaging architecture that builds on the ideas of DFD, CASSI, and spectral-depth PSF-engineering relying on optoelectronic phase and amplitude encoding devices and a single sensor. Specifically, we sync the DM and the DMD – phase and amplitude coding devices– with the sensor exposure time. Here, the DM induces a focal length sweeping while introducing a controlled aberration. The resulting phase-modulated wavefront is spatially modulated and spectrally dispersed by a DMD and a prism. Finally, the resulting measurements –one per focal step– are integrated in the sensor into a single measurement. The latter is coined, coded aperture snapshot spectral-depth imager via depth-from-aberrations – CASSDI-DFA.

4.1. Continuous and discrete sensing model

The proposed CSDI system, coded aperture snapshot spectral depth imaging via depth from coded aberrations (CASSDI-DFA), is depicted in Fig. 33. It is comprised of a DM located at the pupil plane that is responsible for the focal sweep and coding additional aberrations, while the modulated image is passed through a dynamic coded aperture implemented by a DMD that is further dispersed by a prism and projected onto a detector array. Note that CASSDI-DFA is an optical system envisioned to be used in computer vision applications in outdoor environments, i.e., for incoherent illumination.

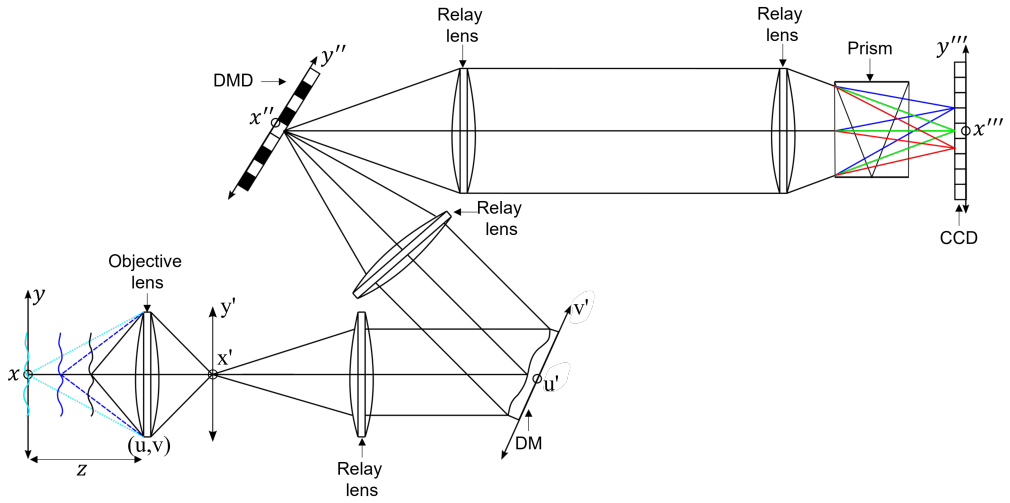
4.1.1. Continuous sensing model The mathematical generative model implemented by the CASSDI-DFA is described as follows. Formally, let $f_o(x, y, \lambda, z)$ represent the spatial-spectral-depth scene in intensity, where (x, y) index the spatial coordinates, λ index the wavelength and z the depth dimension. First, an objective lens forms an image in an intermediate image plane, resulting in the wave field

$$f_1(x', y', \lambda, z_i) = \iiint \gamma(\lambda) f_o(x, y, \lambda, z) h(x - x', y - y', \lambda, z_i, z) dx dy dz, \tag{50}$$

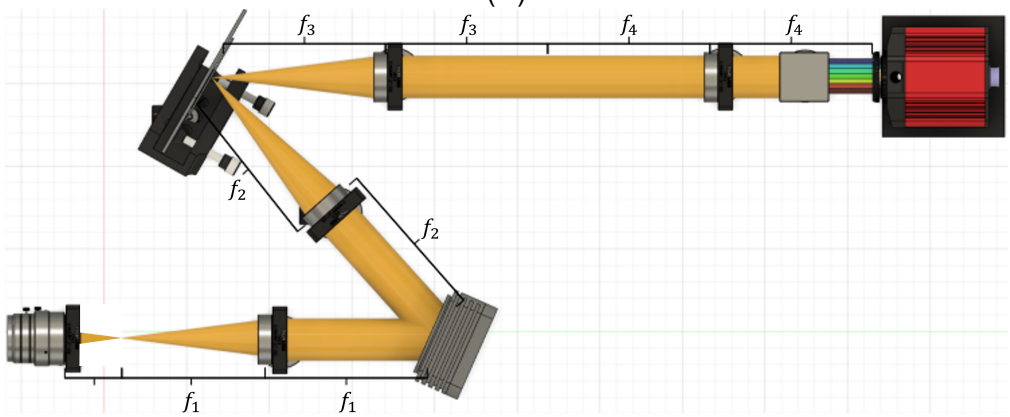
with

$$h(x', y', \lambda, z_i, z) \propto \left| \mathcal{F} \left\{ A(u, v) \text{exp} \left[j\Omega_{\lambda, z_i, z} \left(\frac{u^2 + v^2}{r^2} \right) \right] \right\} |_{\frac{x'}{\lambda z_i}, \frac{y'}{\lambda z_i}} \right|^2, \tag{51}$$

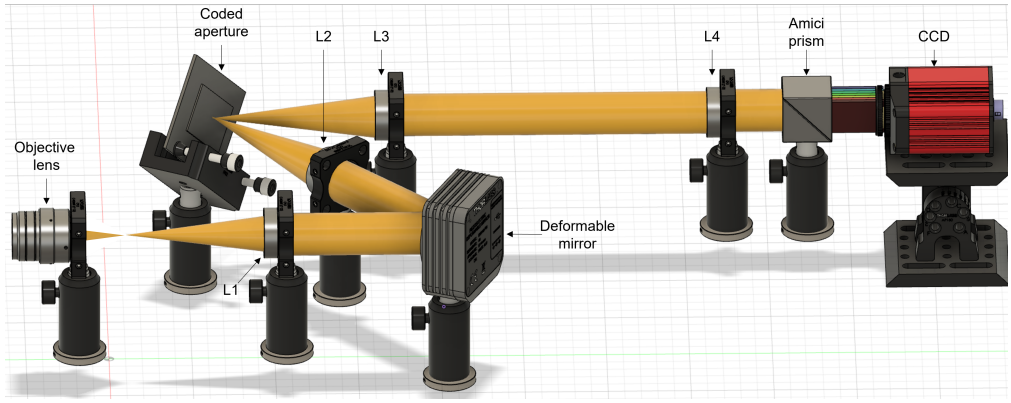
where h is the incoherent PSF of the objective lens, $\mathcal{F}(\cdot)$ is the Fourier transform operator, z_i represents the distance from the exit pupil plane to the image plane, A is the pupil function of the objective lens, $\Omega_{\lambda, z_i, z}$ is the defocus coefficient with $\Omega_{\lambda, z_i, z} = \frac{\pi}{\lambda} \left(\frac{1}{z} + \frac{1}{z_i} - \frac{1}{f_\lambda} \right) r^2$, f_λ is the wavelength dependent focal length, and r is the radius of the pupil; $\gamma(\lambda)$; $\gamma(\lambda)$ represent the spectral source density. Then, the



(A)



(B)



â—˜ (C)

Figura 33. Sketch of the proposed CASSDI-DFA system. Here, the deformable mirror's function is to sweep the depth planes and introduce a phase modulation. The coded aperture and the prism encode and disperse the spatial and spectral information, respectively, to be integrated by the sensor.

field $f_1(x', y', \lambda, z_i)$ is propagated through a 4f system that has a DM located at the pupil plane. The wavefront coding system is imaged onto a coded aperture located at the focal plane. Therefore, the spectral density immediately after the coded aperture plane can be expressed as

$$f_2(x'', y'') = \int k(x'', y'', z') \iint f_1(x', y', z') h_M(x' - x'', y' - y'', \lambda, z') dx' dy' dz', \quad (52)$$

with

$$h_M(x'', y'', \lambda, z') \propto \left| \mathcal{F}\{\mathcal{P}(u', v') \text{exp}\left[j2\pi \mathcal{W}(u', v')\right] \text{exp}\left[j\Omega'_{\lambda, z_i} \left(\frac{u'^2 + v'^2}{r^2} \right) \right] \} |_{\frac{x'}{\lambda z'}, \frac{y'}{\lambda z'}} \right|^2, \quad (53)$$

where $h_M(x'', y'', \lambda, z', z_i)$ is the incoherent PSF introduced by the 4f system, Ω'_{λ, z_i} is the defocus coefficient with $\Omega'_{\lambda, z_i} = \frac{\pi}{\lambda} \left(\frac{1}{z'} + \frac{1}{d - z_i} - \frac{1}{f'_\lambda} \right) r'^2$, $\mathcal{P}$ is the pupil function of the 4f system, f'_λ is the 4f system's wavelength-dependent focal length, r' is the radius of the pupil, d represents the length from the objective lens to the first relay lens, z' represents the distance from the exit pupil plane to the DMD (intermediate image plane), and $k(x'', y'', z_i)$ represent the coded aperture, and $\mathcal{W}(u', v')$ is the wavefront aberration function introduced by the deformable mirror. Particularly, the coded aperture is modeled as

$$k(x'', y'', z_i) = \sum_{i_x, i_y} C_{i_x, i_y, i_z} \text{rect}\left(\frac{x''}{\Delta_c} - i_x, \frac{y''}{\Delta_c} - i_y, \frac{z_i}{\Delta_z} - i_z \right), \quad (54)$$

where $C_{i_x, i_y, i_z} \in \{0, 1\}$ is the coding performed on the $(i_x, i_y)^{th}$ voxel at depth $\lfloor z'/\Delta_z \rfloor \in \mathbb{Z}$; Δ_c, and Δ_d account for the pixel sizes of the CA, with $i_x = \{0, \ldots, N_x - 1\}$, $i_y = \{0, \ldots, N_y - 1\}$ indexing the rows and columns inside each z depth, and $i_z = \{0, \ldots, N_z - 1\}$ indexing the depth dimension. Additionally, the wavefront can be expressed as $W(u', v') = \sum_{k=1}^{\infty} a_k Z_k(u', v')$, where $Z_k(u', v')$ represents the k-th Zernike polynomial (ZP) in the Noll's notation and a_k its amplitude coefficient Robert J

Noll. “Zernike polynomials and atmospheric turbulence”. En: *JOSA A* 66.3 (1976), págs. 207-211. Then, the spatially modulated wavefront propagated through the dispersive element is spectrally decomposed following a wavelength-dependent horizontal shifting $S(\lambda)$ Wagadarikar y col., “Single disperser design for coded aperture snapshot spectral imaging” (please refer to supplementary material Eq. A.20), such that the density can be seen in the detector plane (immediately before being integrated by the sensor) as

$$f_3(x''', y''', \lambda) = \iint f_2(x'', y'', \lambda) h_P(x''' - x'' - S(\lambda), y''' - y'') dx'' dy'', \tag{55}$$

where h_P accounts is the incoherent PSF introduced by the Double Amici prism. Finally, the measurement at the $(i_1, i_2')^{th}$ pixel is represented by

$$g_{i_1, i_2'} = \int_\Lambda \rho(\lambda) \iint f_3(x''', y''', \lambda) p(i_1, i_2'; x''', y''') dx''' dy''' d\lambda, \tag{56}$$

where Λ represents the wavelength axis over the spectral range, p is the pixel sampling function with $p(i_1, i_2'; x''', y''') = \text{rect}\left(\frac{x'''}{\Delta_d} - i_1, \frac{y'''}{\Delta_d} - i_2'\right)$ represents, Δ_d represents the pixel's side length, and $\rho(\lambda)$ represents the detector's normalized quantum efficiency.

4.1.2. Discrete sensing model Let's $\mathbf{C}$ defined as the discrete version of the coded aperture with $\mathbf{C} \in \mathbb{R}^{N_x \times N_y \times N_z}$. Hence, based on $\mathbf{C}$ and Eq. (56), the discretized version of the compressed measurement can be expressed as

$$\mathbf{G} = \sum_{i_z=1}^{N_z} \sum_{i_\lambda=1}^{N_\lambda} ([[\mathbf{F}_{:,:,i_\lambda,i_z} * \mathbf{H}_{:,:,i_\lambda,i_z}] * (\mathbf{H}_M)_{:,:,i_\lambda,i_z}] \circ \mathbf{C}_{:,:,i_z}) * (\mathbf{H}_P)_{:,:,i_\lambda} + \mathbf{E}, \tag{57}$$

where $\mathbf{G} \in \mathbb{R}^{N_x \times (N_y + N_\lambda - 1)}$ is the discrete compressed measurement, $\mathbf{E} \in \mathbb{R}^{N_x \times (N_y + N_\lambda - 1)}$ represents the noise, $\mathbf{H} \in \mathbb{R}^{N_x \times N_y \times N_\lambda \times N_z}$ represents the discrete intensity PSF intro-

duced by the objective lens, $\mathbf{H}_M \in \mathbb{R}^{N_x \times N_y \times N_\lambda \times N_z}$ represents the discrete intensity PSF introduced by the 4f system, and $\mathbf{H}_P \in \mathbb{R}^{N_x \times N_y \times N_\lambda}$ represents the discrete intensity PSF introduced by the prism. In computer vision for the design of recovery algorithms, a widespread approximation is to neglect the lens aberrations by assuming their intensity PSF as delta Dirac. Although this assumption is physically unrealistic, this simple model suffices for several computer vision and commercial applications. Additionally, modeling the noise variable **E** as a colored Gaussian noise distribution has been widely accepted in high-level computer vision tasks, e.g., reconstruction, recognition, classification, detection, and understanding. Because the scope of this thesis is limited to reconstruction tasks and based on the plethora of state-of-the-art computational reconstruction algorithms, in this thesis, the noise is approximated by a Gaussian distribution.

Based on these approximations, Eq. (57) can be rewritten as

$$G_{i_x,i'_y} = \sum_{i_z=1}^{N_z} \sum_{i_\lambda=1}^{N_\lambda} \tilde{\mathbf{F}}_{:,(i'_y-i_\lambda),} \circ \mathbf{C}_{:,(i'_y-i_\lambda),i_z} + \mathbf{E}, \tag{58}$$

with $\tilde{\mathbf{F}}_{:,:,i_\lambda,i_z} = \mathbf{F}_{:,:,i_\lambda,i_z} * (\mathbf{H}_M)_{:,:,i_\lambda,i_z}$, where $\mathbf{G} \in \mathbb{R}^{N_x \times (N_y+N_\lambda-1)}$ is the discrete compressed measurement, $\mathbf{H}_M \in \mathbb{R}^{N_x \times N_y \times N_z}$ is the discrete point-spread-function introduced by the DM, and $\mathbf{F} \in \mathbb{R}^{N_x \times N_y \times N_\lambda \times N_z}$ is the discrete spatio-spectral-depth source. Note that the compressed measurement in Eq. (58) yields a compression ratio of $\frac{(N_y+N_\lambda-1)}{N_y N_\lambda N_z}$, where the spectral and depth dimension are projected onto a 2D image. Here, note that the spectral and depth-sensing model is inspired in the CASSI sensing geometry Wagadarikar y col., “Single disperser design for coded aperture snapshot spectral imaging” and the depth-dependent image formation Chang y Wetzstein, “Deep Optics for Monocular Depth Estimation and 3D Object Detection”, respectively. Based on Eq. (58) , the proposed optical system can be represented as

a linear system of the form,

$$\mathbf{g} = \mathbf{H}\mathbf{\Phi}\mathbf{f} + \boldsymbol{\epsilon}, \tag{59}$$

where $\mathbf{g} \in \mathbb{R}^{m\times 1}$ represents the vectorized version of $\mathbf{G}$ with $m = N_x(N_y + N_\lambda - 1)$, $\mathbf{f} \in \mathbb{R}^{n\times 1}$ represents the vectorized version of $\mathbf{F}$ with $n = N_x N_y N_\lambda N_z$, $\mathbf{\Phi} \in \mathbb{R}^{n\times n}$ is the matrix that models the aberrations introduced by the deformable mirror, $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the sensing matrix that models the spatial codification and spectral dispersion, and $\boldsymbol{\epsilon} \in \mathbb{R}^{m\times 1}$ represents the noise. Note that, in Eq. (59) the defocus effect is encoded as an intrinsic feature of the high dimensional object $\mathbf{f}$, i.e., the spatially varying intensity PSF associated with the depth information is an unknown variable. Specifically, the entries of the $\mathbf{H}$ are given by

$$H_{i,j} = \begin{cases} c_u, & if\ i = mod(j, N_xN_y) + N_x\lfloor \frac{mod(j,N_xN_yN_\lambda)}{N_xN_y} \rfloor \\ 0, & otherwise \end{cases}, \tag{60}$$

for $i = \{0, ..., m-1\}$, $j = \{0, ..., n-1\}$, $u = mod(j, N_xN_y) + N_xN_y\lfloor \frac{j}{N_xN_yN_\lambda} \rfloor$, and c_u are the entries of $\mathbf{c} \in \mathbb{R}^{N_xN_yN_z\times 1}$, which is the vectorized version of $\mathbf{C}$. A conventional approach to obtain an estimation of $\mathbf{f}$ from its compressed measurement $\mathbf{g}$, considering Eq. (59) and the use of sparsity promoting priors, is given by

$$\underset{\boldsymbol{\theta}}{\arg\min} \quad \|\mathbf{g} - \mathbf{H}\mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta}\|_2^2 + \tau\|\boldsymbol{\theta}\|_1, \tag{61}$$

where $\|\cdot\|_1$ represents the ℓ_1-norm, $\boldsymbol{\theta} \in \mathbb{R}^{n\times 1}$ is a sparse representation of $\mathbf{f}$ in the orthonormal basis $\mathbf{\Psi} \in \mathbb{R}^{n\times n}$ with $\mathbf{\Psi}^T\mathbf{\Psi} = \mathbf{I}$, and $\tau \in \mathbb{R}_+$ is a regularization parameter. Here it is important to note that although the optimization problem established in Eq. (61) allows to obtain an approximation of the sparse high dimensional data cube $\boldsymbol{\theta}$, it requires high computational complexity, which is bounded by $O(n^3)$.

4.2. Reconstruction algorithm

Conventional reconstruction algorithms based on multiresolution approaches aim to rewrite a high-complexity optimization problem into a set of low-complexity subproblems. We build on this concept to propose a multiresolution spectral-depth reconstruction methodology, splitting the high dimensional optimization problem into three low-complexity problems that allow estimating an all-in-focus grayscale $\mathbf{f}_g \in \mathbb{R}^{n_g \times 1}$ with $n_g = N_x N_y$, an all-in-focus spectral $\mathbf{f}_s \in \mathbb{R}^{n_s \times 1}$ with $n_s = N_x N_y N_\lambda$, and a grayscale focal stack version $\mathbf{f}_f \in \mathbb{R}^{n_f \times 1}$ with $n_f = N_x N_y N_z$, respectively, from a single compressed measurement $\mathbf{g}$. Having estimated the focal-stack grayscale image $\mathbf{f}_s$, these are used as an input for a pre-trained U-net neural network Olaf Ronneberger, Philipp Fischer y Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". En: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, págs. 234-241 to estimate the depth map information $\mathbf{f}_d \in \mathbb{R}^{n_d \times 1}$ with $n_d = n_g$, as shown Fig. 34. Particularly, U-Net is a convolutional neural network that was developed for biomedical image segmentation. The network consists of a contracting path and an expansive path, which gives it the u-shaped architecture. The contracting path is a typical convolutional network that consists of repeated application of convolutions, each followed by a rectified linear unit (ReLU) and a max pooling operation. During the contraction, the spatial information is reduced while feature information is recovered. The expansive pathway combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path.

4.2.1. Reconstructing $\mathbf{f}_g$ Mathematically, $\mathbf{f}_g$ is defined as $\mathbf{f}_g \approx \mathbf{D}_g \mathbf{f}$, where $\mathbf{D}_g = \mathbf{1}_{N_z}^T \otimes \left[\mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{n_g \times n_g}\right]$ is a depth decimator matrix with $\mathbf{D}_g \in \mathbb{R}^{n_g \times n}$. Then replacing $\mathbf{f}_g$ in Eq. (59), it can be obtained the equivalent sensing model for the all-in-focus

grayscale version as

$$\mathbf{g} = \mathbf{H}\mathbf{\Phi}\mathbf{D}_g^T\mathbf{f}_g + \boldsymbol{\epsilon}_g, \tag{62}$$

where $\boldsymbol{\epsilon}_g = \mathbf{H}\mathbf{\Phi}\left[\mathbf{D}_g^T\mathbf{D}_g - \mathbf{I}\right]^{-1}\mathbf{f} + \boldsymbol{\epsilon}$. Note that Eq. (62) can be solved following an $\ell_2 - \ell_1$-norm approach, where the ℓ_2 model is only optimal when the noise is colored Gaussian, and it tends to over-smooth the image details. In contrast, the ℓ_1-norm approach effectively preserves the edges in the image Huanfeng Shen y col. "Adaptive norm selection for regularized image restoration and super-resolution". En: *IEEE Transactions on Cybernetics* 46.6 (2015), págs. 1388-1399. More precisely, $\mathbf{f}_g$ can be estimated via

$$\begin{aligned} \arg\min_{\mathbf{f}_g,\boldsymbol{\theta}_g,\mathbf{w}} \quad & \tfrac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{\Phi}\mathbf{D}_g^T\mathbf{f}_g\|_2^2 + \tau_w\|\mathbf{w}\|_1 + \tau_g\|\boldsymbol{\theta}_g\|_1, \\ \text{subject to} \quad & \mathbf{w} = \mathcal{T}\mathbf{f}_g,\ \boldsymbol{\theta}_g = \mathbf{\Psi}_g\mathbf{f}_g \end{aligned} \tag{63}$$

for $\|\mathbf{f}_g\|_{TV} = \|\mathcal{T}\mathbf{f}_g\|_1$, where $\mathcal{T}$ is an operator matrix that computes the first-order finite differences of the neighboring features across horizontal/vertical directions, $\mathcal{T}^T\mathcal{T} \approx \mathbf{I}$, $\tau_g \in \mathbb{R}_+$ is a regularization parameter, $\mathbf{\Psi}_g \in \mathbb{R}^{n_g \times n_g}$ is an orthonormal representation basis with $\mathbf{\Psi}_g^T\mathbf{\Psi}_g = \mathbf{I}_{n_g \times n_g}$, and $\boldsymbol{\theta}_g \in \mathbb{R}^{n_g}$ is a sparse representation of $\mathbf{f}_g$ in the basis $\mathbf{\Psi}_g$. Following an ADMM methodology, an alternative form of Eq. (63) can be expressed as

$$\begin{aligned} \arg\min_{\mathbf{f}_g,\mathbf{w},\boldsymbol{\theta}_g} \quad & \tfrac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{\Phi}\mathbf{D}_g^T\mathbf{f}_g\|_2^2 + \alpha_1\|\mathbf{w} - \mathcal{T}\mathbf{f}_g - \boldsymbol{\nu}_1^\iota\|_2^2 + \alpha_2\|\boldsymbol{\theta}_g - \mathbf{\Psi}_g\mathbf{f}_g - \boldsymbol{\nu}_2^\iota\|_2^2 \\ & + \tau_w\|\mathbf{w}\|_1 + \tau_g\|\boldsymbol{\theta}_g\|_1, \end{aligned} \tag{64}$$

where $\{\boldsymbol{\nu}_1, \boldsymbol{\nu}_2\} \in \mathbb{R}^{n_g \times n_g}$ are scaled dual variables with $\boldsymbol{\nu}_1^{\iota+1} = \boldsymbol{\nu}_1^\iota - \left(\mathbf{w}^{\iota+1} - \mathcal{T}\mathbf{f}_g^{\iota+1}\right)$, and $\boldsymbol{\nu}_2^{\iota+1} = \boldsymbol{\nu}_2^\iota - (\boldsymbol{\theta}_g^{\iota+1} - \mathbf{\Psi}_g\mathbf{f}_g^{\iota+1})$, and $\{\alpha_1, \alpha_2\} > 0$ terms are the weights of the augmented Lagrangian term. The method to solve Eq. (63) via Eq. (64) is summarized

in **Algorithm 2**. In summary, the computational complexity of the all-in-focus grayscale image estimation Algorithm is $O(2n^2 n_g + n_g^3 + n_g^2(9+n) + n_g(15+2mn))$, and its computational complexity is bounded by $O(n_g^3)$.

Algorithm 2 All-in-focus grayscale algorithm $\mathbf{f}_g$

Input:$\mathbf{g}, \mathbf{f}_g, \alpha_1, \alpha_2, \tau_2, \tau_g$
Result: $\mathbf{f}_g$
1. $\{\boldsymbol{\nu}_1^0, \boldsymbol{\nu}_2^0, \mathbf{w}^0, \boldsymbol{\theta}_g^0\} \leftarrow \mathbf{0} \in \mathbb{R}^{n_g \times 1}$
for $\iota = 0$ *to* $Iter$ **do**
2. $\mathbf{A} = \left[\mathbf{D}_g \boldsymbol{\Phi}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\Phi} \mathbf{D}_g^T + (\alpha_1 + \alpha_2)\mathbf{I}\right]^{-1}$
3. $\mathbf{f}_g^{\iota+1} = \mathbf{A} \cdot \left[\mathbf{D}_g \boldsymbol{\Phi}^T \mathbf{H}^T \mathbf{g} + \alpha_1 \mathcal{T}(\mathbf{w}^\iota - \boldsymbol{\nu}_1^\iota) + \alpha_2 \boldsymbol{\Psi}_g^T (\boldsymbol{\theta}_g^\iota - \boldsymbol{\nu}_2^\iota)\right]$
4. $\mathbf{w}_1^{\iota+1} = \text{soft}\left(\mathcal{T}\mathbf{f}_g^{\iota+1} + \boldsymbol{\nu}_1^\iota, \tau_w/\alpha_1\right)$
5. $\boldsymbol{\theta}_g^{\iota+1} = \text{soft}\left(\boldsymbol{\Psi}_g \mathbf{f}_g^{\iota+1} + \boldsymbol{\nu}_2^\iota, \tau_g/\alpha_2\right)$
6. $\boldsymbol{\nu}_1^{\iota+1} = \boldsymbol{\nu}_1^\iota - \left(\mathbf{w}^{\iota+1} - \mathcal{T}\mathbf{f}_g^{\iota+1}\right)$
7. $\boldsymbol{\nu}_2^{\iota+1} = \boldsymbol{\nu}_2^\iota - \left(\boldsymbol{\theta}_g^{\iota+1} - \boldsymbol{\Psi}_g \mathbf{f}_g^{\iota+1}\right)$
end

4.2.2. Reconstructing $\mathbf{f}_f$ Similarly to Eq. (62), $\mathbf{f}_f$ can be directly related to $\mathbf{f}$ as $\mathbf{f}_f = \mathbf{D}_f \mathbf{f}$, where $\mathbf{D}_f = \mathbf{I}_{N_z \times N_z} \otimes \left[\mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{n_g \times n_g}\right]$ is a spectral decimator matrix with $\mathbf{D}_f \in \mathbb{R}^{n_f \times n}$. Then replacing $\mathbf{f}_f$ in Eq. (59), we can obtain the equivalent sensing model for the focus stack grayscale version as

$$\mathbf{g} = \mathbf{H}\boldsymbol{\Phi}\mathbf{D}_f^T \mathbf{f}_f + \boldsymbol{\epsilon}_f, \tag{65}$$

where $\boldsymbol{\epsilon}_f = \mathbf{H}\boldsymbol{\Phi}\left[\mathbf{D}_f^T \mathbf{D}_f - \mathbf{I}\right]^{-1} \mathbf{f} + \boldsymbol{\epsilon}$. Having calculated the all-in-focus grayscale version $\mathbf{f}_g$ from $\mathbf{g}$ in Eq. (63), and relating $\mathbf{f}_f$ to $\mathbf{f}$ in Eq. (62), an estimation of the focal stack grayscale version $\mathbf{f}_f$ from $\mathbf{g}$ and $\mathbf{f}_g$ can be obtained. Moreover, to limit the solution space and exploit the spatial correlation, we establish a ℓ_2 fidelity function based on the all-in-focus grayscale version $\mathbf{f}_g$. More precisely, $\mathbf{f}_f$ can be estimated

via

$$\begin{aligned} \arg\min_{\mathbf{f}_f, \boldsymbol{\theta}_f} \quad & \tfrac{1}{2}\|\mathbf{g} - \mathbf{H}\boldsymbol{\Phi}\mathbf{D}_f^T\mathbf{f}_f\|_2^2 + \tfrac{\sigma_f}{2}\|\mathbf{f}_g - \mathbf{B}_f\mathbf{f}_f\|_2^2 + \tau_f\|\boldsymbol{\theta}_f\|_1, \\ \text{subject to} \quad & \boldsymbol{\theta}_f = \boldsymbol{\Psi}_f\mathbf{f}_f \end{aligned} \tag{66}$$

where $\tau_f \in \mathbb{R}_+$ is a regularization parameter, $\mathbf{B}_f \in \mathbb{R}^{n_g \times n_f}$ is a depth decimator matrix with $\mathbf{B}_f = \mathbf{1}_{N_z}^T \otimes \mathbf{I}_{n_g}$, and $\boldsymbol{\Psi}_f \in \mathbb{R}^{n_f \times n_f}$ is an orthonormal representation basis with $\boldsymbol{\Psi}_f^T\boldsymbol{\Psi}_f = \mathbf{I}_{n_f \times n_f}$. Similar to Eq. (64), Eq. (63) can be solved by

$$\begin{aligned} \arg\min_{\mathbf{f}_f, \boldsymbol{\theta}_f} \quad & \tfrac{1}{2}\|\mathbf{g} - \mathbf{H}\boldsymbol{\Phi}\mathbf{D}_f^T\mathbf{f}_f\|_2^2 + \tfrac{\sigma_f}{2}\|\mathbf{f}_g - \mathbf{B}_f\mathbf{f}_f\|_2^2 + \alpha\|\boldsymbol{\theta}_f - \boldsymbol{\Psi}_f\mathbf{f}_f - \boldsymbol{\nu}\|_2^2 \\ & + \tau_f\|\boldsymbol{\theta}_f\|_1, \end{aligned} \tag{67}$$

where $\boldsymbol{\nu} \in \mathbb{R}^{n_f}$ is a scaled dual variable with $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^{\iota} - \left(\boldsymbol{\theta}_f^{\iota+1} - \boldsymbol{\Psi}_f\mathbf{f}_f^{\iota+1}\right)$ and $\alpha > 0$ is the weighting of the augmented Lagrangian term. The method to solve Eq. (66) via Eq. (67) is summarized in **Algorithm 3**. In summary, the computational complexity of the focal-stack grayscale image estimation algorithm is $O(n_f^3 + n^2(2n_f + 1) + n_f^2(n + n_g + 8) + n_f(n + n_g + 2mn + 11) + mn)$, and its computational complexity is bounded by $O(n_f^3)$.

Algorithm 3 Grayscale focal stack algorithm $\mathbf{f}_f$

Input: $\mathbf{g}, \alpha, \sigma, \tau_f$
Result: $\mathbf{f}_f$
1. $\{\boldsymbol{\nu}^0, \boldsymbol{\theta}_f^0\} \leftarrow \mathbf{0} \in \mathbb{R}^{n_f \times 1}$
2. $\mathbf{A} = \left[\mathbf{D}_f\boldsymbol{\Phi}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\Phi}\mathbf{D}_f^T + \sigma\mathbf{B}_f^T\mathbf{B}_f + \alpha\mathbf{I}\right]^{-1}$
for $\iota = 0$ *to* $Iter$ **do**
  3. $\mathbf{f}_f^{\iota+1} = \mathbf{A} \cdot \left[\mathbf{D}_f\boldsymbol{\Phi}^T\mathbf{H}^T\mathbf{g} + \sigma\mathbf{B}^T\mathbf{f}_g + \alpha\boldsymbol{\Psi}_f^T\left(\boldsymbol{\theta}_f^{\iota} - \boldsymbol{\nu}^{\iota}\right)\right]$
  4. $\boldsymbol{\theta}_f^{\iota+1} = \text{soft}\left(\boldsymbol{\Psi}_f\mathbf{f}_f^{\iota+1} + \boldsymbol{\nu}^{\iota}, \tau_f/\alpha\right)$
  5. $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^{\iota} - \left(\boldsymbol{\theta}_f^{\iota+1} - \boldsymbol{\Psi}_f\mathbf{f}_f^{\iota+1}\right)$
end

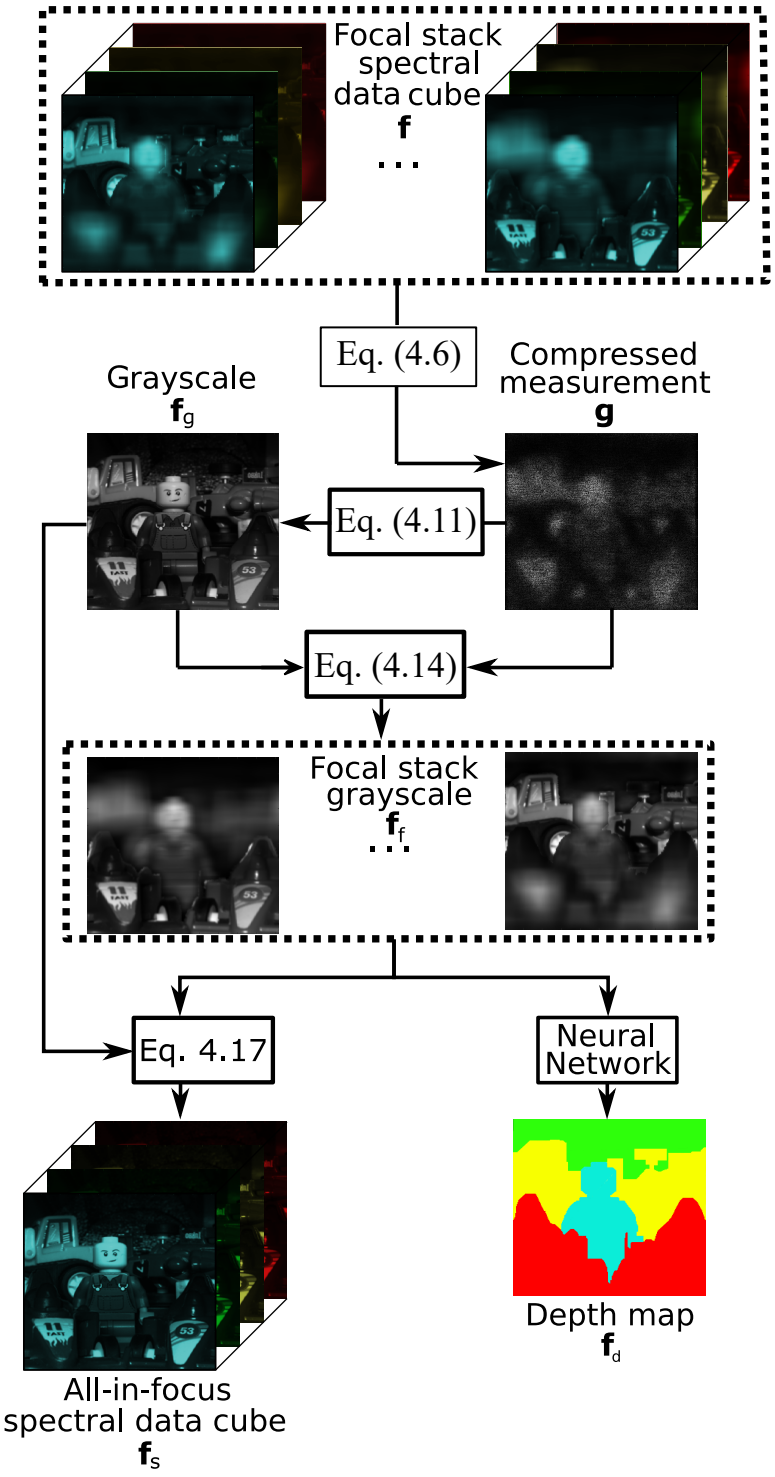


Figura 34. Structure of spectral-depth image reconstruction pipeline

4.2.3. Reconstructing $\mathbf{f}_s$ Finally, $\mathbf{f}_s$ can be defined as $\mathbf{f}_s = \mathbf{D}_s\mathbf{f}$, where $\mathbf{D}_s = \left[\mathbf{1}_{N_z}^T \otimes \mathbf{I}_{n_s \times n_s}\right]$ is a depth decimator matrix with $\mathbf{D}_s \in \mathbb{R}^{n_s \times n}$. Then replacing $\mathbf{f}_s$ in Eq. (59), we can obtain the equivalent sensing model for the all-in-focus spectral data cube as

$$\mathbf{g} = \mathbf{H}\mathbf{\Phi}_s\mathbf{D}_s^T\mathbf{f}_s + \boldsymbol{\epsilon}_s, \tag{68}$$

where $\boldsymbol{\epsilon}_s = \mathbf{H}\mathbf{\Phi}_s\left[\mathbf{D}_s^T\mathbf{D}_s - \mathbf{I}\right]^{-1}\mathbf{f} + \boldsymbol{\epsilon}$, and $\mathbf{\Phi} \in \mathbb{R}^{n \times n}$ is a matrix that models the aberrations introduced by the deformable mirror along with the associated defocus aberration. Specifically, this defocus aberration is obtained from the deformable mirror focal sweeping prior information and the neuronal network's depth map estimation. To improve the spatial quality in the reconstruction of the all-in-focus spectral image, two ℓ_2 fidelity norms were included based on the all-in-focus grayscale $\mathbf{f}_g$ and the focal stack grayscale versions $\mathbf{f}_f$. Then, the optimization problem to estimate the all-in-focus spectral data cube is mathematically expressed as

$$\begin{aligned} \underset{\mathbf{f}_s, \boldsymbol{\theta}_s}{\arg\min} \quad & \|\mathbf{g} - \mathbf{H}\mathbf{\Phi}_s\mathbf{D}_s^T\mathbf{f}_s\|_2^2 + \|\mathbf{B}_f\mathbf{f}_f - \mathbf{B}_s\mathbf{f}_s\|_2^2 + \|\mathbf{f}_g - \mathbf{B}_s\mathbf{f}_s\|_2^2 + \tau_s\|\boldsymbol{\theta}_s\|_1, \\ \text{subject to} \quad & \boldsymbol{\theta}_s = \mathbf{\Psi}_s\mathbf{f}_s \end{aligned} \tag{69}$$

where $\tau_s \in \mathbb{R}_+$ is a regularization parameter, $\mathbf{B}_s \in \mathbb{R}^{n_g \times n_s}$ is a depth decimator matrix with $\mathbf{B}_s = \mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{n_g}$, and $\mathbf{\Psi}_s \in \mathbb{R}^{n_s \times n_s}$ is an orthonormal representation basis with $\mathbf{\Psi}_s^T\mathbf{\Psi}_s = \mathbf{I}_{n_s \times n_s}$. Similar to Eq. (64) and Eq. (67), an ADMM methodology is used to solve Eq. (69) starting with the calculation of the augmented Lagrangian as

$$\begin{aligned} \underset{\mathbf{f}_s, \boldsymbol{\theta}_s}{\arg\min} \quad & \tfrac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{\Phi}_s\mathbf{D}_s^T\mathbf{f}_s\|_2^2 + \alpha_1\|\mathbf{B}_f\mathbf{f}_f - \mathbf{B}_s\mathbf{f}_s\|_2^2 + \alpha_2\|\mathbf{f}_g - \mathbf{B}_s\mathbf{f}_s\|_2^2 + \tau_s\|\boldsymbol{\theta}_s\|_1 \\ & + \alpha_3\|\boldsymbol{\theta}_s - \mathbf{\Psi}_s\mathbf{f}_s - \boldsymbol{\nu}\|_2^2, \end{aligned} \tag{70}$$

where $\boldsymbol{\nu} \in \mathbb{R}^{n_s}$ is a scaled dual variable with $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^{\iota} - \left(\boldsymbol{\theta}_s^{\iota+1} - \boldsymbol{\Psi}_f \mathbf{f}_s^{\iota+1}\right)$, and $\alpha_3 > 0$ is the weighting of the augmented Lagrangian term. The method to solve Eq. (69) via Eq. (70) is summarized in **Algorithm 4**. In summary, the computational complexity of the focal-stack grayscale image estimation algorithm is $O(n_s^3 + n^2(2n_s+1) + n_s^2(n+3) + nm(2n_s+1) + n_g(n_f+3) + 4n_s)$, and its computational complexity is bounded by $O(n_s^3)$. In general, the computational complexity of the proposed reconstruction methodology (illustrated in Fig. 4) for a 4-dimensional spatial-spectral-depth image is bounded by $O\left((N_x N_y \cdot \max\left(N_\lambda, N_z\right))^3\right)$ in contrast to Eq. (61) which is bounded by $O((N_x N_y N_\lambda N_z)^3)$. Therefore, the complexity ratio between Eq. (61) and the proposed reconstruction methodology for a DSI with $N_x = N_y = N_\lambda = N_z = \sqrt[4]{n}$ is $O\left(\frac{1}{\sqrt[4]{n^3}}\right)$.

Algorithm 4 All-in-focus spectral algorithm $\mathbf{f}_s$

Input: $\mathbf{f}_g, \mathbf{f}_f, \mathbf{g}, \alpha_1, \alpha_2, \alpha_3, \tau_s$
Result: $\mathbf{f}_s$
1. $\{\boldsymbol{\nu}^0, \boldsymbol{\theta}_s^0\} \leftarrow \mathbf{0} \in \mathbb{R}^{n_s \times 1}$
2. $\mathbf{A} = \left[\mathbf{D}_s \boldsymbol{\Phi}_s^T \mathbf{H}^T \mathbf{H} \boldsymbol{\Phi}_s \mathbf{D}_s^T + (\alpha_1 + \alpha_2)\mathbf{B}_s^T \mathbf{B}_s + \alpha_3 \mathbf{I}\right]^{-1}$
for $\iota = 0$ *to* $Iter$ **do**
   3. $\mathbf{f}_s^{\iota+1} = \mathbf{A} \cdot \left[\mathbf{D}_s \boldsymbol{\Phi}_s^T \mathbf{H}^T \mathbf{g} + \mathbf{B}_s^T(\alpha_1 \mathbf{B}_f \mathbf{f}_f + \alpha_2 \mathbf{f}_s) + \alpha_3 \boldsymbol{\Psi}_s^T(\boldsymbol{\theta}_s^{\iota} - \boldsymbol{\nu}^{\iota})\right]$
   4. $\boldsymbol{\theta}_s^{\iota+1} = \text{soft}\left(\boldsymbol{\Psi}_s \mathbf{f}_s^{\iota+1} + \boldsymbol{\nu}^{\iota}, \tau_s/\alpha_3\right)$
   5. $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^{\iota} - \left(\boldsymbol{\theta}_s^{\iota+1} - \boldsymbol{\Psi}_s \mathbf{f}_s^{\iota+1}\right)$
end

4.3. Simulation results

The performance of the proposed optical system and reconstruction method is firstly evaluated by simulated compressed measurements using the model described in Eq. (59), and reconstructing them using the proposed ADMM-based reconstruction methodology. For these simulations, the public dataset Nathan Silberman y col. “Indoor segmentation and support inference from rgbd images”. En: *European Conference on Computer Vision*. Springer. 2012, págs. 746-760, and the hierarchical regression network for spectral reconstruction from RGB images Yuzhi Zhao y col.

"Hierarchical Regression Network for Spectral Reconstruction from RGB Images". En: (2020), págs. 422-423 were used to generate a synthetic spectral depth dataset. Then, these spectral images were resized to have spatial and spectral dimensions $N_x \times N_y = 512 \times 512$ and $N_\lambda = 12$, respectively. We analyze the system performance by considering three main aspects. Firstly, the advantage of multiplexing the spectral-depth information is evaluated via a sequential phase-amplitude modulation approach. Secondly, the phase mask structure impacts the rendering fidelity of the 4D information is tested by varying the pure Zernike polynomials and their amplitude coefficients. Third, the impact of introducing controlled aberrations in the focal length sweeping step is evaluated by comparing the proposed CASSDI-DFA system with a depth-from-defocus CASSI system, which is a variation of the proposed CASSDI-DFA system but setting the amplitude coefficient to zero ($a_j = 0$). The depth-from-defocus CASSI approach is simulated by sweeping the focal lengths using pure defocus (Z_4). The focal stack images are solely coded by the CASSI system and then integrated into a single frame. This particular sensing case is called CASSDI-DFD, and its sampling protocol is analogous to using a varifocal lens instead of a DM.

The measurements are simulated using a set of random black-and-white coded apertures with $50\,\%$ transmittance. Finally, a convolutional U-net neural network is trained to obtain a fast and precise depth map segmentation from the focal stack grayscale estimation. Specifically, the network consists of 5 downsampling layers (Conv-BN-ReLU$\times$2$\rightarrow$MaxPool2$\times$2) followed by 5 upsampling layers with skip connections(ConvT + Concat$\rightarrow$Conv-BN-ReLU$\times$2). The output is the predicted depth map at the same spatial resolution as the input image. We use the standard ADAM optimizer with a mean-square-error (MSE) loss on the logarithmic depth. We train the models for 100,000 iterations at a learning rate of 0.0001 and a batch size of 4. The sparse promoting bases are set to be $\mathbf{\Psi}_w = \mathbf{\Psi}_{1D-W} \otimes \mathbf{\Psi}_{1D-W}$, $\mathbf{\Psi}_z = \mathbf{\Psi}_{1D-DCT} \otimes \mathbf{\Psi}_w$, and $\mathbf{\Psi}_f = \mathbf{\Psi}'_{1D-DCT} \otimes \mathbf{\Psi}_w$, where $\mathbf{\Psi}_{1D-W} \in \mathbb{R}^{N_x \times N_x}$- and $\mathbf{\Psi}_{1D-DCT} \in \mathbb{R}^{N_\lambda \times N_\lambda}$-

$\mathbf{\Psi}'_{1D-DCT} \in \mathbb{R}^{N_z \times N_z}$ represent the 1D Wavelet (Symlet 8) basis, and the 1D discrete cosine transform Simon Foucart y Holger Rauhut. “An invitation to compressive sensing”. En: *A mathematical introduction to compressive sensing*. Springer, 2013, págs. 1-39, respectively. All simulations were conducted and timed using an Intel Core i7 3960X 3.30 GHz processor with 32 GB of RAM. To compare the quality of the reconstructions, we use the root mean squared error (RMSE), the spectral angular mapper (SAM), and the structural similarity (SSIM) metrics. Specifically, the SSIM metrics are calculated band-per-band and averaged, the RMSE metric is calculated pixel-wise, and the SAM metric is estimated for each spectral signature and averaged.

4.3.1. Reconstruction performance for given aberrations One of the most important parameters in designing the proposed sensing protocol is the additional optical aberration induced by the deformable mirror. Thus, we estimate this aberration in terms of the ZP that yields the best reconstruction quality. Specifically, this analysis is developed by varying the pure ZPs along with its amplitude coefficient, i.e., $W(x,y) = a_j Z_j(x,y)$ with $j \in \{5,6,7,8,9,10,12,13,14,15\}$ and $a_j \in \{0, 0.1, ..., 0.9, 1\}$. This cross-validation analysis is developed for each one of the three steps of the proposed reconstruction methodology, more precisely, it is developed for the estimation of $\mathbf{f}_g$ [see **Algorithm 2**], $\mathbf{f}_f$ [see **Algorithm 3**], and $\mathbf{f}_s$ [see **Algorithm 4**].
Figure 35 shows a 2D histogram that illustrates the SAM, RMSE, and SSIM metrics of (a) $\mathbf{f}_g$, (B) $\mathbf{f}_f$, and (C) $\mathbf{f}_s$ reconstruction on the set of 50 spectral depth images. In Fig. 35, the axes -x and -y represents the amplitude coefficient and the pure ZP, respectively, and for clarity, the ZP that allows obtaining the best results are bounded with a red box. Although optimal splitting ratios were found for each metric in Fig. 4.3, for real optical implementation purposes, it is desired to establish a unique (optimal) amplitude aberration. Therefore, a general performance metric fuses all the metric results to determine the optimal amplitude aberration is proposed. First, the metrics

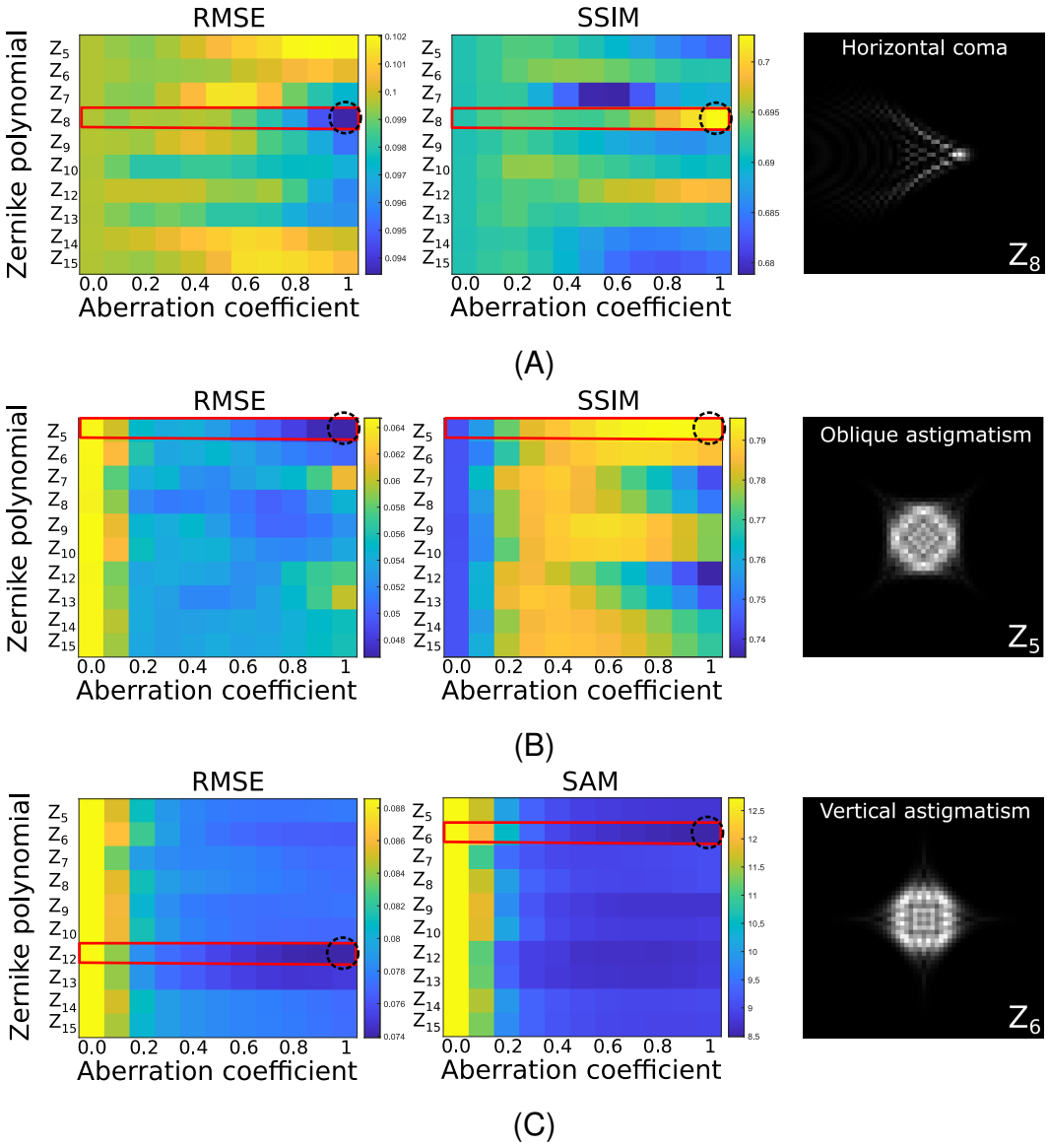


Figura 35. Zernike cross-validation analysis for the reconstruction of (A) $\mathbf{f}_g$ (**Algorithm 2**), (B) $\mathbf{f}_f$ (**Algorithm 3**), and (C) $\mathbf{f}_s$ (**Algorithm 4**). The experiments were performed by using pure ZP per simulation.

are transformed into dimensionless metrics, where the percentage error metric is selected. Then, a percentage error analysis is developed for each pixel of the matrices shown in Fig. 4.3, defined as $\tilde{R}^k_{i,j} = \left| \frac{\hat{R}^k - R^k_{i,j}}{\hat{R}^k} \right|$, where $\hat{R}^k$ represents the optimal metric value for the k-th matrix (i.e., either max or min value) and k index the total of matrix in Fig. 35, i.e., six (6). Finally, the resulting six matrices are average, and the optimal point is estimated. Based on above mentioned general metric, the ZPs that allow obtaining the best reconstruction performance of $\mathbf{f}_g$, $\mathbf{f}_f$, and $\mathbf{f}_s$, are Z_8, Z_5, and $\{Z_6, Z_{12}\}$, respectively. Nevertheless, the polynomial that exhibits the best reconstruction trade-off between $\mathbf{f}_g$, $\mathbf{f}_f$, and $\mathbf{f}_s$ is vertical astigmatism (Z_6) with $a_6 = 1$. Specifically, for the reconstruction of $\mathbf{f}_g$, the best results are achieved when using a horizontal coma (Z_8) with an amplitude coefficient of $a_8 = 1$. Here it is worth noting that the maximum relative absolute error between the best and worst result in terms of the SSIM and RMSE is approximately 1.06 % and 6 %, respectively. In the case of $\mathbf{f}_f$, the optimal reconstruction is achieved by oblique astigmatism (Z_5) with an amplitude aberration of $a_5 = 1$. Here, the second and third ZP that achieves the best reconstruction results are Z_9 and Z_6, respectively, where the maximum absolute error between them are 0.25 %, and 9.45 % in terms of SSIM and RMSE, respectively.

Finally, for the reconstruction of $\mathbf{f}_s$, the best spectral result in terms of SAM metric is achieved when a vertical astigmatism (Z_6) with an amplitude coefficient of $a_6 = 1$ is used. In contrast, in terms of SSIM and RMSE metrics, the best performance is achieved by the use of a vertical secondary astigmatism (Z_{12}) with an amplitude coefficient of $a_{12} = 1$. Here, to select the optimal ZP, we compare the reconstruction performance achieved by Z_6 and Z_{12} in terms of the SAM, SSIM and RMSE metrics, where it is found that the maximum relative absolute error between them is in average 3 % for Z_{12} and 2 % for Z_6. In this manner, the optimal ZP for the reconstruction of $\mathbf{f}_s$ is Z_6 with an amplitude coefficient of $a_6 = 1$. In summary, the optimal ZP to reconstruct the grayscale focus-stack and spectral information are Z_5, and Z_6, respectively, both

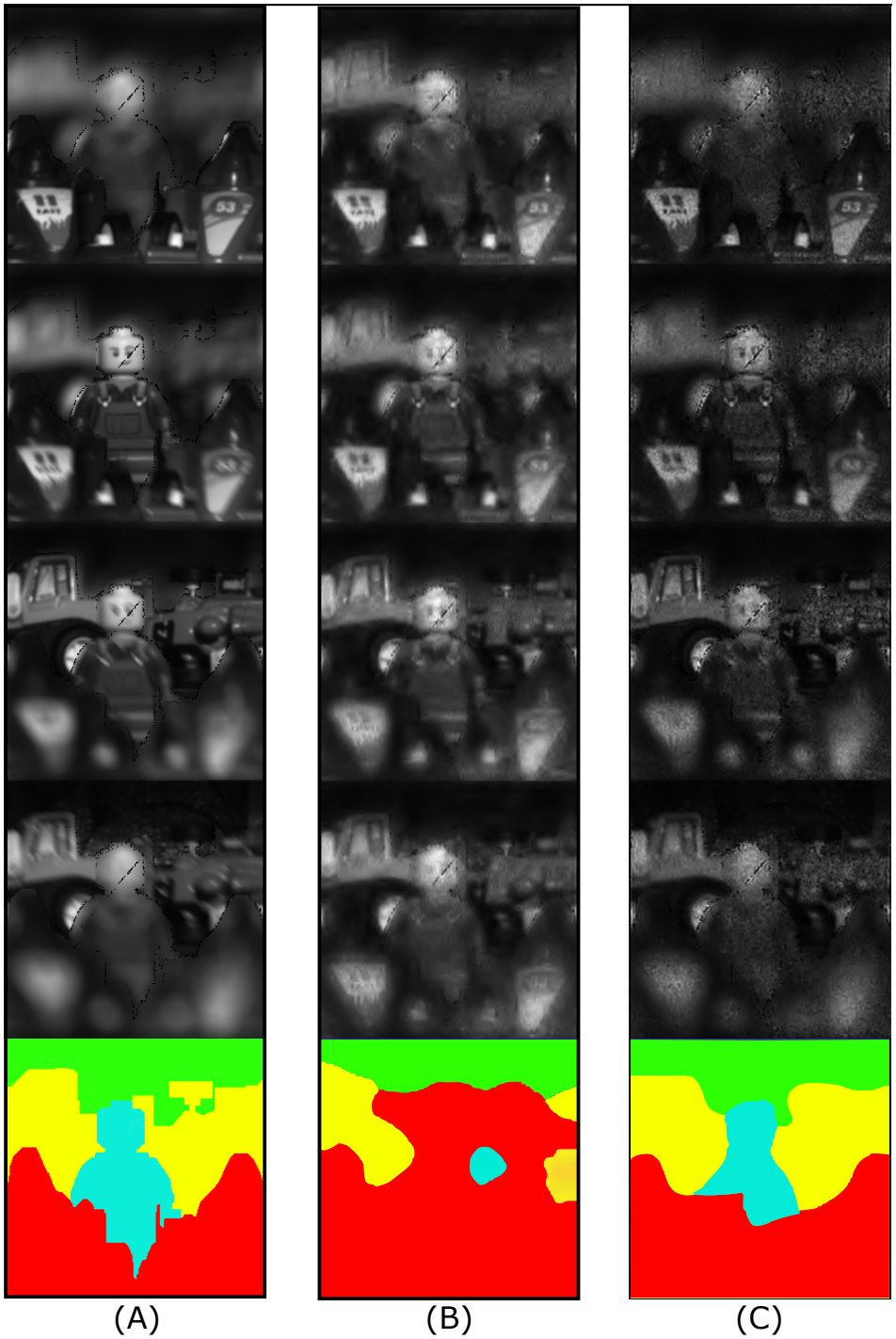


Figura 36. Focal-stack and depth map for the (A) Ground truth, (B) reconstruction via CASSDI-DFD, and (C) CASSDI-DFA approaches.

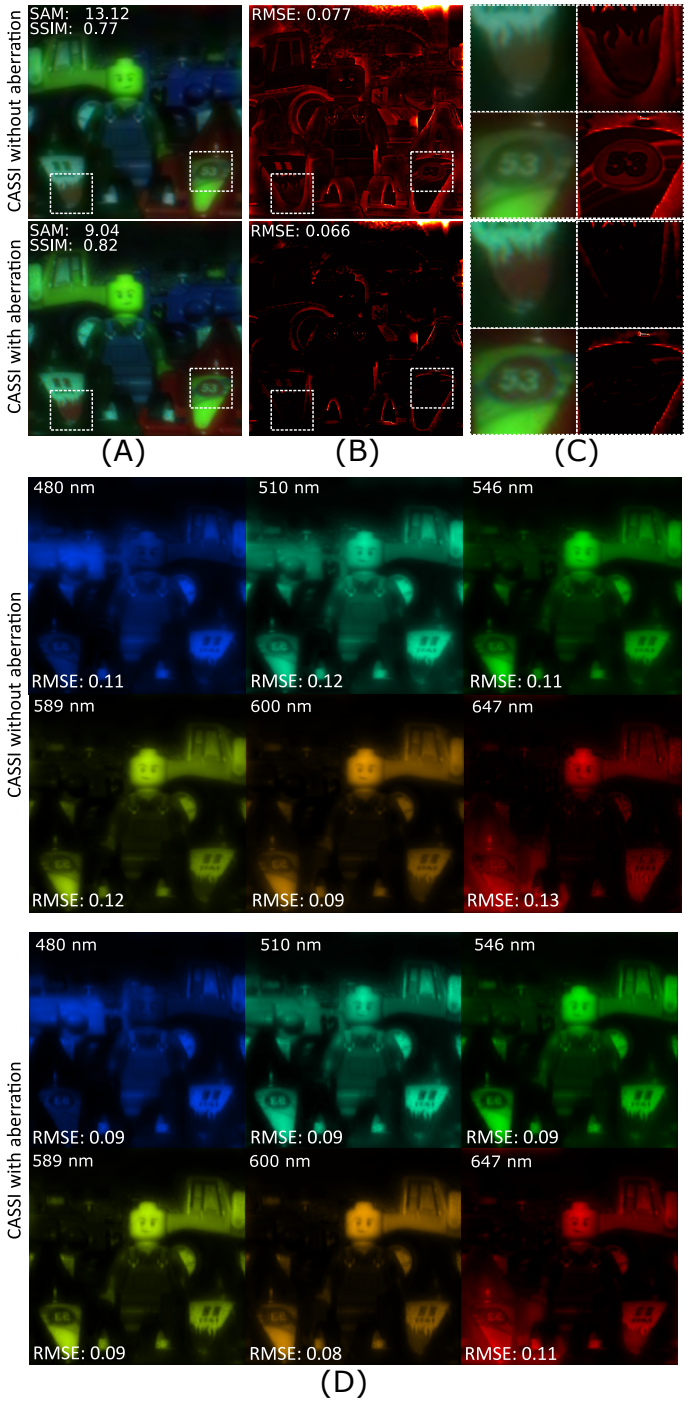


Figura 37. (A) Reconstructed data cubes mapped to RGB. (B) Reconstruction relative error. (C) Two zoomed portion of (A) RGB and (B) error images. (D) Illustration of 6 out of the 12 reconstructed spectral bands of the Cars scene.

with $a_5 = a_6 = 1$. Although the results obtained by the use of Z_5, and Z_6 are close, we select to Z_6 as the optimal pure ZP for the reconstruction of the SDI information, since it establishes the best reconstruction trade-off for $\mathbf{f}_g$, $\mathbf{f}_f$, and $\mathbf{f}_s$. The regularization parameters used on **Algorithms 2-4** were obtained experimentally through cross validation by minimizing the RMSE values.

4.3.2. Depth and spectral performance Figure 36 illustrates the grayscale focal stack and depth map estimation. Here, the depth map estimation is obtained via a pre-trained U-net neural network. Specifically, Fig. 36 in the first column shows the ground-truth focal stack and depth map features of the Cars image, the second and third-colum shows the reconstruction obtained by the CASSDI-DFA and CASSI-DFD approaches, respectively. These results echo the claims presented in Chang y Wetzstein, "Deep Optics for Monocular Depth Estimation and 3D Object Detection"; Henry Arguello y col. "Shift-variant color-coded diffractive spectral imaging system". En: *Optica* 8.11 (2021), págs. 1424-1434; Xiong Dun y col. "Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging". En: *Optica* 7.8 (2020), págs. 913-922, where the recovery process is improved using a different modulation per spectral/depth layer. Thus, the modulation mask leads to a more incoherent encoding operation, contributing to better image quality in reconstruction. Here, it can be noticed that the disambiguation of the focal stack features exhibit better results when the DFA approach is used. In the same manner, the depth map estimation shows improvements by using the DFA sensing approach. To further evaluate the spatial reconstruction quality, in Fig. 37 the results are evaluated in three fronts: first, the illustration of RGB composites of the attained reconstructions along with the SAM and SSIM; second, the cumulative absolute errors per spectral band along with the RMSE metric; and third, an illustration of 6 out of the 12 reconstructed bands is included. It can be noticed that the visual quality attained by performing a phase-amplitude modulation overcomes the results attained by performing just an

amplitude modulation. Moreover, the metric results in terms of SAM, RMSE, and SSIM, validate the visual quality results. This comparison shows that spectral signature reconstruction obtained by the CASSDI-DFA approach overcomes the results obtained by the CASSI-DFD approach.

4.4. Proof of concept experiments

We have built a testbed in our laboratory so as to demonstrate the proposed system through a proof-of-concept prototype, as shown in Fig. 38 (A). This prototype uses a Navitar (12 mm Fixed Focal Length, MVL12M23 - 12 mm EFL, f/1.4) as the objective lens to image the scene onto the focal plane of a relay lens (Achromatic Doublet Lens f=75 mm, Thorlabs, AC254-075-A-ML) to transmit the wavefront onto a deformable mirror (Actuator Piezo DM, Thorlabs, DMP40-P01-40). Then, a standard relay lens (Achromatic Doublet Lens f=75 mm, Thorlabs, AC254-075-A-ML), located at its focal length of the DM, is used to image the scene onto a digital micromirror device (DMD, Texas Instruments, D4120). Then, a standard relay lens (Thorlabs AC254-100-A-ML, f=100 mm, ϕ1") located at 100 mm of the DMD is used to attain two 4F-systems, split by the beam splitter (Thorlabs CCM1-BS013, 30 mm non-polarizing beamsplitter), which transfers half the light to the CASSI-arm (transmissive) and the other half to a side-arm (reflective).

Here it is worth noting that the side-arm is just used for calibration and analysis purposes, but this arm is removed for the final version of the proposed CASSDI-DFA system. In the CASSI-arm, a lens (Thorlabs AC254-100-A-ML, f=100 mm, ϕ1") and a double Amici prism are coupled to a rotation mount (Thorlabs CRM1P, 30 mm cage rotation mount, ϕ1") to precisely adjust the dispersion angle horizontally. A CCD sensor (Stingray F-080B, 4.65 μm pixel size) is located at the focal length of the lens, where the phase-modulated, spatial modulated, and spectral d two-dimensional projection of the scene is acquired. In the reflective arm, the wavefront is propagated

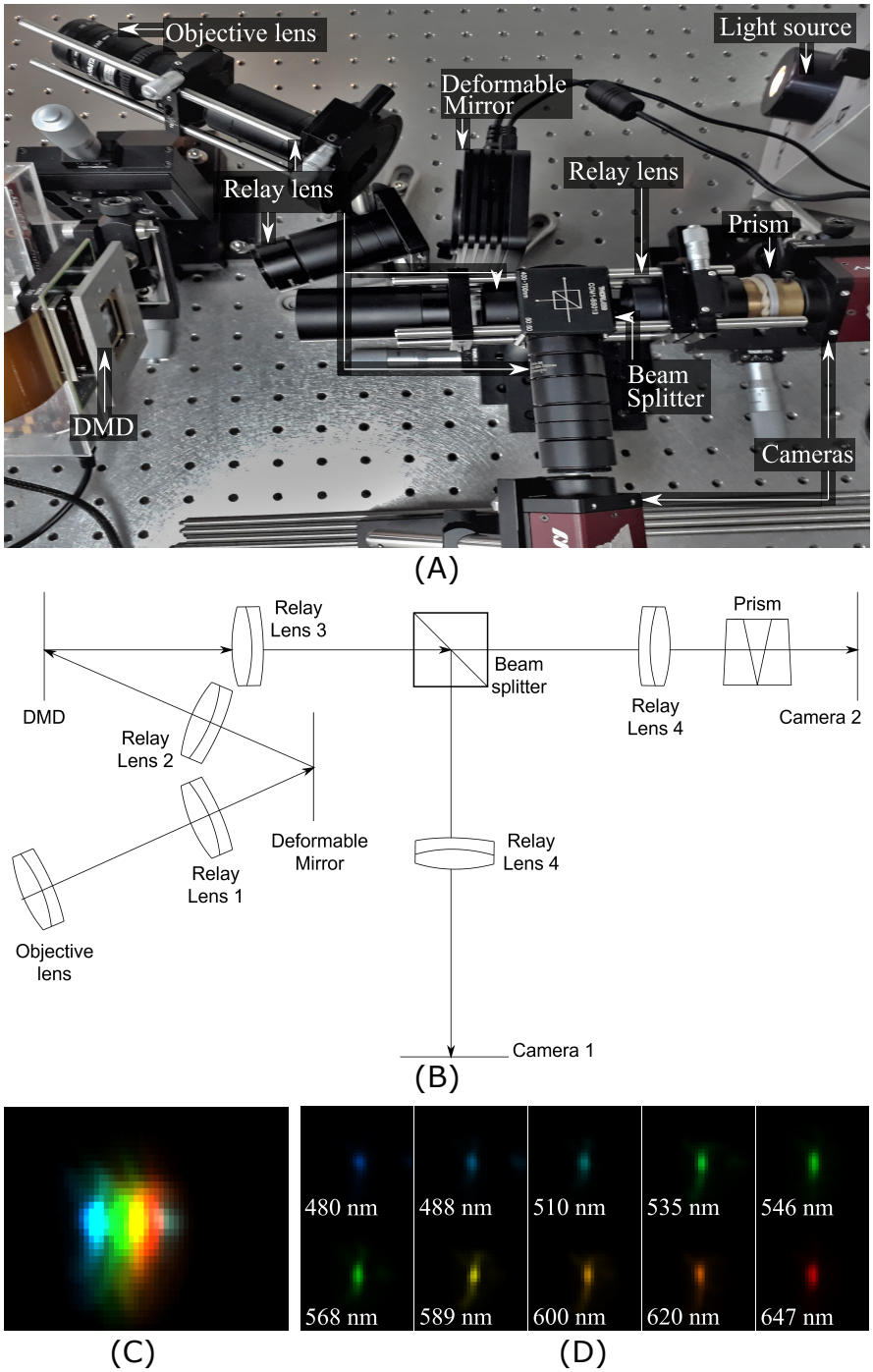


Figura 38. (A) Testbed implementation of the CASSDI-DFA. (B) Geometric sketch of the optical system. (C) Point spread function of the CASSDI-DFA system by setting the deformable mirror with $\{Z_4, Z_6\} = \{0, 0.9\}$. (D) Intensity PSF splitting as a function of the wavelength.

to a relay lens (Thorlabs AC254-100-A-ML, f=100 mm, ϕ1") to image the scene onto a CCD sensor (Stingray F-080B, 4.65 μm pixel size) located at the focal length of the lens. In this section, the reconstructions were attained using the single compressed measurement acquired with the CASSI-arm. For the sensing process, the deformable mirror is configurated with vertical astigmatism (Z_6) with $a_6 = 0.9$. The intensity point-spread function (PSF) of the CASSDI-DFA system is illustrated in Fig. 38(B), where it can be appreciate the vertical distortion and the horizontal dispersion introduced by the DM and the prism, respectively. The intensity PSF is characterized using an optical fiber (Ocean BIF200-UV-VIS) as a point source connected to a monochromator (Newport TLS130B) as a tunable light source. Then, the intensity PSF characterization is developed in function of the spectral response, mirror deformation, and the three fixed depth planes. Figure 38(B) illustrated the spectral description of the intensity PSF in the spectral range from 450 to 700 in steps of 10 nm with a full width at half maximum (FWHM) of 10 nm. Finally, to better appreciate the intensity PSF behavior in Fig. 38(C) are illustrated 10 intensity PSFs. The characterized intensity PSF is then used to construct the matrix $\mathbf{\Phi}$ and matrix $\mathbf{\Phi}_s$.

The experiments consider one target scene named, Flowers, for which three depth planes were acquired using the controlled aberrations $\{Z_4, Z_6\} = \{0, 1\}$ (first plane), $\{Z_4, Z_6\} = \{0.12, 1\}$ (second plane), and $\{Z_4, Z_6\} = \{0.2, 1\}$ (third plane). Here it is worth noting that the polynomial Z_4 is associated with the depth sweep developed by the optical system, and the polynomial Z_6 is associated with the modulation introduced per depth plane. For each depth plane, the coded aperture is varied and set with a transmittance of 0.5, and the resulting compressed measurement is illustrated in Fig. 39(A). Specifically, this target is composed of two colored wooden flowers and a card with the HDSP logo, which are located at 50, 58, and 72 cm from the objective lens, respectively. The raw compressive projection exhibits a spatial resolution of 512×523 pixels, i.e., $L = 523 - 512 + 1 = 12$ spectral bands can be recovered. Follo-

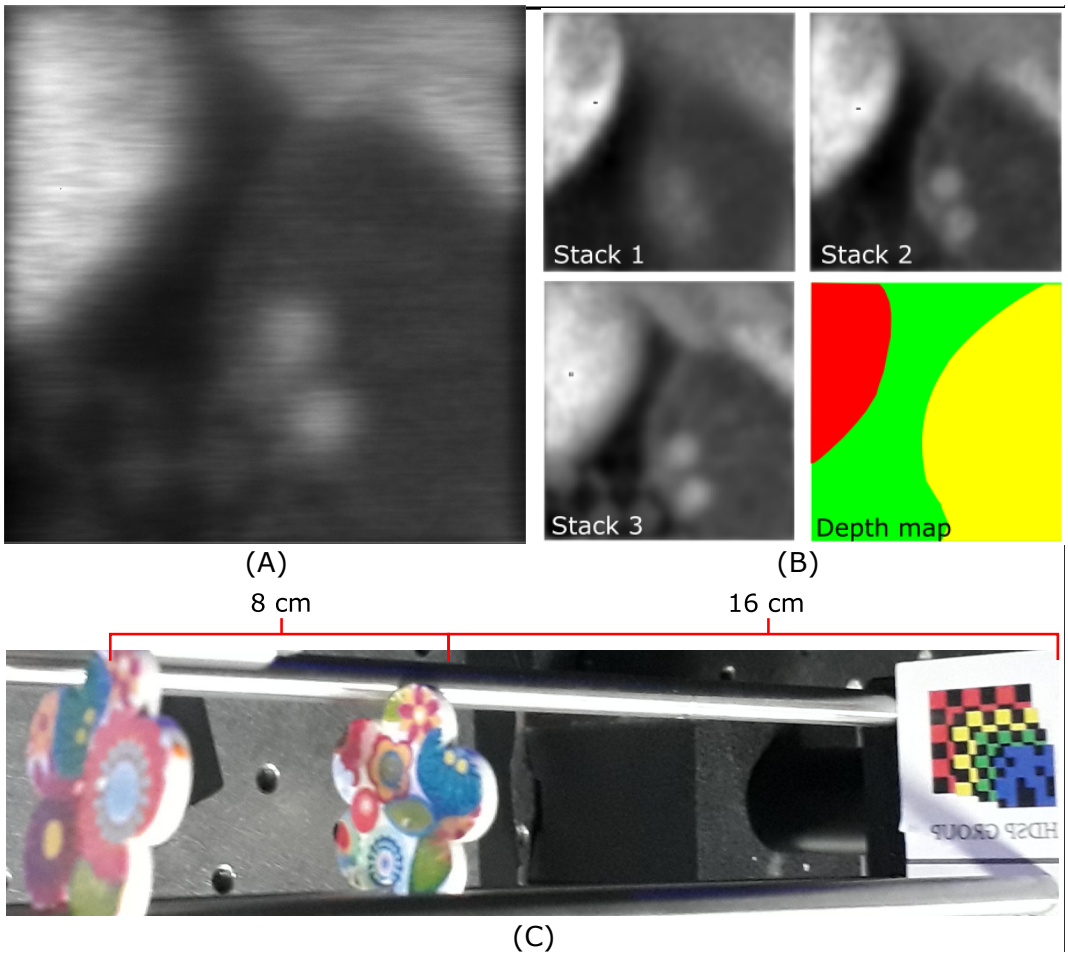


Figura 39. Testbed results. (A) Compressed measurements. (B) Illustration of the three reconstructed focal stack images and the depth map estimation via a pre-trained U-net network Olaf Ronneberger, Philipp Fischer y Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. En: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, págs. 234-241. (C) Top-view of the Flower scene.

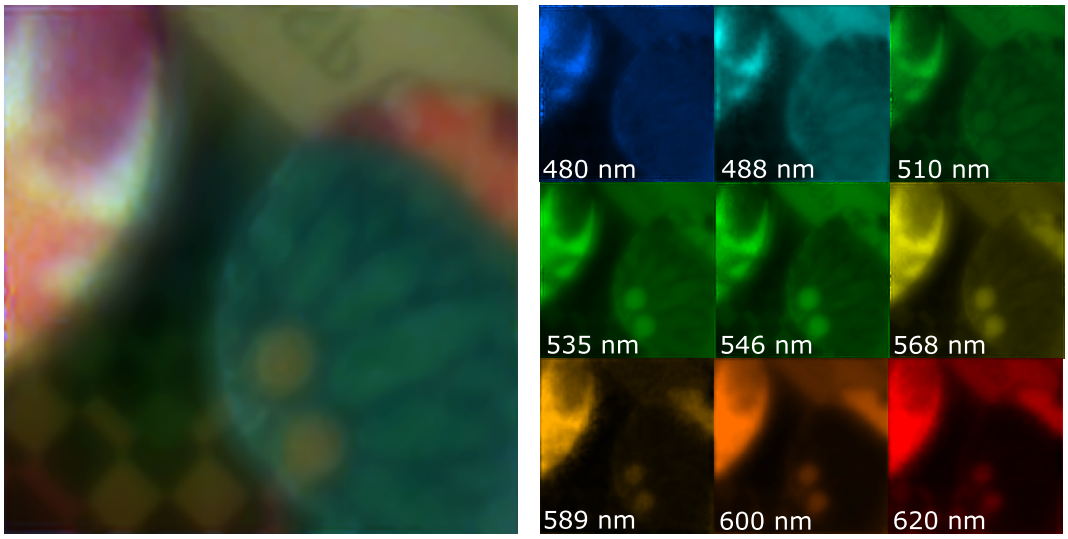


Figura 40. Testbed result. (A) RGB composite of the all-in-focus reconstructions. (B) Illustration of 9 out of the 12 reconstructed spectral bands of the Flowers scene.

wing the proposed reconstruction methodology, the raw measurement –Fig. 39(A)– is given as input to **Algorithm 3**, where the resulting focal-stack grayscale $\mathbf{f}_f$ reconstruction is shown in figure Fig. 39(B). Then, the U-net neural network is used to infer the depth map from the estimated focal-stack grayscale, as is shown in figure Fig. 39(B) (right-bottom picture). Figure 39(C) illustrated the scene which is constructed with three depth levels.

Finally, the compressed measurement along with the estimations $\mathbf{f}_g$ and $\mathbf{f}_f$, are introduced to the **Algorithm 4** to reconstruct the all-in-focus spectral datacube $\mathbf{f}_f$, and the RGB composite of the attained reconstruction is depicted in Fig. 40(A). Figure 40(B) shows the spatial reconstruction per band to evaluate the accuracy of the spectral reconstruction, 9 out of the 12 spectral bands are depicted. In summary, it can be observed that the proposed CASSDI-DFA testbed system allows estimating simultaneously the spectral and depth information of a scene from a single compressed measurement. Note that the **Algorithm 4** returns spectral data cubes with their spectral signatures without illumination corrections, i.e., in intensity. Therefore, an additional offline reconstruction step is the spectral reflectance estimation, see

Supplemental material 9.

4.5. Conclusions

This chapter introduced the CASSDI-DFA system to capture spectral-depth images using a single compressed snapshot measurement. CASSDI-DFA performs dynamic phase and amplitude coding-through a DM and a DMD at different optical path stages during the detector array's integration time. The achieved multiplexing of depth-from-aberration along with a coded-and-dispersed projection allowed the estimation of an all-in-focus grayscale focal-stack and also an all-in-focus spectral version of the scene, which are used as prior information for the spectral image reconstruction. To estimate the low-dimensional projections, we proposed a sequential reconstruction methodology composed of three ADMM-based optimization problems. The proposed CASSDI-DFA system relies on a single sensor, as a potential advantage in contrast with state-of-the-art systems that rely on stereo or multi-sensors. The optimal aberration for the CASSDI-DFA was concluded as vertical astigmatism Z_6 with an amplitude coefficient of $a_6 = 0.9$. Further, the proposed imaging system performance was demonstrated via simulations against a depth-from-defocus sensing alternative of the proposed CASSDI-DFA, and through a proof-of-concept implementation, which confirmed that our proposed approach represents an efficient alternative to capture spectral-depth images with a single sensor in a single snapshot.

As a chapter summary and to assess the CASSDI-DFA's pros and cons, we construct a radar chart with seven relevant CoSDI features: compactness, low costs, low-recovery complexity, spectral recovery, depth recovery, modularity, and encoding flexibility. Here, the bottom and the top contours represent the worst and best scores, respectively, i.e., the ideal CoSDI system would have to reach the upper shell in the seven features. Specifically, Fig. 41 illustrates the CASSDI-DFA's radar chart, showing that CASSDI-DFA offers high coding flexibility for either the amplitude or phase

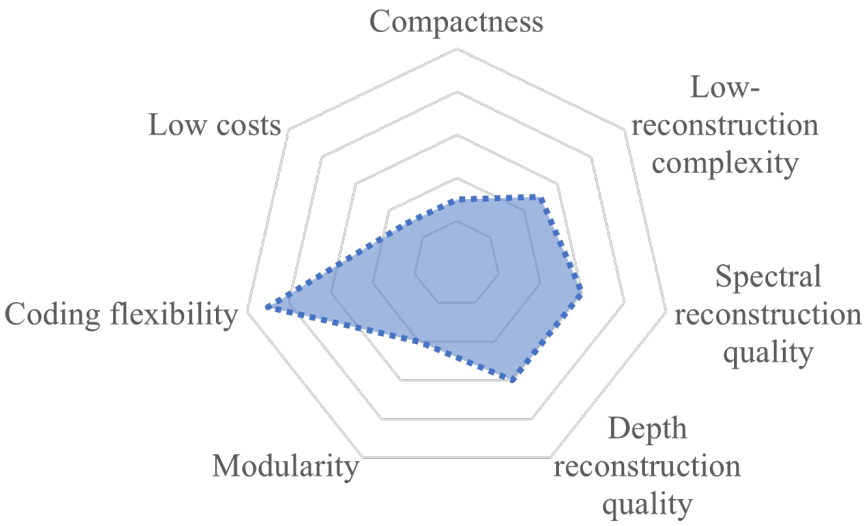


Figura 41. Chart graph for the CASSDI-DFA system
wavefront field at the expense of the modularity and low costs.

5. C^{3}SLFI: Color-coded Compressive Spectral Light Field Imager

Few light field (LF)-based compressive spectral depth imagers (CoSDI) have been proposed in the literature, including hybrid camera systems that use coded apertures, prisms, and photomasks Xiong y col., “Snapshot hyperspectral light field imaging”; Zhu y col., “Hyperspectral Light Field Stereo Matching”; Zhao y col., “Heterogeneous camera array for multispectral light field imaging”. Works in Zhu y col., “Hyperspectral Light Field Stereo Matching” and Zhao y col., “Heterogeneous camera array for multispectral light field imaging” presented camera array-based CoSDI systems that use independent band-pass filters per camera, which allow high spatial, spectral, and angular resolution at the cost of the temporal resolution, angular variance, and implementation costs. This filter-array geometry is inspired by the color-coded aperture approach used in the compressive spectral imaging (CSI) frameworks Claudia V Correa, Henry Arguello y Gonzalo R Arce. “Snapshot colored compressive spectral imager”. En: *JOSA A* 32.10 (2015), págs. 1754-1763. To overcome the cost and temporal resolution shortcomings, in Xiong y col., “Snapshot hyperspectral light field imaging” it was introduced a dual-arm system composed of a beam splitter that divides the incoming wavefront to an LF Georgiev y col., “Lytro camera technology: theory, algorithms, performance analysis” and a CSI Wagadarikar y col., “Single disperser design for coded aperture snapshot spectral imaging” imager, named Lytro-CASSI system. The dual-arm-based CoSDI systems allow high spectral and depth resolution at the cost of the light throughput and computational complexity (i.e., co-registration tasks). In this path, conventional LF-based CoSDI’s reconstruction algorithms demand complex matrix-vector operations, i.e., mapping the multiple dimensions into large vectors or matrices. Works in Nicholas D Sidiropoulos y col. “Tensor decomposition for signal processing and machine learning”. En: *IEEE Transactions on Signal Processing* 65.13 (2017), págs. 3551-3582; Sidi-

ropoulos y col., “Tensor decomposition for signal processing and machine learning”; Andrzej Cichocki y col. “Tensor decompositions for signal processing applications: From two-way to multiway component analysis”. En: *IEEE Signal Processing Magazine* 32.2 (2015), págs. 145-163 has shown that operating the object in its native tensorial structure (e.g., three or four-dimensional) allows exploiting intrinsic multidimensional features such as local similarity and sparsity behavior.

The positive impact of color-coded apertures (CCA) to improve the CSI's reconstruction performance and the tensor models to alleviate the high-dimensional complexity naturally leads to the question: ***Is it possible to develop an LF-based CoSDI system to recover the spectral-depth information by relying on complementary CCAs, lenslet array, and a single sensor?***

To answer this question, this chapter introduces an LF-based CoSDI system that builds on the ideas of CCAs, lenslet array, and tensor models but relies on a single sensor. Specifically, we propose to encode– with two complementary CCA– multiple views of the scene via a lenslet array. Then, a dictionary-alike representation is calculated via the Tucker decomposition. To promote sparsity in the spatial and angular dimensions, resulting in reconstruction improvements. We have coined the latter, color-coded compressive spectral light field imager – C^3SLFI.

5.1. Continuous and tensor model

A spectral light field (SLF), denoted as $f_o(x, y, \lambda, u, v)$, is a 5-dimensional function that represents the spatio-angular-spectral information of a scene. Capturing this 5D information is very demanding with current technology. For instance, it is possible to capture the spatial information (x, y) at different spectral bands λ by tuning a set of optical filters, and the angular information either by varying the view-angle of the camera in a time-sequential manner, or by using an array of similar sensors distributed on a planar or spherical surface. However, these methods demand separately scanning either angularly, spatially or spectrally the scene, to obtain the SLF. To alleviate the scanning requirement, new methods have proposed to multiplex information, so that the 5D SLF can be captured in a single 2D sensor plane. This, in turn, permits to capture the SLF in a single multiplexed snapshot, thus enabling the sensing of dynamic scenes, at the cost of lower spatial or angular resolution. We build on the multiplexing alternatives to propose our SLF imager named color-coded compressive spectral light field imager (C^{3}SLFI). To obtain a SLF, C^{3}SLFI captures coded multiplexed measurements which are then processed by a tensor-based optimization algorithm to reconstruct it. The C^{3}SLFI is composed by an objective lens, an aperture stop, a microlens array, a pair of relay lenses, a CCA and a grayscale focal plane array (FPA) sensor, as detailed in Fig. 42. This architecture works as follows: first, the SLF is focused by the imaging lens inside the proposed device, where the microlens array views and splits the incoming light rays according to their orientation (u, v); second, the relay lenses propagates the different view angles throughout the color-coded aperture, which randomly samples the spatial (x, y) and the spectral domain of the SLF (λ). Finally, the CCD integrates the incoming splitted and spatio-spectrally encoded light field, resulting in a grayscale multiplexed measurement. Note that, that every angular view of the SLF can be modeled independently from the others. Finally, keep in mind that the CASSDI-DFA is an optical system en-

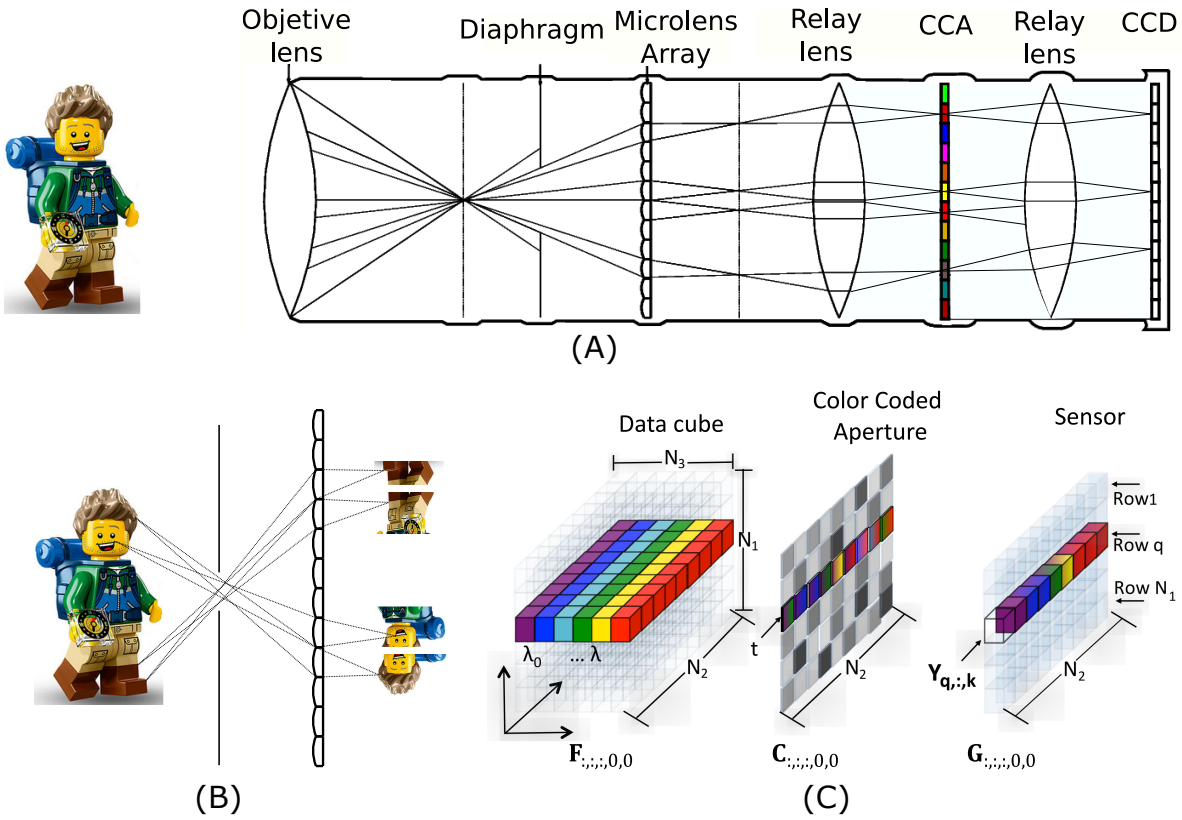


Figura 42. Sketch of the proposed C^{3}SLFI architecture. (A) Ray-tracing of the light propagation through the proposed imager. The objective lens image plane is observed by the microlens array which splits the light-field in different view-angles that are relayed through the CCA and then to the detector. (B) Illustration of the light field propagation through the aperture-stop and microlens array. (C) Sketch of the spectral data flow, where the q-th slice of the (u, v)-th view angle is modulated by the q-th row of the CCA, and then integrated by the CCD.

visioned to be used in computer vision applications in outdoor environments, i.e., for incoherent illumination.

5.1.1. Continuous model The first step is to define the C^{3}SLFI's propagation model in a continuous notation to develop the reconstruction methodology. Formally, let $f_o(x, y, \lambda, u, v)$ be the spatial-spectral-angular scene in intensity, where (x, y) index the spatial coordinates, (u, v) index the view-angle dimension, and λ index the wavelenght. First, an objective lens coupled with a lenslet array (LA) forms an image in an intermediate image plane, resulting in the wave field

$$f_1(x', y', \lambda) = \iiiint \gamma(\lambda) f_o(x, y, \lambda, u, v) h_{LA}(x - x', y - y', \lambda, u, v) dx dy du dv, \quad (71)$$

where $h_{LA}(\cdot)$ is the intensity PSF of the objective lens + LA and $\gamma(\lambda)$ represent the spectral source density. This 4D to 2D parameterization (without considering the spectral dimension) was proposed in Adelson y Wang, "Single lens stereo with a plenoptic camera", in which the view angle dimension is collapsed into the spatial dimension. In this regard, the conventional light field systems, and consequently the proposed C^{3}SLFI system, aim to reconstruct the spectral information in this 2D projection. Then, the spectral density immediately after the coded aperture plane can be expressed as

$$f_2(x'', y'', \lambda) = k(x'', y'', \lambda) \iint f_1(x', y', \lambda) h_1(x' - x'', y' - y'', \lambda) dx' dy', \quad (72)$$

with

$$k(x'', y'', \lambda) = \sum_{i_1, i_2, i_3} \tilde{C}_{j_1, j_2, i_3} \text{rect}\left(\frac{x''}{\Delta_c} - i_1, \frac{y''}{\Delta_c} - i_2, \frac{\lambda}{\Delta_d} - i_3 \right), \quad (73)$$

where $h_1(\cdot)$ represents the intensity PSF of the first relay lens, $k(\cdot)$ represent the color coded aperture, $\tilde{C}_{j_1, j_2, i_3} \in \{0, 1\}$ is the coding performed on the $(j_1, j_2, i_3)^{th}$ SLF voxel, with $j_1 = \{0, \ldots, N_1 N_4 - 1\}$, $j_2 = \{0, \ldots, N_2 N_5 - 1\}$ indexing the rows and

columns inside each angular view, $i_3 = \{0, \ldots, N_3 - 1\}$ indexing the spectral bands; Δ_c, and Δ_d account for the pixel sizes of the CCA and FPA detector, respectively. Then, the spatial-spectral encoded wavefront is propagated by a relay lens onto the detector array, such that the density can be seen in the detector plane (immediately before being integrated by the sensor) as

$$f_3(x''', y''', \lambda) = \iint f_2(x'', y'', \lambda) h_2(x'' - x''', y'' - y''', \lambda) dx'' dy'', \tag{74}$$

where $h_2(\cdot)$ is the intensity PSF introduced by the second relay lens. Finally, the measurement at the $(j_1, j_2)^{th}$ pixel is represented by

$$\tilde{G}_{j_1,j_2} = \int_{\Lambda} \rho(\lambda) \iint f_3(x''', y''', \lambda) p(j_1, j_2; x''', y''') dx''' dy''' d\lambda, \tag{75}$$

where Λ represents the wavelength axis over the spectral range, p is the pixel sampling function with $p(j_1, j_2; x''', y''') = \text{rect}\left(\frac{x'''}{\Delta_d} - j_1, \frac{y'''}{\Delta_d} - j_2\right)$ is the pixel sampling function, Δ_d represents the pixel's side length, and $\rho(\lambda)$ represents the detector's normalized quantum efficiency. Here $\tilde{\mathbf{G}} \in \mathbb{R}^{N_1 N_4 \times N_2 N_5}$ is a 2D measurement composed (in a matrix arranged manner) of the compressive measurement generated from each lenslet array projection. Let's $\mathbf{C}$ defined as the discrete version of the color coded aperture with $\tilde{\mathbf{C}} \in \mathbb{R}^{N_1 N_4 \times N_2 N_5 \times N_3}$. Hence, based on $\tilde{\mathbf{C}}$ and Eq. (75), the discretized version of the compressed measurement can expressed as

$$\tilde{\mathbf{G}} = \left[\sum_{i_3=0}^{N_\lambda - 1} \left[\left[\tilde{\mathbf{F}}_{:,:,i_3} * (\mathbf{H}_1)_{:,:,i_3} \right] \circ \tilde{\mathbf{C}}_{:,:,i_3} \right] * (\mathbf{H}_2)_{:,:,i_3} \right] + \tilde{\mathbf{E}}, \tag{76}$$

where $\circ$ represent the element wise matrix product, $\mathbf{H}_1 \in \mathbb{R}^{N_1 N_4 \times N_2 N_5 \times N_3}$ represents the discrete intensity PSF introduced by the first relay lens, and $\mathbf{H}_2 \in \mathbb{R}^{N_1 N_4 \times N_2 N_5 \times N_3}$ represents the discrete intensity PSF introduced by the second relay lens. In computer vision for the design of recovery algorithms, a widespread approximation is to

neglect the lens aberrations by assuming their intensity PSF as delta Dirac. Although this assumption is physically unrealistic, this simple model suffices for several computer vision and commercial applications. Based on this approximation, Eq. (76) can be rewritten as

$$\tilde{\mathbf{G}} = \left[\sum_{i_3=0}^{N_3-1} \tilde{\mathbf{F}}_{:,:,i_3} \circ \tilde{\mathbf{C}}_{:,:,i_3}\right] + \tilde{\mathbf{E}}. \tag{77}$$

Additionally, approximating the noise variable $\tilde{\mathbf{E}}$ to a approximately Gaussian distribution (i.e., colored Gaussian noise) has been widely accepted in high-level computer vision tasks. Though this simple model is physically unrealistic, it suffices for some computer vision tasks, e.g., reconstruction, recognition, classification, detection, and understanding. Because the scope of this thesis is limited to reconstruction tasks and based on the plethora of state-of-the-art computational optics reconstruction algorithms, in this thesis, the noise is approximated by a colored Gaussian distribution. Finally, since the measurement generated by each view-angle (associated with each element of LA) doesn't intercept/fuse between them into the detector, $\tilde{\mathbf{G}}$ can be rearranged from a 2D tensor to a 4D tensor as follows

$$G_{i_1,i_2,i_4,i_5} = \tilde{G}_{i_1+i_4N_4,i_2+i_5N_5}, \tag{78}$$

for $i_1 = \{0,\ldots,N_1-1\}$, $i_2 = \{0,\ldots,N_2-1\}$, $i_4 = \{0,\ldots,N_4-1\}$, $i_5 = \{0,\ldots,N_5-1\}$, where $\mathbf{G} \in \mathbb{R}^{N_1\times N_2\times N_4\times N_5}$ is a 4D rearranged representation of the compressed measurement $\tilde{\mathbf{G}}$. The output of the system $\mathbf{G}^k$ is an $N_1 \times N_2 \times N_4 \times N_5$ signal, yielding a compression ratio of $1/N_3$.

5.1.2. Tensor model Compressive sensing processes such as that in Eq. (77) have been traditionally modeled as the linear system

$$\mathbf{g} = \mathbf{\Phi}\mathbf{f} + \boldsymbol{\epsilon}, \tag{79}$$

where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ represents the sensing transfer function, with $n = N_1 N_2 N_3 N_4 N_5$, $m = N_1 N_2 K N_4 N_5$ (where K is the number of snapshots) and $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$ are the vectorized versions of the discrete SLF $\mathbf{F}$ and the compressed measurements $\mathbf{G}$, respectively. Note that, such linear system requires to calculate the product between a big sparse matrix $\mathbf{\Phi}$ with a vector, which demands operations on the order of $\mathcal{O}(mn)$.

In this path, tensor representation allows recasting the matrix-vector product to the matrix product between low-dimensional matrices. Moreover, tensors are a widely used data representation style for interaction data in the Machine Learning (ML) application community Andrzej Cichocki y col. “Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions”. En: *Foundations and Trends® in Machine Learning* 9.4-5 (2016), págs. 249-429; Vassilis N Ioannidis y col. “Coupled graphs and tensor factorization for recommender systems and community detection”. En: *IEEE Transactions on Knowledge and Data Engineering* 33.3 (2019), págs. 909-920. In addition to applications in which the data is naturally represented in tensors, another commonly used case is the Wavelet basis representation based tensor sparsity for recovery applications Qi Xie y col. “Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.8 (2017), págs. 1888-1902. To facilitate the understanding of the terminology used throughout the chapter, the tensor notation and preliminaries are relegated to **Appendix 9.1** in the supplementary material. The tensorial representation preserves the original structure of the data, critical for the analysis of their intrinsic properties.

Based on the properties established in **Appendix 9.1** and assuming that the sensing matrix $\mathbf{\Phi}$ is a separable function Yair Rivenson y Adrian Stern. “Compressed imaging with a separable sensing operator”. En: *IEEE Signal Processing Letters* 16.6 (2009), págs. 449-452, Eq. (79) can be rewritten in tensor notation as

$$\underline{\mathbf{G}} = \prod_{j=1}^{5} \underline{\mathbf{F}} \times_j (\mathbf{\Phi}_j) + \underline{\mathbf{E}}, \tag{80}$$

where $\mathbf{\Phi}_1 \in \mathbb{R}^{N_1 \times N_1}$, $\mathbf{\Phi}_2 \in \mathbb{R}^{N_2 \times N_2}$, $\mathbf{\Phi}_3 \in \mathbb{R}^{K \times N_3}$, $\mathbf{\Phi}_4 \in \mathbb{R}^{N_4 \times N_4}$, and $\mathbf{\Phi}_5 \in \mathbb{R}^{N_5 \times N_5}$ represent the Kronecker decomposition of the sensing matrix $\mathbf{\Phi} = (\mathbf{\Phi}_5 \otimes \mathbf{\Phi}_4 \otimes \mathbf{\Phi}_3 \otimes \mathbf{\Phi}_2 \otimes \mathbf{\Phi}_1)$, $\underline{\mathbf{F}} \in \mathbb{R}^{N_1 \times N_2 \times N_3 \times N_4 \times N_5}$ is the 5D tensor representation of the SLF image, $\underline{\mathbf{G}} \in \mathbb{R}^{N_1 \times N_2 \times K \times N_4 \times N_5}$ of the measurements and $\underline{\mathbf{E}}$ is the tensor that models the noise, which is modeled by a Gaussian distribution, i.e., colored Gaussian noise.

5.2. Reconstruction algorithms

5.2.1. Conventional reconstruction problem The tensorial sensing process demands a lower computational complexity on the order of $\mathcal{O}(\max(N_1, N_2, K, N_4, N_5)n)$ in contrast to $\mathcal{O}(KN_1^2N_2^2N_3N_4^2N_5^2)$ from Eq. (80). The conventional problem to obtain an estimation of the SLF from the set of compressed measurements, considering Eq. (80), is given by

$$\underset{\underline{\mathbf{F}}}{\text{argmin}} \ ||\underline{\mathbf{G}} - \prod_{j=1}^{5} \underline{\mathbf{F}} \times_j \mathbf{\Phi}_j||_F^2, \tag{81}$$

where $||\cdot||_F^2$ represents the squared Frobenius-norm, i.e., $||\mathbf{A}||_F = \sqrt[2]{\text{trace}(\mathbf{A}^*\mathbf{A})}$ with $\mathbf{A}$ as an arbitrary matrix and trace$(\cdot)$ is the trace function operator. One way to promote correct solutions of Eq. (81) is via regularization using sparsity priors that

exploit the high-correlations within the SLF. To this end, we rewrite the SLF tensor as

$$\underline{\mathbf{F}} = \prod_{j=1}^{5} \underline{\mathbf{\Theta}} \times_j \mathbf{\Psi}_j, \tag{82}$$

where $\underline{\mathbf{\Theta}} \in \mathbb{R}^{N_1 \times N_2 \times N_3 \times N_4 \times N_5}$ is a sparse 5D representation of $\underline{\mathbf{F}}$ in the bases $\mathbf{\Psi}_1 \in \mathbb{R}^{N_1 \times N_1}$, $\mathbf{\Psi}_2 \in \mathbb{R}^{N_2 \times N_2}$, $\mathbf{\Psi}_3 \in \mathbb{R}^{N_3 \times N_3}$, $\mathbf{\Psi}_4 \in \mathbb{R}^{N_4 \times N_4}$, and $\mathbf{\Psi}_5 \in \mathbb{R}^{N_5 \times N_5}$. Note that $\underline{\mathbf{\Theta}} = \prod_{j=1}^{5} \underline{\mathbf{F}} \times_j \mathbf{\Psi}_j^T$ and $\underline{\mathbf{F}} = \prod_{j=1}^{5} \left(\prod_{j=1}^{5} \underline{\mathbf{F}} \times_j \mathbf{\Psi}_j^T \right) \times_j \mathbf{\Psi}_j = \prod_{j=1}^{5} \underline{\mathbf{F}} \times_j (\mathbf{\Psi}_j^T \mathbf{\Psi}_j)$, i.e., $\mathbf{\Psi}_i^T \mathbf{\Psi}_i = \mathbf{I}_j$, where $\mathbf{I}_j$ represents an identity matrix. Therefore, based on Eq. (82), the reconstruction problem to estimate the sparse SLF can be rewritten as

$$\underset{\underline{\mathbf{\Theta}}}{\text{argmin}} \quad ||\underline{\mathbf{G}} - \prod_{j=1}^{5} \underline{\mathbf{\Theta}} \times_j (\mathbf{\Phi}_j \mathbf{\Psi}_j)||_F^2, \tag{83}$$

$$\text{subject to} \qquad ||\text{vec}(\underline{\mathbf{\Theta}})||_1 \leq S \tag{84}$$

where S is a parameter that controls the sparsity level. Reconstruction accuracy attained by the sparsity priors-based optimization problems' is determined by the representation basis capability to sparsify the scene. Therefore, designing an appropriate representation basis has become an essential step for developing sparsity-based reconstruction algorithms Kareth M León-López, Laura V Galvis Carreno y Henry Arguello Fuentes. "Temporal colored coded aperture design in compressive spectral video sensing". En: *IEEE Transactions on Image Processing* 28.1 (2018), págs. 253-264.

5.2.2. 5D to 3D spectral light field parameterization In natural scenes, it is expected that the radiances at a given arbitrary spatio-spectral position for a set of close view-angles are highly correlated, with some exceptions in light field imaging Manuel Martínez-Corral y col. "Recent advances in the capture and display of macroscopic and microscopic 3-D scenes by integral imaging". En: *Proceedings of the IEEE* 105.5 (2017), págs. 825-836. Taking this into account, the fifth-order SLF ten-

sor can be re-arranged as a third-order tensor with redundant spatial self-similarity of local patch patterns. Figure 43 depicts the 5D to 3D light field parameterization. Mathematically, this 5D to 3D parameterization can be expressed as

$$\tilde{F}_{j_1,j_2,i_3} = F_{\lfloor \frac{j_1}{N_4} \rfloor,\, \lfloor \frac{j_2}{N_5} \rfloor,\, i_3,\, j_1 - N_4 \lfloor \frac{j_1}{N_4} \rfloor,\, j_2 - N_5 \lfloor \frac{j_2}{N_5} \rfloor}, \tag{85}$$

for $j_1 \in (0, \ldots, N_1 N_4 - 1)$ and $j_2 \in (0, \ldots, N_2 N_5 - 1)$, where $\tilde{F}_{j_1,j_2,i_3}$ represents the $(j_1, j_2, i_3)^{th}$ position value of the third-order tensor $\underline{\tilde{\mathbf{F}}} \in \mathbb{R}^{N_1 N_4 \times N_2 N_5 \times N_3}$ and $\lfloor c \rfloor$ returns the greatest integer less than or equal to c. The parameterization described in Eq. (85) can be expressed in matrix form as $\text{vec}(\underline{\tilde{\mathbf{F}}}) = \mathbf{R} \cdot \text{vec}(\underline{\mathbf{F}})$, where $\mathbf{R} \in \mathbb{R}^{n \times n}$ is a permutation matrix with entries $R_{i,j} = 1$ if $j = a + b + c + d$ and 0 otherwise, for $i, j \in (0, \ldots, n-1)$, $a = \text{mod}(i, N_4) N_1 N_2 N_3$, $b = \lfloor \frac{\text{mod}(i, N_1 N_4)}{N_4} \rfloor$, $c = \lfloor \frac{\text{mod}(i, N_1 N_4 N_5)}{N_1 N_4} \rfloor N_1 N_2 N_3 N_4$, and $d = \lfloor \frac{i}{N_1 N_4 N_5} \rfloor N_1$. To keep notation light, we write $\tilde{N}_1 = N_1 N_4$ and $\tilde{N}_2 = N_2 N_5$. Given Eq. (85), Eqs. (82)-(83) can be rewritten as

$$\underline{\tilde{\mathbf{F}}} = \prod_{j=1}^{3} \underline{\tilde{\mathbf{\Theta}}} \times_j \tilde{\mathbf{\Psi}}_j, \tag{86}$$

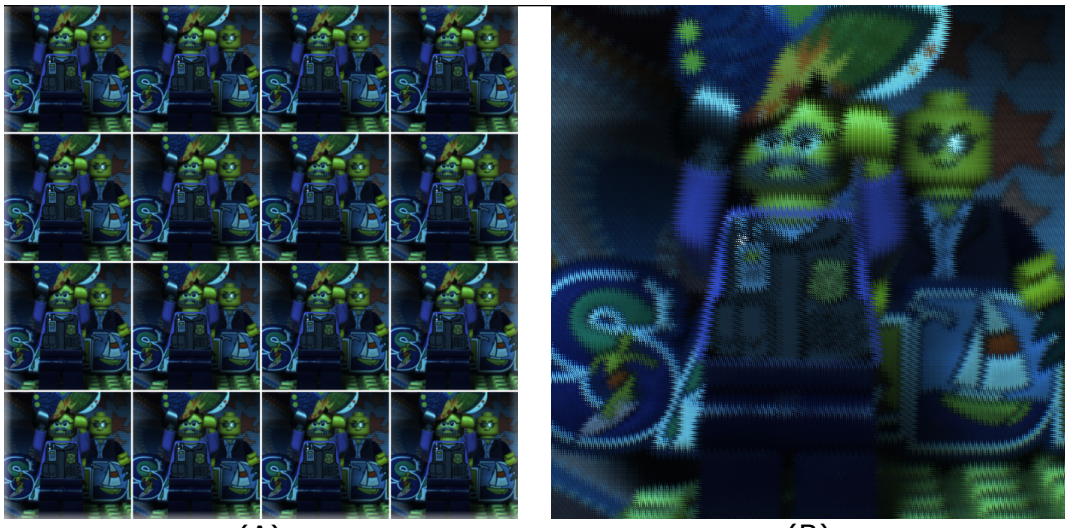
and

$$\begin{aligned} \arg\min_{\underline{\tilde{\mathbf{\Theta}}}} \quad & ||\underline{\tilde{\mathbf{G}}} - \textstyle\prod_{j=1}^{3} \underline{\tilde{\mathbf{\Theta}}} \times_j (\tilde{\mathbf{\Phi}}_j \tilde{\mathbf{\Psi}}_j)||_F^2, \\ \text{subject to} \quad & ||\text{vec}(\underline{\tilde{\mathbf{\Theta}}})||_1 \leq S \end{aligned} \tag{87}$$

with

$$\underline{\tilde{\mathbf{G}}} = \prod_{j=1}^{3} \underline{\tilde{\mathbf{\Theta}}} \times_j (\tilde{\mathbf{\Phi}}_j \tilde{\mathbf{\Psi}}_j) + \underline{\tilde{\mathbf{E}}}, \tag{88}$$

and $\underline{\tilde{\mathbf{E}}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3}$ irepresents the noise, $\underline{\tilde{\mathbf{\Theta}}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3}$ is a sparse representation of the third-order tensor $\underline{\tilde{\mathbf{F}}}$ in the Kronecker representation basis $\tilde{\mathbf{\Psi}}_1 \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_1}$,



(A) (B)

Figura 43. Graphic representation of the (A) 5D SLF and (B) its 3D parameterization. Spectral bands have been RGB-mapped.

$\tilde{\mathbf{\Psi}}_2 \in \mathbb{R}^{\tilde{N}_2 \times \tilde{N}_2}$, and $\tilde{\mathbf{\Psi}}_3 \in \mathbb{R}^{N_3 \times N_3}$ and $\underline{\tilde{\mathbf{G}}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times K}$ represents a third-order tensor containing the compressive measurements acquired with the re-arranged matrices $\tilde{\mathbf{\Phi}}_1 \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_1}$, $\tilde{\mathbf{\Phi}}_2 \in \mathbb{R}^{\tilde{N}_2 \times \tilde{N}_2}$ and $\tilde{\mathbf{\Phi}}_3 \in \mathbb{R}^{K \times N_3}$.

5.2.3. Online image representation To solve the optimization problem in Eq. (87), traditional reconstruction algorithms assume that the data are sparse when represented via universal bases, such as, Wavelets or the discrete cosine transform (DCT) Piotr Porwik y Agnieszka Lisowska. “The Haar-wavelet transform in digital image processing: its status and achievements”. En: *Machine Graphics and Vision* 13.1/2 (2004), págs. 79-98. In particular, a sparse version of the SLF on the Wavelet-DCT (WDCT) basis can be expressed as $\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\mathbf{F}}} \times_3 \tilde{\mathbf{\Psi}}_{DCT} \times_2 \tilde{\mathbf{\Psi}}_W \times_1 \tilde{\mathbf{\Psi}}_W$, where $\tilde{\mathbf{\Psi}}_{DCT} \in \mathbb{R}^{N_3 \times N_3}$ and $\tilde{\mathbf{\Psi}}_W \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_1}$ represent the DCT and Wavelet bases, respectively. Furthermore, recent works have demonstrated that learned dictionaries promote better sparse representations than the sparsification level achieved by the universal basis (UB), which translates to better reconstruction performance. Taking into account that C^{3}SLFI allows us to design the structure of the CCA Mejia y Argue-

llo, “Binary codification design for compressive imaging by uniform sensing”; Claudia V Correa, Henry Arguello y Gonzalo R Arce. “Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging”. En: *JOSA A* 33.12 (2016), págs. 2312-2322, and that learned dictionaries are preferred, we propose a methodology to learn a spatial basis for the SLF from the compressed measurements and iteratively learn the spectral basis. For this purpose, we first show through Lemma 1 that the grayscale version of $\underline{\tilde{\mathbf{F}}}$ can be obtained from $\underline{\tilde{\mathbf{G}}}$, without additional acquisitions from auxiliary cameras. Afterwards, using Lemma 2, we show that a self-learned spatial representation basis can be obtained from the grayscale image.

Lemma 1 *Let* $\mathbf{Z} = \underline{\tilde{\mathbf{F}}} \times_3 \mathbf{1}_{N_3}^T$ *with* $\mathbf{Z} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2}$ *being the grayscale version of the SLF* $\underline{\tilde{\mathbf{F}}}$ *and* $\underline{\tilde{\mathbf{C}}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3 \times K}$ *be a CCA with* $\underline{\tilde{C}}_{i_1,i_2,i_3,k} \in \{0,1\}$*, satisfying* $\sum_{k=1}^{K} \underline{\tilde{C}}_{i_1,i_2,i_3,k} = 1$*, i.e, complementary coded apertures. Then,*

$$\mathbf{Z} = \underline{\tilde{\mathbf{F}}} \times_3 \mathbf{1}_{N_3}^T = \underline{\tilde{\mathbf{G}}} \times_3 \mathbf{1}_K^T, \tag{89}$$

where $\mathbf{1}_{N_3}$ *and* $\mathbf{1}_K$ *are one-valued vectors of size* N_3 *and* K*, respectively.*

The proof of Lemma 1 is relegated to **Appendix 9.2** in the supplementary material. To calculate a spatial representation basis for $\underline{\tilde{\mathbf{F}}}$ from $\mathbf{Z}$, first we demonstrate that the tensorial product satisfies the commutative property for a specific case.

Lemma 2 *Let* $\mathbf{d} \in \mathbb{R}^{N_3}$ *be an arbitrary vector, then*

$$\underline{\tilde{\mathbf{\Theta}}} \times_3 \mathbf{d}^T \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \tilde{\mathbf{\Psi}}_2 = \underline{\tilde{\mathbf{\Theta}}} \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \tilde{\mathbf{\Psi}}_2 \times_3 \mathbf{d}^T. \tag{90}$$

The proof of Lemma 2 is relegated to **Appendix 9.3** in the supplementary material. Based on Lemmas 1 and 2 and using Eq. (86), a direct relation between the tensor $\mathbf{Z}$ and the self-learned spatial representation basis of $\underline{\tilde{\mathbf{F}}}$ is established in Corollary 1.

Corollary 1 *From Lemmas 1 and 2, and replacing Eq.* (86) *in Eq.* (89)*, the grayscale*

version of $\underline{\tilde{\mathbf{F}}}$ *can be expressed as*

$$\begin{aligned}\mathbf{Z} &= \left[\underline{\tilde{\mathbf{\Theta}}} \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \tilde{\mathbf{\Psi}}_2 \times_3 \tilde{\mathbf{\Psi}}_3\right] \times_3 \mathbf{1}_{N_3}^T \\ &= \underline{\tilde{\mathbf{\Theta}}} \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \tilde{\mathbf{\Psi}}_2 \times_3 \mathbf{1}_{N_3}^T \tilde{\mathbf{\Psi}}_3 \\ &= \tilde{\mathbf{\Theta}}_{2D} \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \tilde{\mathbf{\Psi}}_2 \\ &= \tilde{\mathbf{\Psi}}_2 \tilde{\mathbf{\Theta}}_{2D}^T \tilde{\mathbf{\Psi}}_1^T, \end{aligned} \tag{91}$$

where $\tilde{\mathbf{\Theta}}_{2D} = \underline{\tilde{\mathbf{\Theta}}} \times_3 \mathbf{1}_{N_3}^T \tilde{\mathbf{\Psi}}_3$ *is an auxiliary 2D tensor,* $\tilde{\mathbf{\Theta}}_{2D} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2}$*, the first equality comes after replacing Eq.* (86) *into Eq.* (89)*, the second results from using the unfolding n-mode Eq.* (193)*, the third equality after using Lemma 2, and the last equality results from the unfolding n-mode tensor product.*

In Corollary 1, we have established a direct relationship between the tensor $\mathbf{Z}$ and the spatial representation bases $\tilde{\mathbf{\Psi}}_1$ and $\tilde{\mathbf{\Psi}}_2$ in Eq. (86), thus proving that a dictionary can be learned online from a set of compressed measurements of C^{3}SLFI. Based on Eqs. (89) and (91), it is worth noting that the proposed system requires at least two acquisitions, which must be obtained using complementary coded apertures, so as to enable the estimation of the grayscale version of the scene, and in turn, the online dictionary learning (DL).

5.2.4. Proposed reconstruction algorithm The goal of the proposed algorithm, summarized in Algorithm 5, is to reconstruct $\underline{\tilde{\mathbf{\Theta}}}$ from its compressive measurements $\underline{\tilde{\mathbf{G}}}$, using the online trained representation basis $\tilde{\mathbf{\Psi}}_1$-$\tilde{\mathbf{\Psi}}_2$ obtained from $\mathbf{Z}$. In this section, we describe the algorithm in detail, which includes two main steps: (A) Initialize $\tilde{\mathbf{\Psi}}_3$ with a DCT basis and estimate the representation bases $\tilde{\mathbf{\Psi}}_1$ and $\tilde{\mathbf{\Psi}}_2$ from $\mathbf{Z}$ using the Tucker decomposition; (B) Solve the optimization problem Eq. (87) via ADMM and update the representation basis $\tilde{\mathbf{\Psi}}_3$ using the estimated scene.

Finding $\tilde{\mathbf{\Psi}}_1$, $\tilde{\mathbf{\Psi}}_2$ and $\tilde{\mathbf{\Psi}}_3$ Since $\mathbf{Z}$ is a second-order tensor and its mode-1 unfolding version is represented by $\mathbf{Z}_1 = \mathbf{Z}^T$, Eq. (91) can be rewritten as $\mathbf{Z}_1 = \tilde{\mathbf{\Psi}}_1 \tilde{\mathbf{\Theta}}_{2D} \tilde{\mathbf{\Psi}}_2^T$ with $\mathbf{Z}_1 \in \mathbb{R}^{\tilde{N}_2 \times \tilde{N}_1}$. Then, matrices $\tilde{\mathbf{\Psi}}_1$ and $\tilde{\mathbf{\Psi}}_2$ can be estimated from $\mathbf{Z}_1$ via

$$\begin{aligned} \min_{\mathbf{\Theta}_{2D}, \tilde{\mathbf{\Psi}}_1, \tilde{\mathbf{\Psi}}_2} \quad & ||\mathbf{Z}_1 - \tilde{\mathbf{\Psi}}_1 \tilde{\mathbf{\Theta}}_{2D} \tilde{\mathbf{\Psi}}_2^T||_F^2 + ||\tilde{\mathbf{\Theta}}_{2D}||_1 \\ \text{Subject to} \quad & \tilde{\mathbf{\Psi}}_1^T \tilde{\mathbf{\Psi}}_1 = \mathbf{I}, \;\; \tilde{\mathbf{\Psi}}_2^T \tilde{\mathbf{\Psi}}_2 = \mathbf{I}. \end{aligned} \tag{92}$$

Problem in Eq. (92) can be solved by alternating between

$$\arg\min_{\tilde{\mathbf{\Psi}}_1} \quad \|\mathbf{Z}_1^T - \tilde{\mathbf{\Psi}}_2^{\iota} \tilde{\mathbf{\Theta}}_{2D}^{\iota} \tilde{\mathbf{\Psi}}_1^T\|_F^2 + \beta_1 \|\tilde{\mathbf{\Psi}}_1^T \tilde{\mathbf{\Psi}}_1 - \mathbf{I}\|_F^2, \tag{93}$$

$$\arg\min_{\tilde{\mathbf{\Psi}}_2} \quad \|\mathbf{Z}_1 - \tilde{\mathbf{\Psi}}_1^{\iota+1} \tilde{\mathbf{\Theta}}_{2D}^{\iota} \tilde{\mathbf{\Psi}}_2^T\|_F^2 + \beta_2 \|\tilde{\mathbf{\Psi}}_2^T \tilde{\mathbf{\Psi}}_2 - \mathbf{I}\|_F^2, \tag{94}$$

$$\arg\min_{\tilde{\mathbf{\Theta}}_{2D}} \quad \|\mathbf{Z}_1 - \tilde{\mathbf{\Psi}}_1^{\iota+1} \tilde{\mathbf{\Theta}}_{2D} (\tilde{\mathbf{\Psi}}_2^{\iota+1})^T\|_F^2 + \|\tilde{\mathbf{\Theta}}_{2D}\|_1, \tag{95}$$

where $\{\beta_1, \beta_2\} > 0$ are regularization parameters. Since there is no knowledge of the spectral information, **Algorithm 5** uses $\tilde{\mathbf{\Psi}}_3 = \tilde{\mathbf{\Psi}}_{DCT}$ in the first iteration.

Reconstructing $\underline{\tilde{\mathbf{\Theta}}}$ With $\tilde{\mathbf{\Psi}}_1$, $\tilde{\mathbf{\Psi}}_2$ calculated and $\tilde{\mathbf{\Psi}}_3 = \tilde{\mathbf{\Psi}}_{DCT}$, **Algorithm 5** solves Eq. (87) and updates $\tilde{\mathbf{\Psi}}_3$ via a proximal alternating optimization methodology which finds $\underline{\tilde{\mathbf{\Theta}}}$ and $\underline{\tilde{\mathbf{\Psi}}}_3$ as

$$\arg\min_{\underline{\tilde{\mathbf{\Theta}}}} \quad ||\underline{\tilde{\mathbf{G}}} - \textstyle\prod_{j=1}^{3} \underline{\tilde{\mathbf{\Theta}}} \times_j (\tilde{\mathbf{\Phi}}_j \tilde{\mathbf{\Psi}}_j)||_F^2 + \beta ||\underline{\tilde{\mathbf{\Theta}}} - \underline{\tilde{\mathbf{\Theta}}}^{\iota}||_F^2 + \lambda ||\underline{\tilde{\mathbf{\Theta}}}||_1, \tag{96}$$

$$\arg\min_{\underline{\tilde{\mathbf{\Psi}}}_3} \quad ||\underline{\tilde{\mathbf{G}}} - \textstyle\prod_{j=1}^{3} \underline{\tilde{\mathbf{\Theta}}} \times_j (\tilde{\mathbf{\Phi}}_j \tilde{\mathbf{\Psi}}_j)||_F^2 + \beta ||\tilde{\mathbf{\Psi}}_3^T \tilde{\mathbf{\Psi}}_3 - \mathbf{I}||_F^2, \tag{97}$$

where $\beta > 0$ and $\lambda > 0$ are regularization parameters.

Solving Eq. (96) To solve Eq. (96) an ADMM approach is used. We introduce the functions $f_1(\tilde{\underline{\mathbf{\Theta}}}) = ||\tilde{\underline{\mathbf{G}}} - \prod_{j=1}^{3} \tilde{\underline{\mathbf{\Theta}}} \times_j (\tilde{\mathbf{\Phi}}_j \tilde{\mathbf{\Psi}}_j)||_F^2$ and $f_2(\underline{\mathbf{\Omega}}) = \beta||\underline{\mathbf{\Omega}} - \underline{\mathbf{\Omega}}^{\iota}||_F^2 + \lambda||\underline{\mathbf{\Omega}}||_1$ with the splitting variable $\underline{\mathbf{\Omega}} = \tilde{\underline{\mathbf{\Theta}}}$, such that Eq. (96) can be rewritten as

$$\begin{aligned} \underset{\tilde{\underline{\mathbf{\Theta}}}, \underline{\mathbf{\Omega}}}{\arg\min} \quad & f_1(\tilde{\underline{\mathbf{\Theta}}}) + f_2(\underline{\mathbf{\Omega}}), \\ \text{subject to} \quad & \underline{\mathbf{\Omega}} = \tilde{\underline{\mathbf{\Theta}}} \end{aligned} \tag{98}$$

The augmented Lagrangian (AL) of Eq. (98) is given by

$$\mathcal{L}(\tilde{\underline{\mathbf{\Theta}}}, \underline{\mathbf{\Omega}}, \nu) = f_1(\tilde{\underline{\mathbf{\Theta}}}) + f_2(\underline{\mathbf{\Omega}}) + \mu\|\tilde{\underline{\mathbf{\Theta}}} - \underline{\mathbf{\Omega}} - \underline{\boldsymbol{\nu}}\|_F^2, \tag{99}$$

where $\underline{\boldsymbol{\nu}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3}$ is the Lagrange multiplier and $\mu \geq 0$ is called the AL penalty parameter Manya V Afonso, José M Bioucas-Dias y Mário AT Figueiredo. “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems”. En: *IEEE Transactions on Image Processing* 20.3 (2011), págs. 681-695. Based on Eq. (99), the alternative form of Eq. (98) can be expressed as

$$\begin{aligned} \underset{\tilde{\underline{\mathbf{\Theta}}}, \underline{\mathbf{\Omega}}}{\arg\min} \quad & f_1(\tilde{\underline{\mathbf{\Theta}}}) + f_2(\underline{\mathbf{\Omega}}) + \mu\|\tilde{\underline{\mathbf{\Theta}}}^{\iota} - \underline{\mathbf{\Omega}}^{\iota} - \underline{\boldsymbol{\nu}}^{\iota}\|_F^2, \\ & \underline{\boldsymbol{\nu}}^{\iota+1} = \underline{\boldsymbol{\nu}}^{\iota} - (\tilde{\underline{\mathbf{\Theta}}}^{\iota+1} - \underline{\mathbf{\Omega}}^{\iota+1}). \end{aligned} \tag{100}$$

Additionally, Eq. (100) can be decoupled in two independent optimization problems

$$\tilde{\underline{\mathbf{\Theta}}}^{\iota+1} = \underset{\tilde{\underline{\mathbf{\Theta}}}}{\arg\min} \|\tilde{\underline{\mathbf{T}}} - \prod_{j=1}^{3} \tilde{\underline{\mathbf{\Theta}}} \times_j (\tilde{\mathbf{\Phi}}_j \tilde{\mathbf{\Psi}}_j)\|_F^2 + \mu\|\tilde{\underline{\mathbf{\Theta}}} - \underline{\mathbf{\Omega}}^{\iota} - \underline{\boldsymbol{\nu}}^{\iota}\|_F^2, \tag{101}$$

Algorithm 5 *C³SLF image reconstruction via online-training dictionary (OTD) Tensors*

Data: $\underline{\tilde{\mathbf{C}}}$, $\underline{\tilde{\mathbf{G}}}$, μ, $\underline{\mathbf{G}}$
Result: $\underline{\tilde{\mathbf{F}}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3}$
1. $\mathbf{Z} \leftarrow \underline{\mathbf{G}} \times_3 \mathbf{1}_K^T$
2. $\mathbf{Z}_1 \leftarrow$ mode-1 unfolding version of $\mathbf{Z}$
3. $[\tilde{\mathbf{\Psi}}_1, -, -] = \text{Tucker}(\mathbf{Z}_1)$
4. $[\tilde{\mathbf{\Psi}}_2, -, -] = \text{Tucker}(\mathbf{Z}_1^T)$
5. $\tilde{\mathbf{\Psi}}_3 = \tilde{\mathbf{\Psi}}_{DCT} \in \mathbb{R}^{N_3 \times N_3}$
6. $\underline{\mathbf{Q}} = \text{Arrange}\left[\underline{\tilde{\mathbf{C}}} \times_3 \mathbf{1}_{N_3}^T\right]$
for *{*$i_1 = 1$, $i_2 = 1$, $k = 1$*} to {*$\tilde{N}_1$, $\tilde{N}_2$, K*}* **do**
  7. $\underline{H}_{i_1,i_2,k} = (1 + \underline{Q}_{i_1,i_2,k})^{-1}$
end
repeat
  8. $\iota \leftarrow 0$,
  9. $[\underline{\tilde{\mathbf{\Theta}}}^{\iota}, \underline{\mathbf{\Omega}}^{\iota}, \underline{\boldsymbol{\eta}}^{\iota}] \leftarrow \mathbf{0} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3}$
  repeat
    10. Update $\underline{\mathbf{T}}$ with Eq. (105)
    11. Update $\underline{\mathbf{R}}$ with Eq. (104)
    12. Update $\underline{\tilde{\mathbf{\Theta}}}^{\iota+1}$ via Eq. (103)
    13. Update $\underline{\mathbf{\Omega}}^{\iota+1}$ via Eq. (106) with soft-thresholding
    14. $\underline{\boldsymbol{\nu}}^{\iota+1} = \underline{\boldsymbol{\nu}}^{\iota} - (\underline{\tilde{\mathbf{\Theta}}}^{\iota+1} - \underline{\mathbf{\Omega}}^{\iota+1})$
    15. $\iota \leftarrow \iota + 1$
  until *some stopping criterion is satisfied*;
  16. $\underline{\tilde{\mathbf{F}}} \leftarrow \prod_{j=1}^{3} \left(\underline{\mathbf{\Omega}}^{\iota+1} \times_j \tilde{\mathbf{\Psi}}_j\right)$
  17. $\tilde{\mathbf{F}}_{(3)} \leftarrow$ mode-3 unfolding version of $\prod_{j=1}^{2} \underline{\tilde{\mathbf{F}}} \times_j \tilde{\mathbf{\Psi}}_j$
  18. $[\tilde{\mathbf{\Psi}}_3, -, -] = \text{Tucker}(\tilde{\mathbf{F}}_{(3)})$
until *some stopping criterion is satisfied*;

and

$$\underline{\mathbf{\Omega}}^{\iota+1} = \arg\min_{\underline{\mathbf{\Omega}}} \beta ||\underline{\mathbf{\Omega}} - \underline{\mathbf{\Omega}}^{\iota}||_F^2 + \lambda ||\underline{\mathbf{\Omega}}||_1 + \mu ||\underline{\tilde{\mathbf{\Theta}}}^{\iota} - \underline{\mathbf{\Omega}} - \underline{\boldsymbol{\nu}}^{\iota}||_F^2. \tag{102}$$

To solve Eq. (101), we exploit the sensing geometry of C³SLFI and use Lemma 1 to obtain the closed-form solution

$$\underline{\tilde{\mathbf{\Theta}}}^{\iota+1} = \prod_{j=1}^{3} \left[\underline{\mathbf{G}} - \sum_{k=1}^{K} \left[(\underline{\mathbf{R}}_{::k} \times_3 \mathbf{1}_{N_3}) \circ \underline{\tilde{\mathbf{C}}}_{:::k}\right]\right] \times_j \tilde{\mathbf{\Psi}}_j^T, \tag{103}$$

where,

$$\underline{\mathbf{R}}_{:,:,k_1} = \sum_{k_2=1}^{K} \left[\underline{\mathbf{H}}_{:,:,k_1,k_2} \circ \left(\left[\underline{\mathbf{G}} \circ \underline{\tilde{\mathbf{C}}}_{:,:,:,k_2} \right] \times_3 \mathbf{1}_{N_3}^T \right) \right], \tag{104}$$

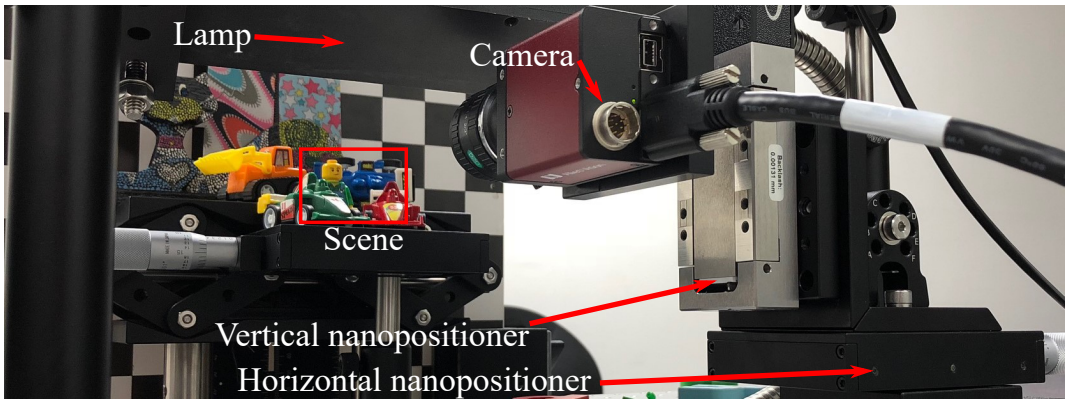
$$\underline{\mathbf{T}} = \sum_{k=1}^{K} \left[\underline{\tilde{\mathbf{C}}}_{:::k} \circ \left(\underline{\tilde{\mathbf{G}}}_{::k} \times_3 \mathbf{1}_{\tilde{N}_3} \right) \right] + \mu \prod_{j=1}^{3} (\underline{\boldsymbol{\Omega}} + \underline{\boldsymbol{\nu}}) \times_j \tilde{\boldsymbol{\Psi}}_j, \tag{105}$$

and the entries of $\underline{\mathbf{H}}$ are given by $\underline{\mathbf{H}}_{i_1,i_2,:,:} = \text{Rearrange}((\text{Arrange}(Q_{i_1,i_2,:,:}))^{-1})$ with $Q_{i_1,i_2,k_1,k_2} = \left[\sum_{i_3=1}^{N_3} \tilde{C}_{i_1,i_2,k_1} \tilde{C}_{i_1,i_2,k_2} \right] + \left(1 - e^{-\delta(k_1 - k_2)}\right)$, where $\text{Arrange}(\cdot) : \mathbb{R}^{1\times1\times K\times K} \rightarrow \mathbb{R}^{K\times K}$ is a function that collapses the first and second dimension, $\text{Rearrange}(\cdot) : \mathbb{R}^{K\times K} \rightarrow \mathbb{R}^{1\times1\times K\times K}$ is a function that rearranges a second-order tensor as a fourth-order tensor, and $\delta(\cdot)$ represents the Dirac delta function. A detailed description on how to obtain the closed-form solution of Eq. (101) is deferred to **Appendix 9.4** in the supplementary material. To solve the optimization problem in Eq. (102), a soft-thresholding approach is used, which entails the closed-form solution

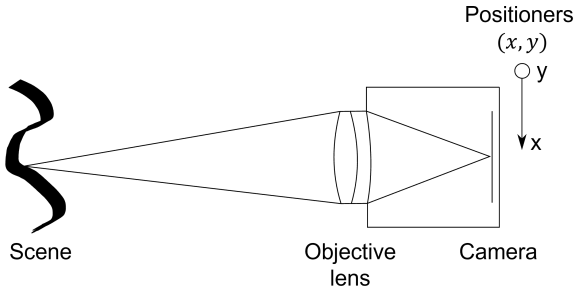
$$\underline{\Omega}_{i_1,i_2,i_3} = \text{sign}(\underline{P}_{i_1,i_2,i_3}) \text{max}(|\underline{P}_{i_1,i_2,i_3}| - b, 0), \tag{106}$$

for $b = \frac{\lambda}{(4\mu+2\beta)}$, where $\underline{P}_{i_1,i_2,i_3}$ represents the elements of the tensor $\underline{\mathbf{P}} = (\underline{\tilde{\boldsymbol{\Theta}}} - \underline{\boldsymbol{\nu}})$ with $\underline{\mathbf{P}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3}$.

Solving Eq. (97) After having an initial estimation of $\underline{\tilde{\boldsymbol{\Theta}}}$, we update $\tilde{\boldsymbol{\Psi}}_3$ by solving Eq. (97) via the Tucker decomposition. This methodology decomposes the estimated $\underline{\tilde{\mathbf{F}}}$ as a tensor product between the sparse tensor $\underline{\tilde{\boldsymbol{\Theta}}}$ and the three orthogonal matrices $\tilde{\boldsymbol{\Psi}}_1$, $\tilde{\boldsymbol{\Psi}}_2$, and $\tilde{\boldsymbol{\Psi}}_3$. Since $\underline{\tilde{\mathbf{F}}}$, $\tilde{\boldsymbol{\Psi}}_1$ and $\tilde{\boldsymbol{\Psi}}_2$ have been already calculated, Eq. (97) has the closed-form solution $[\tilde{\boldsymbol{\Psi}}_3, -, -] = \text{Tucker}(\mathbf{F}_{(3)}^T)$, where $\mathbf{F}_{(3)} \in \mathbb{R}^{\tilde{N}_1 \tilde{N}_2 \times N_3}$ represents the mode-3 unfolding version of $\prod_{j=1}^{2} \underline{\tilde{\mathbf{F}}} \times_j \tilde{\boldsymbol{\Psi}}_j$. In general, the computational



(A)



(B)

Figura 44. Raster scanner implementation to simulate the C^{3}SLFI. (B) Geometric sketch of the optical system.

complexity for a T-dimensional tensor with $N_1 = N_2 = ... = N_T = \sqrt[T]{n}$ is bounded by $O(Tn\sqrt[T]{n^2})$. A detailed description of the complexity and convergence analysis of the proposed algorithm is deferred to **Appendix 9.5**. It is worth noting that the proposed ideas can be extended to other imaging architectures such as 3D-CASSI Gehm y col., “Single-shot compressive spectral imaging with a dual-disperser architecture”, DD-CASSI Lin y col., “Dual-coded compressive hyperspectral imaging”, or SSCSI Xing Lin y col. “Spatial-spectral encoded compressive hyperspectral imaging”. En: *ACM Transactions on Graphics (TOG)* 33.6 (2014), pág. 233 by setting $N_4 = N_5 = 1$.

5.3. Simulation results

The performance of the proposed optical system and methodology is evaluated by sensing SLF images via simulated compressive measurements using the model described in Eq. (88) and reconstructing them with **Algorithm 5**. For these simulations, three SLF databases were captured in our laboratory, named Cars, Police, and Flower. These databases were captured using a CCD camera (Stingray F-080B, 4.65 μm pixel size) with a Navitar (MVL8M23, 8 mm EFL, f/1.4) objective lens, mounted on a $x-y$ nanopositioner (Newport MFA-VP25 with ESP301 motion controller) used for high-precision angular scanning, as shown in Fig. 44. The camera is moved along the horizontal and vertical axes of the scene, and in each step a tunable light source (Oriel TLS-300XR) illuminates the scene with wavelength steps of 3 nm between the 400-700 nm spectral range. The resulting SLF databases exhibit a spatial resolution of $N_1 \times N_2 = 256 \times 256$ pixels, $N_4 \times N_5 = 4 \times 4$ view angles, and $N_3 = 100$ spectral bands. Specifically, between any image and its neighbor, the camera was horizontally/vertically displaced by 2 millimeters. This sensing process allows the acquisition of a spectral depth database by emulating the multiview sensing approach. This configuration is mainly inspired by the Stanford Multi-Camera Array system, where the nanopositioner replaces the need for a multi-camera array.

In the experiments we compare the performance of the proposed system and methodology in three fronts. First, to test the proposed low-complexity online-training dictionary (OTD) representation basis expression power and representation capabilities, we include comparisons with a UB [via a WDCT] and a conventional DL [via overcomplete dictionaries]. Second, we evaluate the advantages of reconstructing either the 5D in Eqs. (83) or the 3D parameterized SLF in Eq. (87). Third, we compare the proposed architecture and methodology with state of the art work that uses side information reconstruction methodology with a dual-arm Lytro-CASSI architecture Xiong y col., “Snapshot hyperspectral light field imaging”. Specifically, the Lytro-CASSI ar-

chitecture is simulated as a multiple snapshot system, where the Lytro arm acquires an RGB representation of all the view angles, while the CASSI arm acquires K compressive measurements of the full field of view of the SLF. The CASSI measurements are simulated by using a time-varying black-and-white coded aperture with $50\,\%$ transmittance. The Lytro measurement is used to train a dictionary with 2D patches of size 8×8, selecting 10^4 patches at random, while a discrete cosine basis is used for the spectral dimension.

For clarity, the WDCT bases used to solve Eqs. (83) and (87) are fixed as $\mathbf{\Psi} = \mathbf{I}_{N_5 \times N_5} \otimes \mathbf{I}_{N_4 \times N_4} \otimes \mathbf{\Psi}_{DCT} \otimes \mathbf{\Psi}_W \otimes \mathbf{\Psi}_W$ and $\tilde{\mathbf{\Psi}} = \mathbf{\Psi}_{DCT} \otimes \tilde{\mathbf{\Psi}}_W \otimes \tilde{\mathbf{\Psi}}_W$, respectively, with $\mathbf{\Psi}_W \in \mathbb{R}^{N_1 \times N_1}$, $\tilde{\mathbf{\Psi}}_W \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_1}$, $\mathbf{\Psi}_{DCT} \in \mathbb{R}^{N_3 \times N_3}$. Here, the 2D spatial representation basis, $\mathbf{\Psi}_W$, was set to be the 2D-Wavelet Symmlet 8 basis. For the case of DL, we train the dictionary with 3D randomly sampled patches of size $8 \times 8 \times 100$ and 10^5 via the K-SVD algorithm. Furthermore, we choose a "2$\times$.ºver-complete dictionary (6200 atoms) which provides adequate representation of these particular training datasets Lin y col., "Spatial-spectral encoded compressive hyperspectral imaging". Here it is worth noting that to reconstruct the Flowers, Cars and Police SLF images, a dictionary was trained for each image by randomly sampling patches from the other SLF images; that is, Police and Cars for Flowers, Police and Flowers for Cars, and Flowers and Cars for the Police SLF, respectively. The latter procedure was followed due to the unavailability of public 5D SLF datasets. All simulations are conducted and timed using an Intel Core i7 3960X 3.30 GHz processor with 32 GB RAM memory. To compare the quality of the reconstructions, we use the peak signal to noise ratio (PSNR) and the spectral angular mapper (SAM) metrics.

5.3.1. WDCT and DL versus OTD representation basis In this subsection, we measure the quality of a representation basis by its power to sparsify a signal, that is, by measuring the number of nonzero elements required to accurately represent most of the signal energy. For that, the SLF is first projected onto the WDCT, DL or

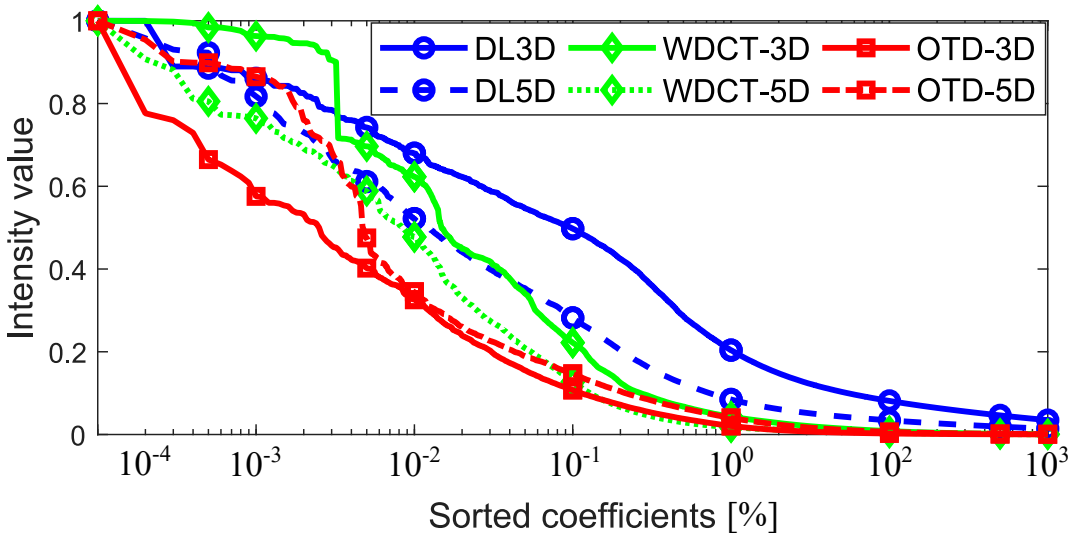


Figura 45. Sparse representation of a Car scene in its 5D and 3D structure, comparing the WDCT and DL bases with the proposed OTD basis.

the proposed OTD representation bases, then, the attained coefficients are sorted by magnitude, and finally we analyze the curve that is below the others. Two scenarios are evaluated, including the 3D and 5D parameterization, and the attained results are summarized in Fig. 45, where solid lines represent the 3D parameterization, dotted lines the 5D parameterization, and different colors account for the different representation bases. It can be seen by comparing pairs of solid versus dotted lines of the same color, that the proposed OTD basis attains a better representation of the signal compared with the traditional WDCT and DL. Finally, by comparing the solid and dotted lines it can be noted that the 3D parameterization entails an even better sparse representation than its original 5D version. Therefore, the use of the OTD basis with the 3D SLF representation attains the sparsest results overall. Although Fig. 45 indicates that UB outperform DL in its sparsification power, this behavior is attributed to the limited availability of SLF datasets for dictionary training, since public SLF with 100 spectral bands were not available at the time of conducting these simulations. Another reason may be the tuning parameters for the K-SVD, however, these were selected via a brute-force cross-validation methodology.

Dataset	Methods	Compression ratio (C)											
		$C = 9\,\%$						$C = 5\,\%$					
		15 [dB]		20 [dB]		50 [dB]		15 [dB]		20 [dB]		50 [dB]	
		PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM
Cars	OTD-3D	**30.5**	**9.1**	**32.8**	**7.1**	**38.5**	**3.6**	**28.2**	**11.9**	**31.6**	**10.0**	**33.8**	**4.9**
	OTD-5D	27.6	11.9	29.7	9.9	37.5	4.5	25.3	15.0	27.2	12.5	33.4	6.4
	DL-3D	28.4	16.0	30.24	11.5	31.0	5.2	27.6	12.3	28.2	15.6	29.6	6.1
	DL-5D	30	14.8	32.3	10.2	34.9	4.3	26.8	20.6	30.6	14.5	33.3	5.0
	WDCT-3D	27.8	17.6	29.8	9.8	30.1	6.5	24.8	23.0	26.0	16.3	27.3	13.5
	WDCT-5D	25.6	17.8	27.8	13.2	33.2	8.4	23.3	24.3	25.2	17.2	28.9	12.2
	LytroCASSI	27.3	17.3	29.9	12.3	34.0	7.3	25.9	16.0	28.3	14.9	30.1	11.0
Police	OTD-3D	**31.4**	**9.5**	**33.8**	**6.0**	**40.0**	**2.6**	**29.3**	**10.6**	**31.4**	**8.1**	**35.7**	**4.1**
	OTD-5D	28.6	10.5	30.7	8.4	39.3	3.2	26.6	13.3	28.6	10.8	34.9	4.9
	DL-3D	28.7	12.0	30.6	8.5	31.0	5.5	28.2	12.5	28.6	11.5	29.8	6.5
	DL-5D	30.2	11.0	33.2	7.4	33.8	4.7	27.0	15.7	30.6	10.2	33.1	5.8
	WDCT-3D	27.6	11.9	28.5	9.4	30.7	6.4	25.7	14.2	26.9	11.8	30.8	6.0
	WDCT-5D	26.7	12.6	28.8	10.2	34.5	4.1	24.5	15.8	26.2	13.1	30.0	6.2
	LytroCASSI	29.2	11.5	31.1	9.4	35.7	5.5	27.9	12.7	30.5	10.5	32.1	8.4
Flower	OTD-3D	**27.9**	**8.9**	29.9	**7.1**	**35.2**	**3.6**	**25.8**	**11.2**	**28.8**	**9.0**	**31.4**	**5.1**
	OTD-5D	25.3	12.2	27.4	10.3	34.5	4.9	23.4	14.3	25.0	12.4	30.7	6.2
	DL-3D	26.3	14.1	28.3	10.5	32.2	5.2	23.5	18.8	26.2	14.0	29.8	10.5
	DL-5D	27.7	12.9	**31.2**	9.1	32.6	4.7	24.4	17.7	28.3	12.5	31.0	5.6
	WDCT-3D	24.9	13.0	26.0	10.1	28.2	7.2	23.2	14.7	24.5	12.4	25.9	8.9
	WDCT-5D	24.1	13.1	26.0	11.4	32.4	5.0	21.9	16.1	23.7	14.2	28.3	7.2
	LytroCASSI	24.0	14.9	27.7	11.3	31.8	7.2	24.2	14.3	26.3	14.4	28.2	11.1

Tabla 1. Reconstruction results of the WDCT, DL and OTD bases when used over the 5D or 3D SLF parameterization, in terms of PSNR and SAM. The performance of the state-of-the-art Lytro-CASSI architecture is also included. Two compression ratios, $C \in \{5\,\%, 9\,\%\}$, and three noise power levels, SNR $\in \{15, 20, 50\}$ [dB], are evaluated.

5.3.2. Performance evaluation for different compression ratios and additive noise scenarios This test compares the reconstruction quality for different compression ratios, $C = \frac{m}{n}$, and levels of additive noise in the compressive measurements, that follow a colored Gaussian distribution with mean and variance given by a specific power of the signal to noise ratio (SNR). In particular, three SNR levels $\{15, 20, 50\}$[dBs] and two compression ratios $\{5\,\%, 9\,\%\}$ (corresponding to $K = 5$ and $K = 9$ snapshots, respectively) are evaluated. Table 1 illustrates the averaged spatial and spectral reconstruction quality measured in terms of PSNR and SAM. For clarity, the best results are bold-faced and the second best are underlined. To evaluate the impact of the representation basis in the quality of reconstruction, we analyze the performance of the OTD, WDCT, and DL bases on their 3D (first, third, and fifth rows of Table 1) and 5D (second, fourth, and sixth rows of Table 1) SLF image parame-

terization. There, it can be seen that the OTD basis overcomes the results obtained with the DL and WDCT bases overall, for both 5D and 3D cases. Specifically, for the 3D parameterization, the OTD gains up to 3.2 and 4.9 dB in averaged PSNR against the DL and WDCT bases, respectively, and attains 3.5° and 4.5° less spectral distortion in averaged SAM against the DL and WDCT bases, respectively. Regarding of the 5D parameterization, the DL gains up to 0.8 and 3.2 dB in averaged PSNR against the OTD and WDCT bases, and the OTD attains 0.84° and 2.8° less spectral distortion in averaged SAM against the DL and WDCT bases.

To evaluate the advantage of the C^{3}SFLI against the state-of-the-art work, we compare the performance of acquiring and reconstructing the SLF image against the dual-arm Lytro-CASSI architecture proposed in Marwah y col., "Compressive light field photography using overcomplete dictionaries and optimized projections". Specifically, the results attained with Lytro-CASSI are summarized in Table 1 (seventh row of each SLF). It can be seen that the OTD-3D gains up to 2.9 dB in averaged PSNR and attains 4.73° less spectral distortion in averaged SAM against the Lytro-CASSI. In summary, in terms of averaged PSNR, the OTD-3D method gains up to 1.4, 2.2, 2.9, 3.2, 4.6, and 4.8 dB against the DL-5D, OTD-5D, Lytro-CASSI, DL-3D, WDCT-5D, and WDCT-3D methods, respectively. In terms of SAM, the OTD-3D method attains 2.17, 3, 3.6, 4.3, 4.5 and 5° less spectral distortion in averaged SAM against the OTD-5D, DL-5D, DL-3D, Lytro-CASSI, WDCT-3D, and WDCT-5D methods, respectively.

5.3.3. Spatial, angular, and spectral performance To further evaluate the spatial reconstruction quality, Fig. 46 shows the cumulative absolute error between the ground-truth Cars SLFI and the reconstruction via OTD-3D, OTD-5D, DL-3D, DL-5D, WDCT-3D, WDCT-5D, and Lytro-CASSI. In particular, these reconstructions are obtained from $C = 5\,\%$ measurements and a noise power level SNR=20. It can be observed that the error of the OTD-3D method is lower than that of Lytro-CASSI,

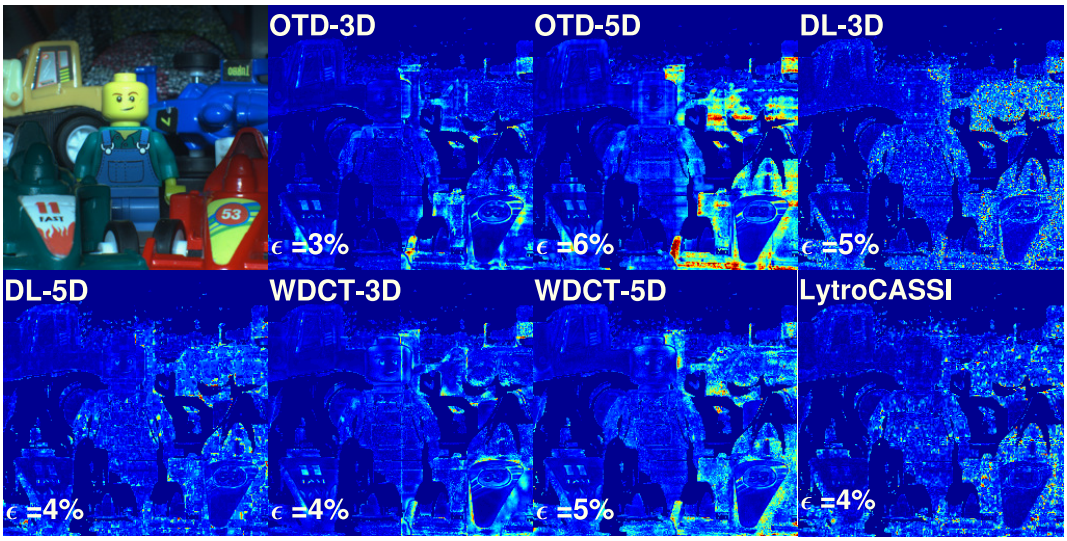


Figura 46. Reconstruction relative error (ϵ) for $C = 5\,\%$ and SNR=20 [dB].
DL-5D, WDCT-3D, OTD-3D and WDCT-5D methodologies, respectively.

To evaluate the quality of the angular reconstruction, Fig. 47 illustrates the depth map of the Cars SLF, calculated via the algorithm in Jeon y col., “Accurate depth map estimation from a lenslet light field camera”. Specifically, Fig. 47(A) shows a top-view of the setup. To evaluate the distance between the objects, the root mean squared error (RMSE) between the ground-truth depth maps and those attained with the OTD-3D/5D, DL-3D/5D, WDCT-3D/5D and the Lytro-CASSI methods is reported. It can be seen that better depth maps are estimated by the 5D parameterization methods compared to the 3D counterparts. Note however, that the proposed OTD-5D method overcomes the results of DL-5D, WDCT-5D, and Lytro-CASSI, respectively. These reconstructions were attained with a compression ratio $C = 9\,\%$ and a noise level SNR $= 50$ [dB].

To evaluate the spectral accuracy of the reconstructions, two different points of the Cars scene are illustrated in Fig. 48. This comparison shows that despite the good approximation obtained by all the approaches, the OTD-3D and DL-5D provide better approximations. In particular, the best results are obtained by the OTD-3D approach.

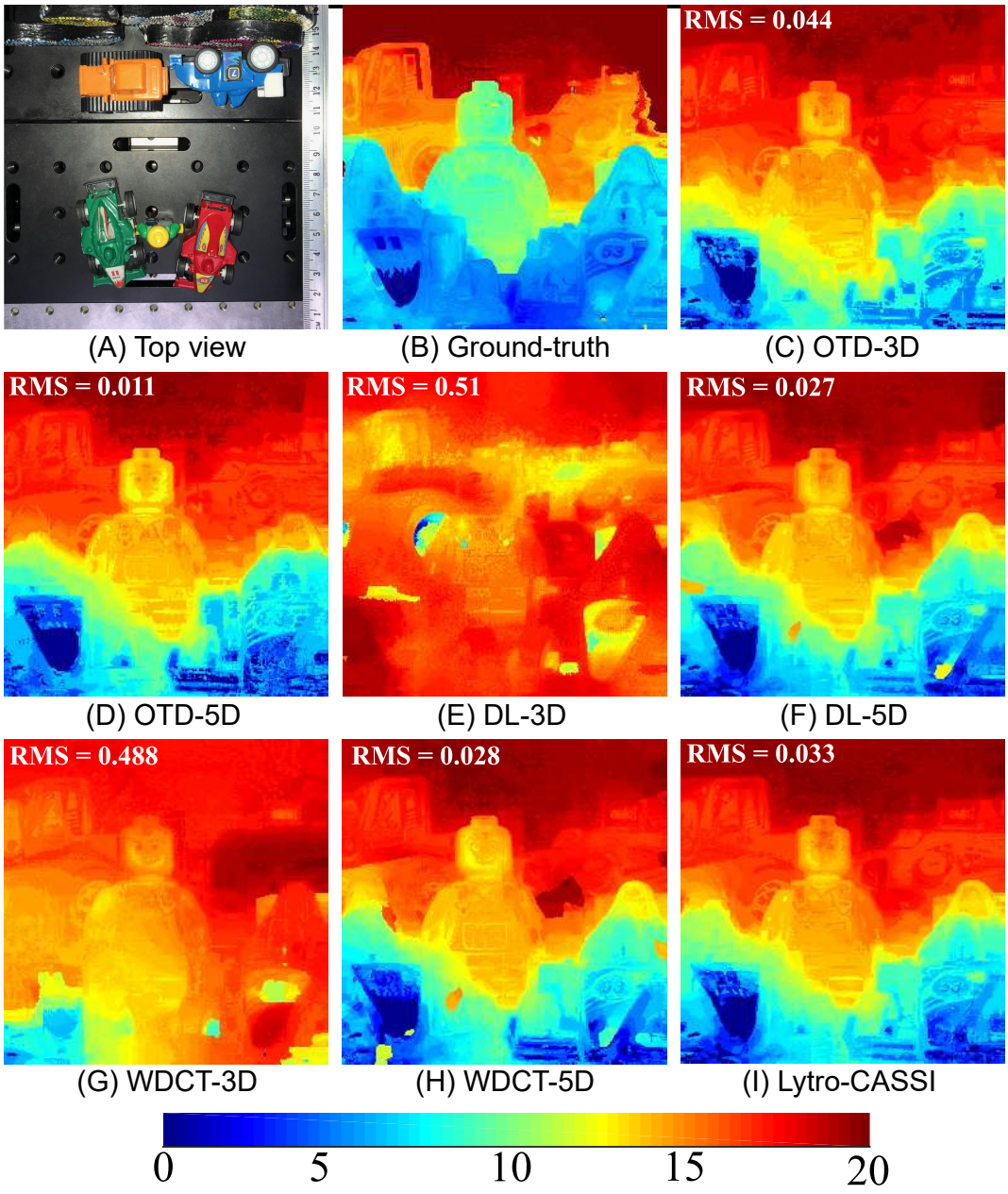


Figura 47. Reconstructed depth maps for the Cars SLF.

5.4. Proof of concept experiments

We built a testbed in our laboratory, so as to demonstrate the proposed ideas, through a proof-of-concept prototype, as shown in Fig. 49. This prototype uses a pinhole between a standard objective lens (Canon zoom lens EF-s 18-55mm 1:3.5-5.6 IS

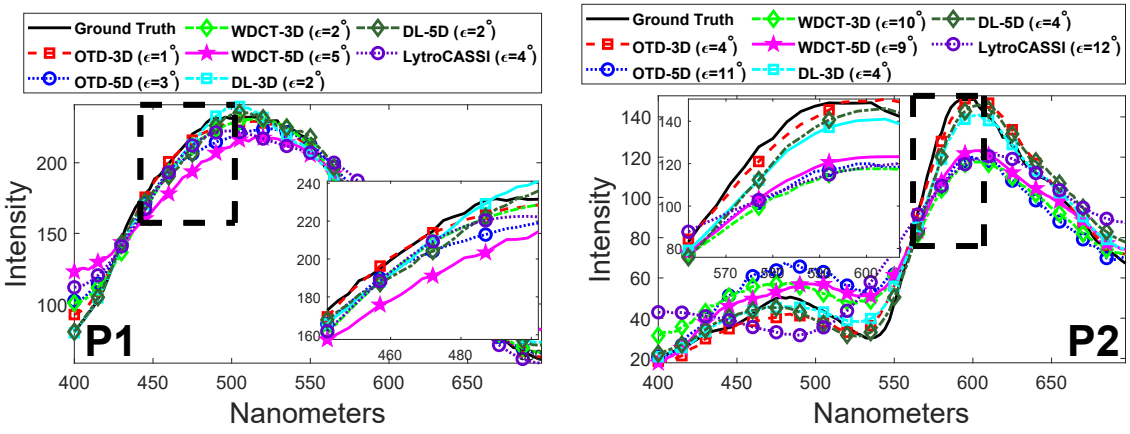
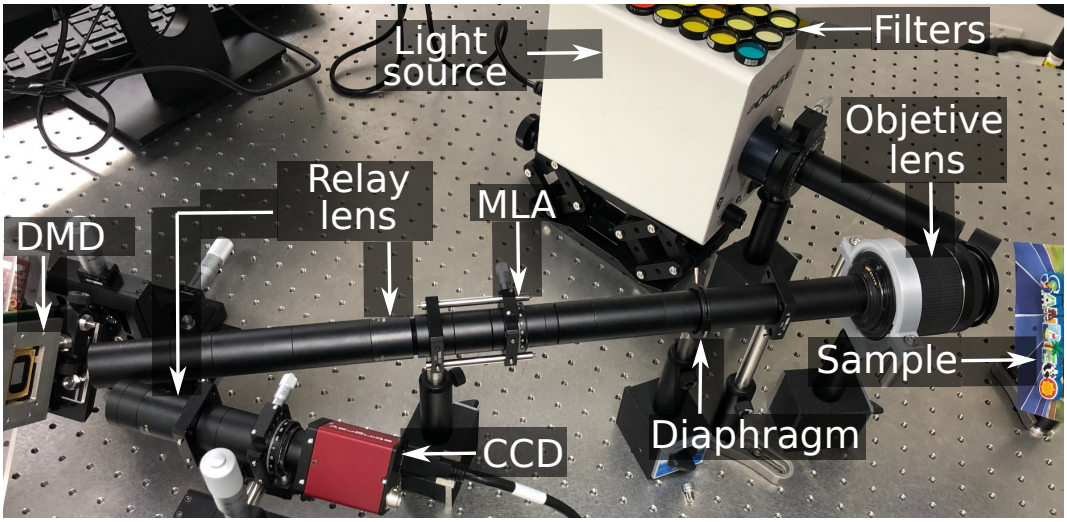
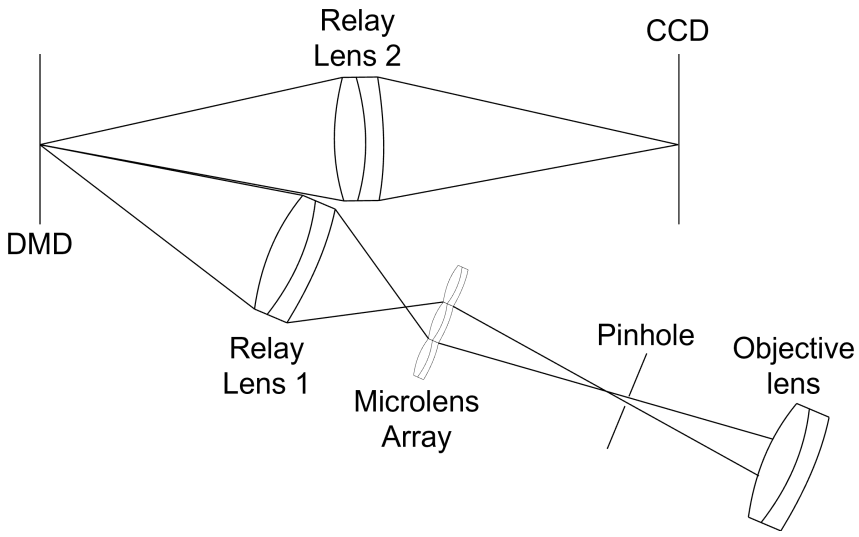


Figura 48. Spectral signatures for two spatial points of the Cars SLF, using $C = 5\,\%$ and SNR=20 [dB].

STM) and a microlens array (Thorlabs, MLA300-14AR) to create the light field image. The MLA has 27$\times$27 square refractive lenses in a 10$\times$10 mm area with a microlens interval of 300 μm and a focal length of 14.6 mm. Then, a standard relay lens (Thorlabs, MAP051919-A BBAR 400-700nm, with f1=f2=100 mm) is used to transfer the light field image to a digital micromirror device (Texas Instruments, D4120). Finally, a second relay lens (Thorlabs, MAP051919-A BBAR 400-700nm, with f1=f2=100 mm) is used to transfer the spatially modulated light field image to the CCD (Stingray F-080B, 4.65 μm pixel size). The experiments consider two target scenes named, Surfer and Boat, for which two compressive projections ($K = 2$) using complementary CCAs were acquired with this testbed. Note that, we have built on the ideas of Rueda et. al. Rueda, Arguello y Arce, "DMD-based implementation of patterned optical filter arrays for compressive spectral imaging" to emulate a CCA through a set of optical filters paired with black-and-white coded apertures, realized via a digital micro-mirror device. This DMD-based approach, in addition, allows the acquisition of multiple snapshots, as required by the proposed approach, without incurring in further calibration procedures. Figure 50 shows an example of the two complementary CCAs, emulated in the laboratory, as seen by the proposed system when they are illuminated with a white light source. Note that, the red and blue pixels refer



(A)



(B)

Figura 49. (A) A testbed implementation of the C^{3}SLFI architecture. (B) Geometric sketch of the optical system.

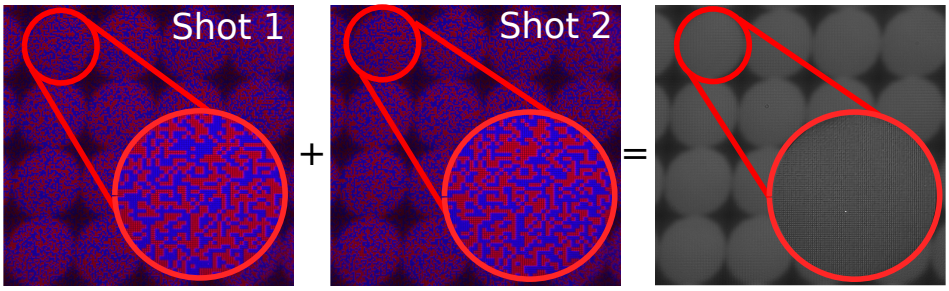


Figura 50. Examples of the complementary CCAs. Blue pixels represent low-pass filters whereas red are high-pass filters. Note that these coded apertures are considered complementary since they add-up to 1, per pixel, as shown on the right image.

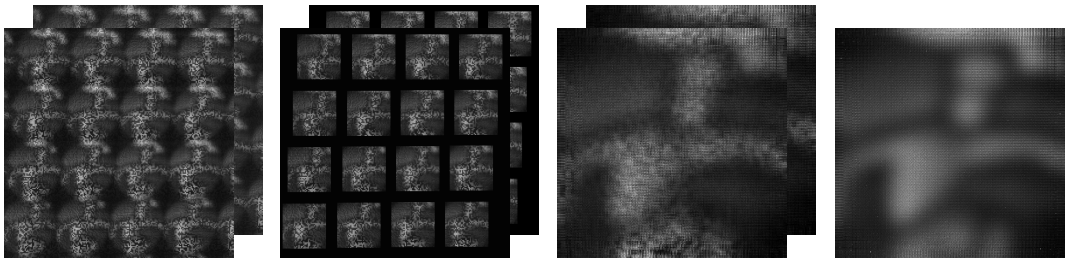


Figura 51. Raw data pre-processing including the 5D to 3D parameterization and the grayscale image estimation.

to low-pass (with spectral range $486 - 550nm$) and high-pass (with spectral range $558 - 671nm$) filters, respectively, and the addition of the complementary apertures entail an all-pass filter. The raw compressive measurements, acquired by the proposed testbed, exhibit a spatial resolution of 761×789 pixels, with $N_4 \times N_5 = 4 \times 4$ view angles, as shown in Fig. 51(A). These raw measurements are then processed to extract the region of interest for each view angle, as depicted in Fig. 51(B). Finally, based on Eq. 91, we estimate the 3D parameterization of the measurements $\underline{\tilde{\mathbf{Y}}}$ and the grayscale version $\mathbf{Z}$, as presented in Figs. 51(C) and 51(D), respectively. These processed data are then introduced into the reconstruction algorithm (using OTD-3D) and the obtained SLF is shown in Fig. 52. Note that in this figure, the RGB composite of the attained reconstruction is depicted, along with eighteen out of the twenty reconstructed spectral bands, that allow to confirm visually the spatial and

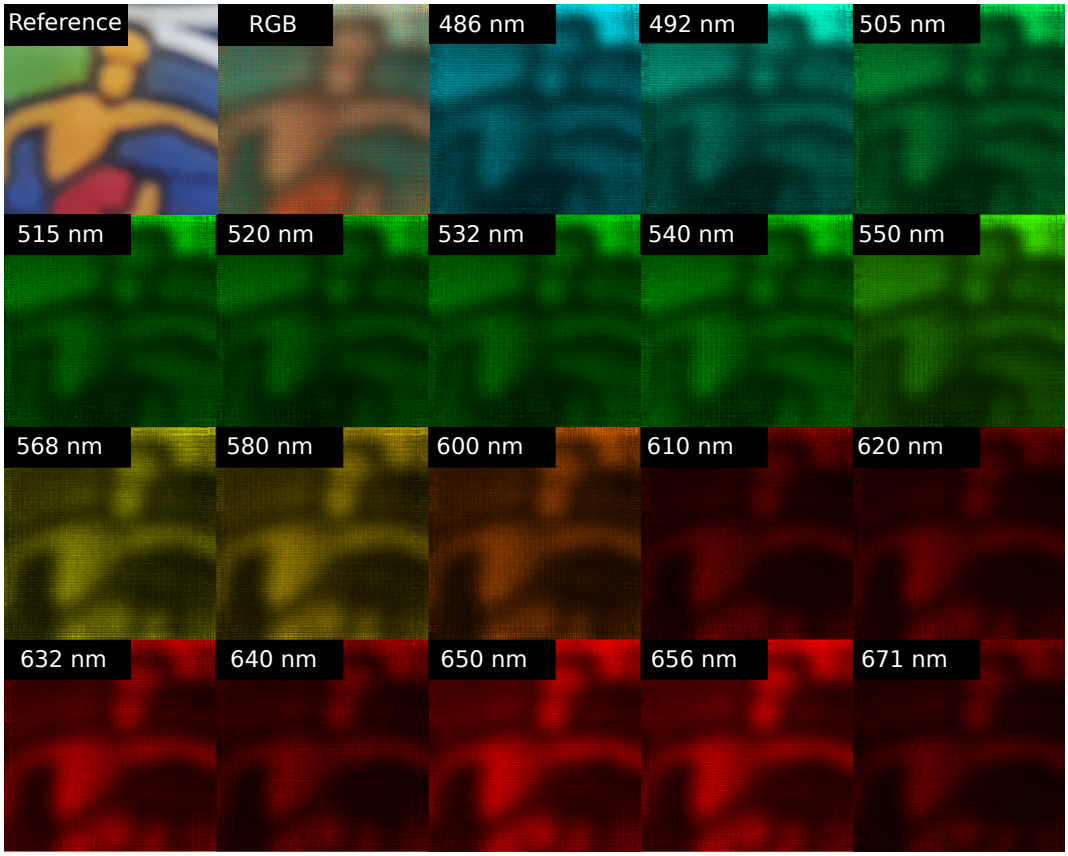


Figura 52. Surfer reconstruction mapped to an RGB profile, along with 18 out of the 20 reconstructed spectral bands.

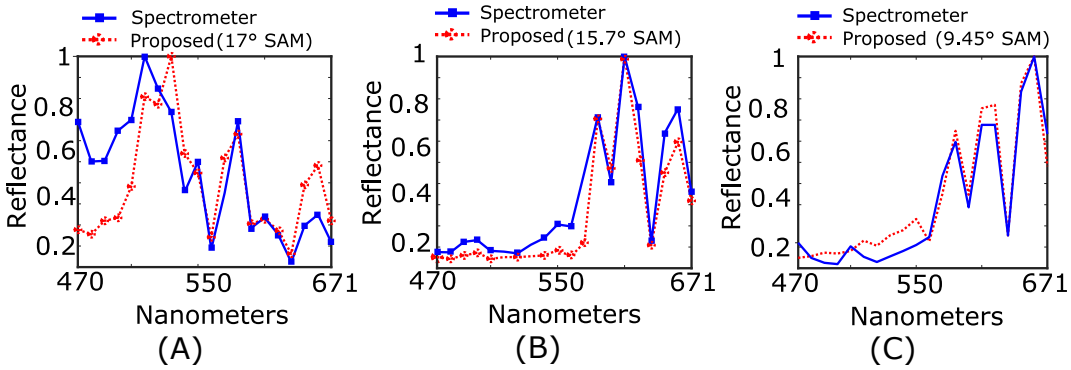


Figura 53. Spectral signatures for three spatial points of the Surfer target (A) blue background, (B) skin, and (C) red short.

spectral reconstruction accuracy. There, it can be noticed that the intensity of the red shorts and the blue background spans principally along the wavelengths in the range between 558-671 nm and 486-550 nm, respectively. To corroborate the accuracy of the spectral reconstruction, Fig. 53 shows the spectral signatures of three spatial points that were measured with a spectrometer (Ocean Optics Flame S-VIS-NIR-ES spectrometer), assumed to be the ground truth, and compared with the ones obtained with the proposed architecture using the OTD-3D representation basis. In this figure, it should be noted that the reconstructed spectral signatures resemble those obtained with the spectrometer, with SAM values smaller than 17°. Note that the **Algorithm 5** returns spectral data cubes with their spectral signatures without illumination corrections, i.e., in intensity. Therefore, an additional offline reconstruction step is the spectral reflectance estimation, see Supplemental material 9.

5.5. Conclusions

This chapter introduced the C^{3}SLFI system that aims to capture SLF images via compressed measurements. Alongside, a reconstruction algorithm, that exploits the inherent correlations of SLF by treating them as tensors, was proposed to reconstruct the SLF from few measurements. The usage of tensors promoted a low-complexity computational algorithm since the SLF is modeled as is, contrary to conventional matrix representations that rearrange the data such that their intrinsic properties are lost and require expensive matrix-vector operations. Furthermore, a 3D parameterization of the 5D SLF was proposed to promote higher-spatial correlations, and, an online-dictionary was learned, so that a representation basis for the data was estimated directly from the compressed measurements. This in turn, jointly with the 3D parameterization, entailed a much better sparse approximation of the SLF, which translated to better reconstruction quality. The proposed imaging system was demonstrated via simulations, and through a proof-of-concept implementation, which confirmed that the proposal represents an efficient alternative to capture SLF. As a chapter summary and to assess the C^{3}SLFI's pros and cons, we construct a radar chart with seven relevant CoSDI features: compactness, low costs, low-recovery complexity, spectral recovery, depth recovery, modularity, and encoding flexibility. Here the bottom and the top contours represent the worst and best scores, respectively, i.e., the ideal CoSDI system would have to reach the upper shell in the seven features. Specifically, Fig. 54 illustrates the C^{3}SLFI's radar chart, showing that C^{3}SLFI achieves high spectral, and depth reconstruction quality at the expense of compactness, encoding flexibility, and implementation costs.

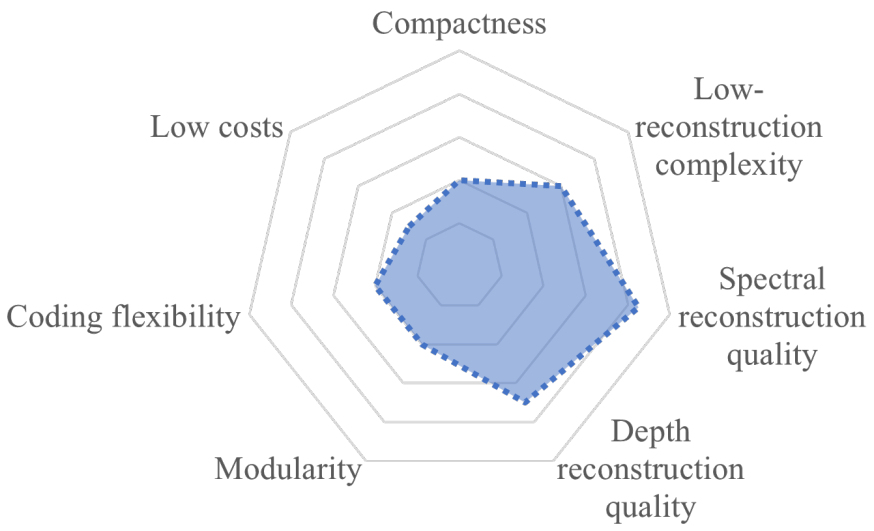


Figura 54. Chart graph for the C^{3}SLFI system

6. CS^{2}LI: Compressive Spectral Structured Light Imager

Structured light (SL)-based compressive spectral depth imagers (CoSDIs) are classified into stereoscopy multi-sensor or color filter mosaic detectors (CFMD) systems. The multi-tensor framework aims to split the depth and spectral sensing tasks into two independent CSI [e.g., CASSI sensor] and DI [e.g., grayscale sensor] imagers, where the recovered adjacent spectral/depth dimensions are merged into a spectral depth image Huijie Zhao y col. “A high throughput integrated hyperspectral imaging and 3D measurement system”. En: *Sensors* 18.4 (2018), pág. 1068. The CFMD-based SL approach aims to redesign the SL-based CoSDI sensing paradigm into a monocular system at the cost of spectral resolution heist20185d. The spatial and spectral degradation directly results from using a spectral mosaic composed of a fixed bandpass filter amount ($\sim$3 or 25 bandpass filters). Although the CFMD approach provides spectral-depth (SD) images with low spectral resolution and high spatial/depth resolution, this approach has become a popular solution in the computer vision community Heist y col., “5D hyperspectral imaging: fast and accurate measurement of surface shape and spectral characteristics using structured light”. To improve the spectral reconstruction performance, few works have introduced monocular SL-based CoSDI systems that integrate the coded aperture snapshot spectral imaging (CASSI) paradigm with the CFMD technology. This approach improves the spectral reconstruction performance at the cost of depth accuracy. Thus, SL-based CoSDI state-of-the-art works have focused on systems based on the CFMD sensing paradigm Chunyu Li y col. “Pro-cam ssfm: Projector-camera system for structure and spectral reflectance from motion”. En: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, págs. 2414-2423.

The positive impact of CCA to acquire measurements that preserves the encoded

depth features and the fact that the use of dispersive optical elements in CSI allows achieving spectral resolution improvements lead naturally to the question: ***Is it possible to develop a SL-based CoSDI system based on a CFMD and a single dispersive element that preserves the depth and spectral features without relying on stereoscopy setups?***

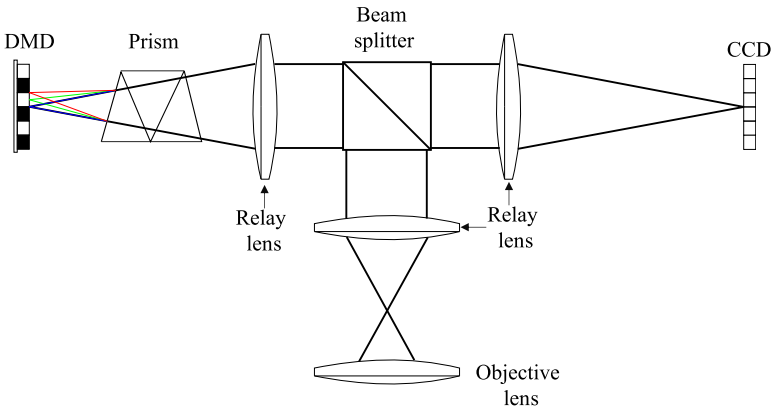
To answer this question, this chapter introduces a structured light-based compressive spectral imaging architecture that builds on the ideas of CCA and the dual-dispersive sensing geometry paradigm but relies on a single prism and sensor. Specifically, we proposed a double-dispersive CASSI system just using a single double-Amici prism and a coded mirror. In the proposed system, a lens is coupled to the double-Amici prism to relay a spectrally-dispersed version of the scene onto the coded mirror, which spatially encodes and reflects back the wavefront through the prism. The sensor, then, integrates the resulting dispersed-encoded-dispersed wavefront into a single measurement. This sensing process is carried out for each structured light projection pattern. We have coined the latter, compressive spectral structured light imager â€“ CS^{2}LI.

6.1. Continuous and discrete sensing model

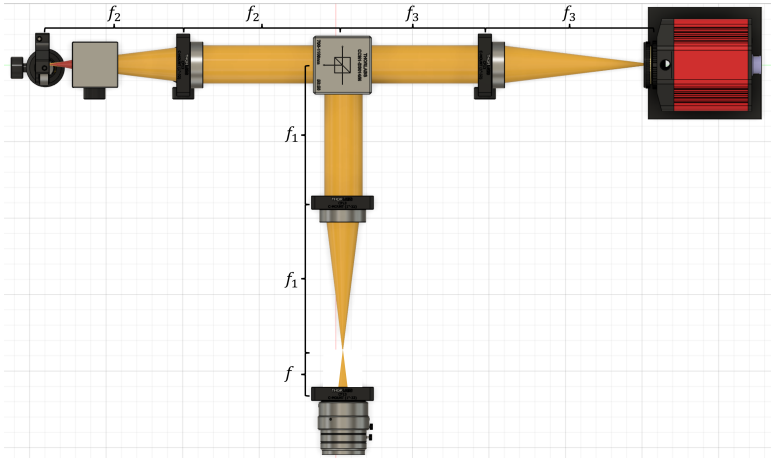
Compressive spectral depth imaging (CoSDI) systems based on active illumination approaches, such as structured light (SL) and time-of-flight (ToF) systems, are the favorite option in applications where high-precision surface shape estimation is required, in contrast to the introduced passive illumination-based CASSDI-DFA [refer to section 4] and C^{3}SLFI [refer to section 5] systems. Although ToF systems have shown reliability in acquiring high-resolution depth image, the needing for custom sensors or complex electronics makes it difficult or costly to modify these specialized devices to acquire additional dimensions. Thus, SL has become the most popular non-contact 3D shape measurement technique that overcomes ToF techniques in terms of simple hardware configuration, high measurement accuracy, high point density, high speed, low cost, and its versatility to be modified to acquire additional dimensions. Specifically, SL-based CoSDI systems aim to encode the 3D surface shape based on the deformations of light patterns [refer to section 2.2.2]. The light patterns' structure is constructed following a sinusoidal model (see Eqs. (11) and (21)), and the 3D surface recovery is obtained via an arctangent formulation along with spatial phase unwrapping methodology (see Eq. (18)).

Conventional SL-based CoSDI systems rely on multiple high-resolution colored-mosaic detectors. Few works have proposed replacing the CFMD-based sensor with a CASSI system to solve the color mosaic (CM) sensors dependence. The CASSI+SL-based CoSDI approach acquires spectral-depth information at the cost of depth reconstruction accuracy, associated with the spatial distortion introduced by the dispersive element. To overcome these shortcomings, CS^{2}LI is inspired by the dual-dispersive CASSI sensing paradigm, and the customized double Amici prism (see **Appendix 9**) to build a CoSDI system that relies on a single prism and sensor.

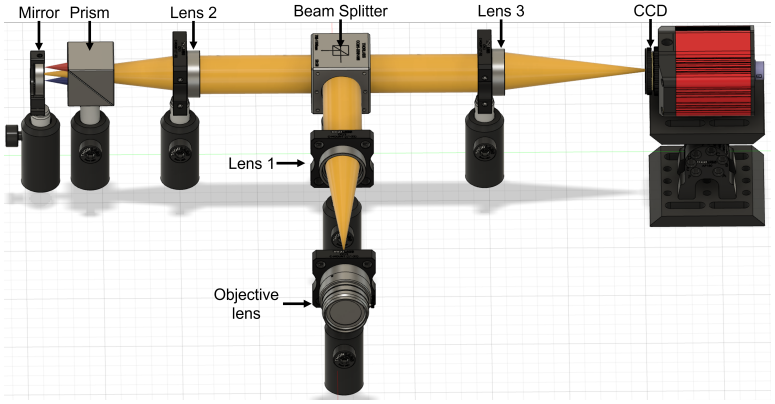
The CS^{2}LI's propagation model is divided into three main stages: spectral dispersion, spatially coding, and a second spectral dispersion. The first stage relies on



(A)



(B)



(C)

Figura 55. (A) Spectral and (B) depth reconstruction results obtained for the optical systems SCCI Ray tracing sketch of the proposed CS^2L architecture. (B)-(C) Upper and lateral view where the distance configuration and the optical element labels are specified, respectively.

an objective lens that images the scene onto the input focal plane of a double 4f system composed of three lenses with a beam splitter placed at its Fourier plane, as is shown in Fig. 55. The first relay lens propagates the incoming wavefront through the beam splitter until the second relay lens, which propagates the incoming wavefront through a prism before focusing onto the CCA attached to a broadband dielectric mirror. Second, taking advantage of the mirror surface, the dispersed-and-coded spectral wavefront is reflected back to the prism and then to the second relay lens. Here, the prism introduce a second spectral dispersion, which results in a coded-in-spectrum wavefront. Finally, the second relay lens propagates the dispersed-encoded-dispersed wavefront through the beam splitter redirecting this to a third relay lens that focuses the wavefront onto the CCD sensor.

6.1.1. Continuous sensing model Let $f_o(x, y, \lambda)$ be the object, where (x, y) index the spatial coordinates and λ index the wavelength. The object's depth map is encoded by illuminating it with a structured light pattern, where the resulting pattern deformations are associated to the object shape. The pattern deformation can be modeled as a phase modulation, where the object's shape acts as the phase mask. Based on Eq. (11), the object shape-modulated structured light can be modeled as

$$p(x, y, \lambda, t) = f_o(x, y, \lambda) \left[\cos\left(\varphi(x, y) + \omega_0 t\right) + b\right] + a(x, y), \tag{107}$$

where $\varphi(x, y)$ is the object phase/shape information, w_0 represents the temporal-carrier, $a(x, y)$ is the background, and b is a balance constant that prevents $p(x, y, \lambda, t)$ achieve negative values with $b = 1$. First, an objective lens forms an image in an intermediate image plane, resulting in the wave field,

$$f_1(x', y', \lambda, t) = \iint \gamma(\lambda) h_1(x - x', y - y', \lambda) p(x, y, \lambda, t) dx dy, \tag{108}$$

where $h_1(\cdot)$ is the intensity PSF of the objective lens and $\gamma(\lambda)$ represent the spectral source density. The standard objective lens forms image in the input plane of a 4f system, which propagates the incoming wavefront through a prism that introduces a wavelength-dependent horizontal shifting to form a dispersed image version over a coded aperture attached to a mirror (or DMD). The resulting dispersed-encoded wavefront can be seen in the detector plane (immediately before the sensor) as

$$f_2(u,v,\lambda,t) = k(u,v,\lambda) \iint h_2(x'-u-S_1(\lambda), y'-v, \lambda) f_1(x',y',\lambda,t) dx'dy', \quad (109)$$

with

$$k(u,v,\lambda) = \frac{1}{2} \sum_{i'_1,i_2,i_3} \text{rect}\left(\frac{u}{\Delta_c} - i_1, \frac{v}{\Delta_c} - i_2, \frac{\lambda}{\Delta_\lambda} - i_3\right) C_{i_1,i_2,i_3}, \quad (110)$$

where rect$(\cdot)$ represents the rectangle function, $S_1(\lambda)$ is a wavelength-dependent function for the propagation from the relay lens to the mirror; Δ_c denotes the pixel size, Δ_λ denotes the spectral bandwidth, $\mathbf{C} \in \mathbb{R}^{N_y \times (N_x+N_\lambda-1) \times N_\lambda}$ is a color-coded aperture with $C_{i'_1,i_2,i_3} \in \{0,1\}$, and $i'_1 = \{0,\ldots,(N_x+N_\lambda-2)\}$, $i_2 = \{0,\ldots,N_y-1\}$ and $i_3 = \{0,...,N_\lambda-1\}$ indexing the rows, columns, and spectral bands, respectively. Then, the color-coded mirror reflects back the dispersed-modulated wavefront to the prism, which introduce a second spectral dispersion. The resulting dispersed-coded-dispersed wavefront is propagated and focused into a detector array. This sensing process can be expressed as

$$f_3(x'',y'',\lambda,t) = \iint f_2(u,v,\lambda,t) h_3(u-x''+S_2(\lambda), v-y'', \lambda) du dv \quad (111)$$

where h_3 represents the intensity PSF and $S_2(\lambda)$ is a wavelength-dependent function for the propagation from the mirror to the relay lens (i.e., backpropagating through

the prism). Finally, the measurement at the $(i_1, i_2)^{th}$ pixel is represented by

$$g_{i_1,i_2} = \int_{\Lambda} \rho(\lambda) \iint f_3(x'', y'', \lambda) \text{rect}\left(\frac{x''}{\Delta_d} - i_1, \frac{y''}{\Delta_d} - i_2\right) dx'', dy'' d\lambda, \tag{112}$$

where Λ represents the wavelength axis over the spectral range, Δ_p represents the pixel's side length, and $\rho(\lambda)$ represents the detector's normalized quantum efficiency.

6.1.2. Discrete sensing model Based on Eq. (112), the discrete sensing model of the proposed system can be expressed as,

$$G_{i_1,i_2,i_4} = \sum_{i_3=0}^{N_\lambda - 1} C_{i_1,(i_2-i_3),i_3} P_{i_1,i_2,i_3,i_4} + \epsilon_{i_1,i_2,i_4}, \tag{113}$$

where $\mathbf{G} \in \mathbb{R}^{N_x \times N_y \times N_t}$ represents the compressed measurement with N_t as the total amoung of pattern projections, $\mathbf{P} \in \mathbb{R}^{N_x \times N_y \times N_\lambda \times N_t}$ is the 4D datacube that contains the spectral response of the scene for the N_t structured light projections, and $\boldsymbol{\epsilon} \in \mathbb{R}^{N_x \times N_y \times N_t}$ represents the noise. Based on the discrete sensing model, the linear sensing model can be expressed as

$$\mathbf{g} = \mathbf{H}\mathbf{p} + \boldsymbol{\epsilon}, \tag{114}$$

where $\mathbf{g} \in \mathbb{R}^m$ is the vectorial version of the compressed measurements with $m = N_x N_y N_t$, $\mathbf{p} \in \mathbb{R}^n$ is the vectorial representation of $\mathbf{P}$ with $n = N_x N_y N_\lambda N_t$, $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the sensing matrix. In particular, $\mathbf{H}$ can be expressed as $\mathbf{H} = (\mathbf{I}_{N_t} \otimes \mathbf{A})$ with $\otimes$ representing the Kronecker product operator, $\mathbf{I}_{N_t} \in \mathbb{R}^{N_t \times N_t}$ is an identity matrix, and $\mathbf{A} \in \mathbb{R}^{\frac{m}{N_t} \times n}$ represents the sensing matrix for a fixed coded aperture $\mathbf{C}$ and its entries

are given by

$$A_{i,j} = \begin{cases} c_u, & \text{if } i = mod(j, N_x N_y) \\ 0, & \text{otherwise} \end{cases}, \tag{115}$$

for $u = mod(j, N_x N_y) + \lfloor \frac{j}{N_x N_y} \rfloor (N_y (N_x + N_\lambda - 1))$, $j = \{0, ..., \frac{m}{N_t} - 1\}$, $i = \{0, ..., N_x N_y - 1\}$, where c_u are the entries of $\mathbf{c} \in \mathbb{R}^{N_x(N_t+N_\lambda-1)\times 1}$, the vectorized version of the coded aperture $\mathbf{C}$ defined in Eq. (110).

6.2. Reconstruction algorithm

The CSI principle aims to recover a time-invariant three-dimensional datacube from two-dimensional measurements, i.e., without illumination variations. Thus, the SL-based CoSDI frameworks tackled the recovery problem by reconstructing the spectral datacube independently for each structured light patterned measurement. To overcome this shortcoming, we propose a hierarchical-based Plug-and-Play alternating direction method of multipliers (PnP-ADMM) reconstruction algorithm. Specifically, the proposed reconstruction methodology relies on complementary structured light patterns and a fixed CA to recast the traditional four-dimensional datacube (i.e., $\mathbf{P}$) reconstruction problem into two low-complex three-dimensional reconstruction subproblems: the spectral datacube $\mathbf{F} \in \mathbb{R}^{N_x \times N_y \times N_\lambda}$ and the grayscale stripe patterns $\mathbf{Z} \in \mathbb{R}^{N_x \times N_y \times N_T}$. Then, the depth map $\mathbf{\Phi}$ [or depth map] is estimated directly from **Z** via an N-step least-square methodology followed by a third-party unwrapping algorithm. Specifically, this thesis used the Phase Unwrapping Max-flow algorithm (PUMA) Gonçalo Valadao y José Bioucas-Dias. “PUMA: phase unwrapping via max flows”. En: *Proceedings of Conference on Telecommunicationsâ€”ConfTele*. Citeseer. 2007, págs. 609-612. This algorithm relies on an energy minimization framework for phase unwrapping, where the minimization is carried out by a sequence

of max-flow/min-cut calculations. The objective functions considered are first-order Markov random fields with pairwise interactions. The associated energy is, therefore, a generalization of the classical norm used in phase unwrapping.

6.2.1. Traditional ADMM optimization problem As a baseline and based on Eq. (114), we introduce a traditional PnP-ADMM optimization problem to reconstruct the four-dimensional datacube, modeled as

$$\arg\min_{\mathbf{p},\mathbf{w}} \quad \|\mathbf{g}-\mathbf{H}\mathbf{p}\|_2^2+\|\mathbf{p}-\mathbf{w}-\boldsymbol{\eta}\|_2^2+\mu\mathcal{R}(\mathbf{w}), \tag{116}$$

where $\mu \geq 0$ is a regularization parameter, $\mathbf{w} \in \mathbb{R}^n$ is an auxiliary parameter, $\mathcal{R}(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is a prior function that plays the role of an arbitrary regularizer, and $\boldsymbol{\eta} \in \mathbb{R}^n$ is the scaled dual variable. Inspired by Stanley H Chan, Xiran Wang y Omar A Elgendy. “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications”. En: *IEEE Transactions on Computational Imaging* 3.1 (2016), págs. 84-98, the optimization problem in Eq. (116) can be decoupled in the following two subproblems

$$\arg\min_{\mathbf{p}} \quad \|\mathbf{g}-\mathbf{H}\mathbf{p}\|_2^2+\tfrac{\rho}{2}\|\mathbf{p}-\mathbf{w}^{\iota}-\boldsymbol{\eta}^{\iota}\|_2^2, \tag{117}$$

$$\arg\min_{\mathbf{w}} \quad \mathcal{R}(\mathbf{w}^{\iota})+\tfrac{\rho}{2}\|\mathbf{p}^{\iota+1}-\mathbf{w}^{\iota}-\boldsymbol{\eta}^{\iota}\|_2^2, \tag{118}$$

where $\boldsymbol{\eta}^{\iota+1}=\boldsymbol{\eta}^{\iota}+\rho\left(\mathbf{p}^{\iota+1}-\mathbf{w}^{\iota+1}\right)$, $\rho \geq 0$ is the regularization parameter, $(^{\iota})$ denotes the superscript of the iteration number. The main drawback of this approach –i.e., Eq. (116)– lies in the computational complexity that grows linearly with the total number of structured light patterns. Then, the spectral datacube $\mathbf{f} \in \mathbb{R}^{N_x N_y N_\lambda}$ and the phase

information $\boldsymbol{\varphi} \in \mathbb{R}^{N_x N_y}$ can be estimated by closed-form solutions

$$\mathbf{f} = \mathbf{D}_t \mathbf{p}, \tag{119}$$

and

$$\boldsymbol{\varphi} = \phi(\mathbf{D}_\lambda \mathbf{p}), \tag{120}$$

respectively, where $\mathbf{D}_t \in \mathbb{R}^{\ell_\lambda \times n}$ is a temporal decimation matrix with $\mathbf{D}_t = \left(\mathbf{1}_{N_t}^T \otimes \mathbf{I}_{\ell_\lambda}\right)$ and $\ell_\lambda = N_x N_y N_\lambda$, $\mathbf{D}_\lambda \in \mathbb{R}^{\ell_t \times n}$ is a spectral decimation matrix with $\mathbf{D}_\lambda = \mathbf{I}_{N_t} \otimes \left(\mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{N_x N_y}\right)$ and $\ell_t = N_x N_y N_t$, and $\phi(\cdot) : \mathbb{R}^{N_x N_y N_t} \to \mathbb{R}^{N_x N_y}$ is a phase-unwrapping algorithm (PUMA) that returns a depth map wrapped version. Note that $\mathbf{f}$ can be obtained via Eq. (119) when the structured light patterns are complementary.

6.2.2. Hierarchical-based PnP-ADMM reconstruction algorithm To circumvent the reconstruction of $\mathbf{p}$, the first step is to recast and split the sensing model in Eq. (114) – for a four-dimensional datacube – into two sensing models – for two three-dimensional datacube – as follows

$$\mathbf{g} = \mathbf{H}\mathbf{D}_t^T \mathbf{f} + \boldsymbol{\epsilon}, \tag{121}$$

and

$$\mathbf{g} = \mathbf{H}\mathbf{D}_\lambda^T \mathbf{z} + \boldsymbol{\epsilon}, \tag{122}$$

where $\mathbf{f} \in \mathbb{R}^{\ell_\lambda}$ is the vector version of $\mathbf{F}$ and $\mathbf{z} \in \mathbb{R}^{\ell_t}$ is the vector version of $\mathbf{Z}$. Based on the Eqs. (121) and (122), we proceed to establish the optimization problem

$$\begin{aligned} \underset{\mathbf{f},\mathbf{z},\boldsymbol{\theta}_f,\boldsymbol{\theta}_z}{\operatorname{argmín}} \quad & \|\mathbf{g} - \mathbf{H}\mathbf{D}_\lambda^T\mathbf{z}\|_2^2 + \|\mathbf{g} - \mathbf{H}\mathbf{D}_t^T\mathbf{f}\|_2^2 + h_f(\boldsymbol{\theta}_f) + h_z(\boldsymbol{\theta}_z) \\ \text{Subject to} \quad & \mathbf{D}_t\mathbf{f} = \mathbf{D}_\lambda\mathbf{z},\ \boldsymbol{\theta}_f = \mathbf{f},\ \boldsymbol{\theta}_z = \mathbf{z}, \end{aligned} \tag{123}$$

where $\{h_f(\cdot), h_z(\cdot)\}$ are two prior restriction functions for which their structure do not need to be specified in the PnP-ADMM approaches. Following the ADMM paradigm, Eq. (123) is split into four optimization subproblems – two for the spectral image and two for the grayscale structured light patterns– which are solved into three stages: spectral datacube reconstruction, grayscale structured fringe patterns estimation, and dual-variables update.

Spectral datacube reconstruction The optimization problem for $\mathbf{f}$ is composed of three ℓ_2 terms: the fidelity term, the augmented Lagrangian term for $\boldsymbol{\eta}_2$, and the spatial regularization term between $\mathbf{f}$ and $\mathbf{z}$. On the other hand, the optimization problem for $\boldsymbol{\theta}_f$ is composed of a ℓ_2-norm and an arbitrary restriction function $h_f(\cdot)$. Particularly, these subproblems are expressed as

$$\arg\min_{\mathbf{f}} \quad \|\mathbf{g} - \mathbf{H}\mathbf{D}_t^T\mathbf{f}\|_2^2 + \alpha_1\|\boldsymbol{\theta}_f^\iota - \mathbf{f} - \boldsymbol{\eta}_1^\iota\|_2^2 + \alpha_2\|\mathbf{D}_\lambda\mathbf{z}^\iota - \mathbf{D}_\lambda\mathbf{f} - \boldsymbol{\eta}_2^\iota\|_2^2, \tag{124}$$

and

$$\arg\min_{\boldsymbol{\theta}_f} \quad \alpha_1\|\boldsymbol{\theta}_f - \mathbf{f}^\iota - \boldsymbol{\eta}_1^\iota\|_2^2 + h_f(\boldsymbol{\theta}_f), \tag{125}$$

where $\{\alpha_1, \alpha_2\} > 0$ are the augmented Lagrangian weighting related to the spectral datacube, and $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2\}$ are the dual variables with $\boldsymbol{\eta}_1 \in \mathbb{R}^{\ell_\lambda}$ and $\boldsymbol{\eta}_2 \in \mathbb{R}^{N_x N_y}$, respectively. Note that in Eq. (124), the analytical inverse model of $\mathbf{f}$ refers to a quadratic

problem, while in Eq. (125) the analytical model of $\boldsymbol{\theta}_f$ is a denoising problem. To solve the first inverse model in Eq. (124), we exploit the Sherman Woodbury Morrison (SWM) matrix inversion lemma and the full-column rank properties to obtain the closed-form solution that simultaneously involves the sensing process, the sensing matrix conditionality, and the gradient solution:

$$\mathbf{f}^{\iota+1} = \left[\mathbf{B}^T\mathbf{B}\right]^{-1}\left[\mathbf{D}_t\mathbf{H}^T\mathbf{g} + \alpha_1\left(\boldsymbol{\theta}_p^{\iota} - \boldsymbol{\eta}_1^{\iota}\right) + \alpha_2\left(\mathbf{D}_\lambda\mathbf{z}^{\iota} - \boldsymbol{\eta}_2^{\iota}\right)\right], \tag{126}$$

where $\mathbf{B} \in \mathbb{R}^{(2\ell_\lambda+m)\times\ell_\lambda}$ is an auxiliary matrix with $\mathbf{B} = \left[\left(\mathbf{H}\mathbf{D}_t\right)^T, \sqrt[2]{\alpha_1}\mathbf{I}_{\ell_\lambda\times\ell_\lambda}, \sqrt[2]{\alpha_2}\mathbf{D}_\lambda^T\right]^T$. To solve the second inverse model in Eq. (125) we use an off-the-shelf denoising algorithm, to yield

$$\boldsymbol{\theta}_f^{\iota+1} = \mathcal{D}_\sigma\left(\mathbf{f}^{\iota+1} + \frac{1}{\alpha_1}\boldsymbol{\eta}_1^{\iota}\right), \tag{127}$$

where $\mathcal{D}_\sigma(\cdot)$ denotes the denoiser being used with σ being the standard deviation of the assumed additive colored Gaussian noise Chan, Wang y Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications".

Grayscale structured patterns estimation Similarly to the spectral datacube recovery problem, the optimization problem for $\mathbf{z}$ is composed of three ℓ_2 terms: the fidelity term, the augmented Lagrangian term for $\boldsymbol{\eta}_1$, and the spatial regularization between $\mathbf{f}$ and $\mathbf{z}$. The optimization problem for $\boldsymbol{\theta}_z$ is composed of an arbitrary resctriction function $h_z(\cdot)$ and a ℓ_2-norm. Specifically, the optimization problems are mathematically expressed as

$$\underset{\mathbf{z}}{\arg\min} \quad \|\mathbf{g} - \mathbf{H}\mathbf{D}_\lambda^T\mathbf{z}\|_2^2 + \alpha_3\|\boldsymbol{\theta}_z^{\iota} - \mathbf{z} - \boldsymbol{\eta}_3^{\iota}\|_2^2 + \alpha_4\|\mathbf{D}_\lambda\mathbf{f}^{\iota} - \mathbf{D}_z\mathbf{z} - \boldsymbol{\eta}_4^{\iota}\|_2^2, \tag{128}$$

and

$$\arg\min_{\boldsymbol{\theta}_z} \quad \alpha_3\|\boldsymbol{\theta}_z - \mathbf{z}^{\iota+1} - \boldsymbol{\eta}_3^{\iota}\|_2^2 + h_z(\boldsymbol{\theta}_z), \tag{129}$$

where $\{\alpha_3, \alpha_4\} > 0$ are the augmented Lagrangian weighting terms related to $\mathbf{z}$, and $\{\boldsymbol{\eta}_3, \boldsymbol{\eta}_4\}$ are the dual variables with $\boldsymbol{\eta}_3 \in \mathbb{R}^{\ell_t}$ and $\boldsymbol{\eta}_4 \in \mathbb{R}^{N_x N_y}$, respectively. Inspired by Sherman Woodbury Morrison (SWM) and following the same process developed in Eq. (126), the closed-form solution for Eq. (128) can be expressed as

$$\mathbf{z}^{\iota+1} = \left[\mathbf{E}^T\mathbf{E}\right]^{-1} \cdot \left[\mathbf{D}_\lambda\mathbf{H}^T\mathbf{g} + \alpha_3\mathbf{I}\left(\boldsymbol{\theta}_z^{\iota} - \boldsymbol{\eta}_3^{\iota}\right) + \alpha_4\mathbf{D}_t^T\left(\mathbf{D}_\lambda\mathbf{f}^{\iota} - \boldsymbol{\eta}_4^{\iota}\right)\right], \tag{130}$$

where $\mathbf{E} \in \mathbb{R}^{(2\ell_t+m)\times\ell_t}$ is an auxiliary matrix with $\mathbf{E} = [\left(\mathbf{H}\mathbf{D}_\lambda^T\right)^T, \sqrt[2]{\alpha_3}\mathbf{I}_{\ell_t\times\ell_t}, \sqrt[4]{\alpha_4}\mathbf{D}_t^T]^T$. Equation (129) is solved via an off-the-shelf denoising algorithm, which is expressed as

$$\boldsymbol{\theta}_z^{\iota+1} = \mathcal{D}_{\sigma_z}\left(\mathbf{z}^{\iota+1} + \frac{1}{\alpha_3}\boldsymbol{\eta}_3^{\iota}\right), \tag{131}$$

where $\mathcal{D}_{\sigma_z}(\cdot)$ and σ_z are the denoiser and the standard deviation associated with the grayscale structured patterns recovery problem.

Dual variables update Finally, the dual variables are updated as follows:

$$\begin{cases} \boldsymbol{\eta}_1^{\iota+1} = \boldsymbol{\eta}_1^{\iota} - \left(\boldsymbol{\theta}_f^{\iota+1} - \mathbf{f}^{\iota+1}\right), \\ \boldsymbol{\eta}_2^{\iota+1} = \boldsymbol{\eta}_2^{\iota} - \left(\mathbf{D}_t\mathbf{z}^{\iota+1} - \mathbf{D}_\lambda\mathbf{f}^{\iota+1}\right), \\ \boldsymbol{\eta}_3^{\iota+1} = \boldsymbol{\eta}_3^{\iota} - \left(\boldsymbol{\theta}_z^{\iota+1} - \mathbf{z}^{\iota+1}\right) \\ \boldsymbol{\eta}_4^{\iota+1} = \boldsymbol{\eta}_4^{\iota} - \left(\mathbf{D}_\lambda\mathbf{f}^{\iota+1} - \mathbf{D}_t\mathbf{z}^{\iota+1}\right) \end{cases} \tag{132}$$

which are updated at the end of each iteration.

Algorithm 6 PnP-ADMM algorithm

Input: $\mathbf{g} \in \mathbb{R}^m$, $\mathbf{H} \in \mathbb{R}^{m \times n}$, $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} \in [0, 1]$
Result: $\boldsymbol{\theta}_f$, $\boldsymbol{\theta}_z$
1. $\{\mathbf{f}, \boldsymbol{\theta}_f, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\} \leftarrow \mathbf{0} \in \mathbb{R}^{\ell_\lambda \times 1}$
2. $\{\mathbf{z}, \boldsymbol{\theta}_z, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4\} \leftarrow \mathbf{0} \in \mathbb{R}^{\ell_t \times 1}$
3. $\mathbf{D}_\lambda = \mathbf{I}_{N_\lambda} \otimes \left(\mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{N_x N_y}\right)$;
4. $\mathbf{D}_t = \left(\mathbf{1}_{N_t}^T \otimes \mathbf{I}_{\ell_\lambda}\right)$;
5. $\mathbf{B} = \left[\left(\mathbf{H}\mathbf{D}_t\right)^T, \sqrt[2]{\alpha_1}\mathbf{I}_{\ell_\lambda \times \ell_\lambda}, \sqrt[2]{\alpha_2}\mathbf{D}_\lambda^T\right]^T$
6. $\mathbf{E} = [\left(\mathbf{H}\mathbf{D}_\lambda^T\right)^T, \sqrt[2]{\alpha_3}\mathbf{I}_{\ell_t \times \ell_t}, \sqrt[4]{\alpha_4}\mathbf{D}_t^T]^T$

for $\iota \leftarrow 1$, *Iter* **do**

7. $\boldsymbol{\gamma}_1 = \left[\mathbf{D}_t\mathbf{H}^T\mathbf{g} + \alpha_1\left(\boldsymbol{\theta}_p^\iota - \boldsymbol{\eta}_1^\iota\right) + \alpha_2\left(\mathbf{D}_\lambda\mathbf{z}^\iota - \boldsymbol{\eta}_2^\iota\right)\right]$
8. $\boldsymbol{\gamma}_2 = \left[\mathbf{D}_\lambda\mathbf{H}^T\mathbf{g} + \alpha_3\mathbf{I}\left(\boldsymbol{\theta}_z^\iota - \boldsymbol{\eta}_3^\iota\right) + \alpha_4\mathbf{D}_t^T\left(\mathbf{D}_\lambda\mathbf{f}^\iota - \boldsymbol{\eta}_4^\iota\right)\right]$
9. $\mathbf{F}^{\iota+1} = \left[\mathbf{B}^T\mathbf{B}_1\right]^{-1}\boldsymbol{\gamma}_1$
10. $\mathbf{z}^{\iota+1} = \left[\mathbf{E}^T\mathbf{E}\right]^{-1}\boldsymbol{\gamma}_2$
11. $\boldsymbol{\theta}_f^{\iota+1} = \mathcal{D}_\sigma\left(\mathbf{f}^{\iota+1} + \frac{1}{\alpha_1}\boldsymbol{\eta}_1^\iota\right)$
12. $\boldsymbol{\theta}_z^{\iota+1} = \mathcal{D}_{\sigma_z}\left(\mathbf{z}^{\iota+1} + \frac{1}{\alpha_3}\boldsymbol{\eta}_3^\iota\right)$
13. $\boldsymbol{\eta}_1^{\iota+1} = \boldsymbol{\eta}_1^\iota - \left(\boldsymbol{\theta}_f^{\iota+1} - \mathbf{f}^{\iota+1}\right)$
14. $\boldsymbol{\eta}_2^{\iota+1} = \boldsymbol{\eta}_2^\iota - \left(\mathbf{D}_t\mathbf{z}^{\iota+1} - \mathbf{D}_\lambda\mathbf{f}^{\iota+1}\right)$
15. $\boldsymbol{\eta}_3^{\iota+1} = \boldsymbol{\eta}_3^\iota - \left(\boldsymbol{\theta}_z^{\iota+1} - \mathbf{z}^{\iota+1}\right)$
16. $\boldsymbol{\eta}_4^{\iota+1} = \boldsymbol{\eta}_4^\iota - \left(\mathbf{D}_\lambda\mathbf{f}^{\iota+1} - \mathbf{D}_t\mathbf{z}^{\iota+1}\right)$

end

6.2.3. Depth map recovery via N-step least-square PSA principle Inspired by the PSA principle (see section 2.2.2), the depth map can be calculated directly from $\mathbf{z}$ by

$$\tilde{\boldsymbol{\Phi}} = \arctan\left[\left(\mathbf{D}_t\mathbf{B}_S\mathbf{z}\right) \circ \left(\mathbf{D}_S\mathbf{B}_C\mathbf{z}\right)^{\circ -1}\right], \tag{133}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N_x \times N_y}$ is the discontinuous phase, $\circ$ represents the Hadamard product, $\circ - 1$ is the inverse Hadamard product, $\mathbf{D}_S \in \mathbb{R}^{N_x N_y \times \ell_t}$ is a temporal decimation matrix with $\mathbf{D}_S = \left(\mathbf{1}_{N_t}^T \otimes \mathbf{I}_{N_x N_y}\right)$, and $\mathbf{B}_S \in \mathbb{R}^{\ell_t \times \ell_t}$-$\mathbf{B}_C \in \mathbb{R}^{\ell_t \times \ell_t}$ are sines/cosines weight diagonal matrices. In particular, the entries of the matrices $\mathbf{B}_S$ and $\mathbf{B}_C$ are

given by

$$(B_S)_{k,k'} = \begin{cases} \sin\left(\frac{2\pi}{N_t}\lfloor\frac{k'}{N_xN_y}\rfloor\right), & \text{if } k = k' \\ 0, & \text{otherwise} \end{cases}, \tag{134}$$

and

$$(B_C)_{k,k'} = \begin{cases} \cos\left(\frac{2\pi}{N_t}\lfloor\frac{k'}{N_xN_y}\rfloor\right), & \text{if } k = k' \\ 0, & \text{otherwise} \end{cases}, \tag{135}$$

where $k = \{0, ..., \ell_t - 1\}$ and $k' = \{0, ..., \ell_t - 1\}$. Finally, the object phase information $\mathbf{\Phi}$ is obtained from $\tilde{\mathbf{\Phi}}$ by using an unwrapping algorithm Valadao y Bioucas-Dias, "PUMA: phase unwrapping via max flows". For scenarios where the structured light patterns are generated using multiples frequencies, the multifrequency heterodyne principle is the more suitable method to remove the 2π ambiguity Yin y col., "Temporal phase unwrapping using deep learning".

6.3. Simulation results

To generated synthetic spectral-depth images, we used the public dataset Silberman y col., "Indoor segmentation and support inference from rgbd images" and the hierarchical regression network for spectral reconstruction from RGB images Zhao y col., "Hierarchical Regression Network for Spectral Reconstruction from RGB Images". We randomly select one-hundred RGB-depth images to be converted to spectral-depth images, which were resized and cropped to have spatial and spectral dimensions $N_x \times N_y = 512 \times 512$ and $N_\lambda = 25$, respectively. We used the boxplot format to display the simulation results obtained –in different analyses– for the one hundred 4D datacubes. Note that the halfway mark indicates the median on each box, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively.

The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol.

All simulations were conducted and timed using an Intel Core i7 3960X 3.30 GHz processor with 32 GB RAM memory. Based on Eqs. (107) and (114), the compressed measurements were generated using two frequencies –high and low following the multifrequency heterodyne principle– with four patterns per-frequency, and a fixed random color-coded aperture.

6.3.1. Performance Evaluation for Different Additive Noise Scenarios We evaluated the performance of the proposed CS^{2}LI system in three different noisy scenarios with signal-to-noise-ratio (SNR) levels of $20, 30, 50$, respectively. The noise follows a colored Gaussian noise distribution. For the first analysis, Fig. 56 (A) illustrates –in a boxplot format– the spectral reconstruction accuracy, we compared the root mean square error (RMSE), structural similarity index measure (SSIM), and spectral angular mapper (SAM) of the reconstructions of one hundred scenes by using the SCCI (blue box), CS^{2}LI with binary CA (black box), and CS^{2}LI with color CA (green box) imager. The CS^{2}LI achieves comparable SAM and RMSE results to the SCCI but gains up to 0.10 and 0.08 SSIM on average with binary and color CA, respectively. Similarly to the first analysis, Fig. 56 (B) illustrates –in a boxplot format– the depth reconstruction accuracy by comparing the peak signal-to-noise ratios (PSNR), SSIM, and RMSE of the reconstructions of one hundred scenes by using the SCCI (blue box), CS^{2}LI with binary CA (black box), and CS^{2}LI with color CA (green box) imager. The CS^{2}LI gains up to 4 dB and 0.02 SSIM on average with respect to the SCCI, and reduces the RMSE by up to 0.02 on average. Although CS^{2}LI with binary and color CA achieves –on average– comparable results, the CS^{2}LI with color CA gains up to 1 dB and reduces the SAM metric up to 0.50° on average against the CS^{2}LI with binary CA.

To further compare the image quality, a representative reconstructed result of the

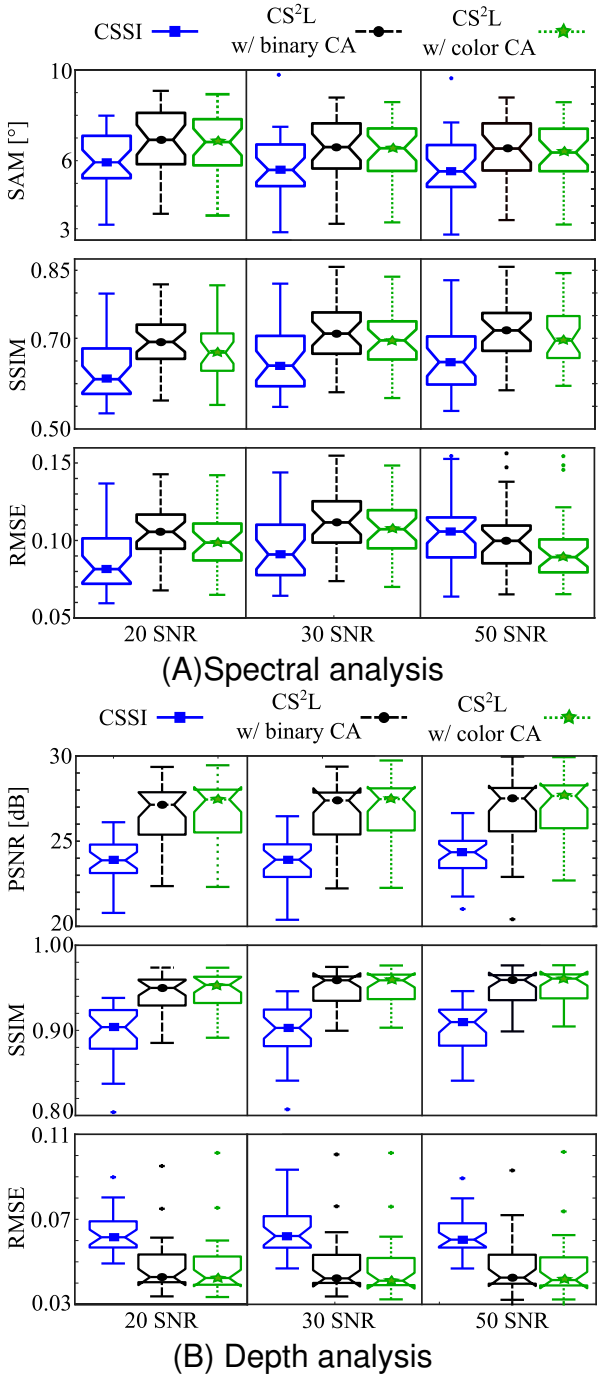


Figura 56. (A) Spectral and (B) depth reconstruction results obtained by the optical systems SCCI (blue straight line), CS²LI w/ binary CA (black dashed line) and CS²LI w/ color CA (green dotted line) in three noise level 20 (left column), 30 (center column), and 50 (right column).

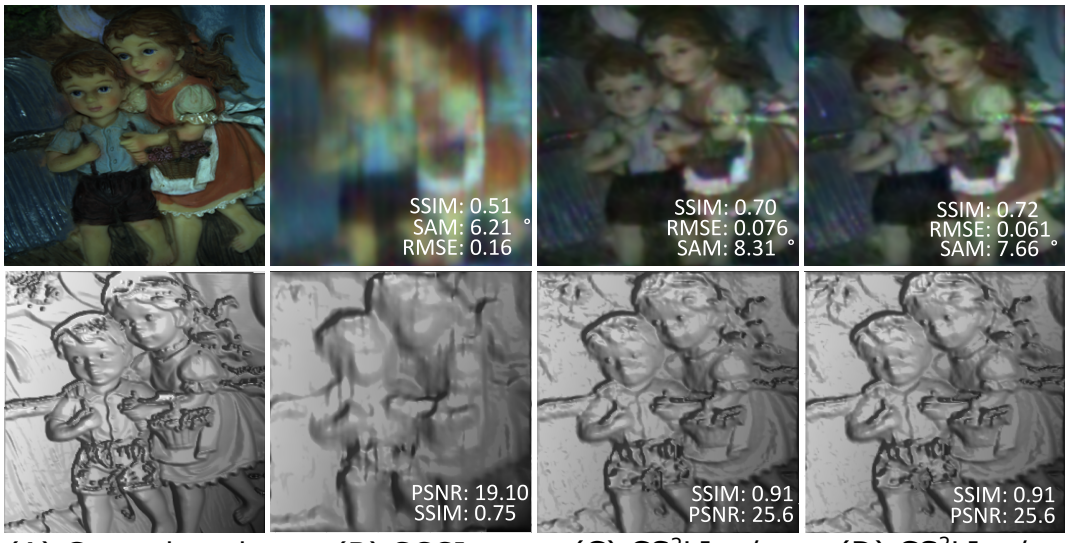


Figura 57. RGB composite reconstruction as well as the phase recovery from measurements with 20 dB SNR.

"Kids"scene is shown in Fig. 57. In particular, an RGB composited version of the ground truth (GT) is compared with their corresponding reconstructed by the CS-SI and the CS^{2}LI with binary and color CA. The depth map estimations were also included to highlight CS^{2}LI's depth reconstruction accuracy. These results support CS^{2}LI's claim that spatial recovery performance affects depth reconstruction accuracy.

6.3.2. Color coded aperture design The CCA-based CS^{2}LI's main constituents are the number of filters (referred to as $\#_f$) and the transmittance level (referred to as κ) that compose the CCA. Thus, we conducted an ablation analysis of the number of filters and transmittance to determine the CCA composition that yields the best trade-off between the spatial, spectral and depth reconstruction quality. Figure 58 illustrates CS^{2}LI's (A) spectral and (B) depth reconstruction accuracy for twelve CCA compositions. To echo and support the results shown in Fig. 56, the spectral and depth results were evaluated using the same four metrics: SAM, SSIM, RMSE, and

PSNR.

For the spectral analysis, Fig. 58(A) illustrates that a CCA constructed with $\#_f = N_\lambda$ and $\kappa = 25\,\%$ is an optimal configuration in terms of the SAM and RMSE metrics. In terms of the SSIM metric, an optimal CCA configuration is achieved with $\#_f = N_\lambda$ and $\kappa = 75\,\%$. For the depth analysis, Fig. 58(B) shows that a CCA constructed with $\#_f = 1$ and $\kappa = 75\,\%$ is an optimal configuration since the three metrics agree. When the CA is constructed using $\#_f = N_\lambda$ and $\kappa = 25\,\%$, the CS^{2}LI system achieves the higher spectral encoding freedom, i.e., achieves up to an ideal narrow filter per spatial location. This configuration and results echo the claims presented in the 3D CCA-based CSI literature León-López, Carreno y Fuentes, "Temporal colored coded aperture design in compressive spectral video sensing"; Correa, Arguello y Arce, "Snapshot colored compressive spectral imager"; Chen Fu y col. "Compressive spectral polarization imaging by a pixelized polarizer and colored patterned detector". En: *JOSA A* 32.11 (2015), págs. 2178-2188. In contrast, the CS^{2}LI performs an optimal depth encoding when $\#_f$ and κ tend to 1 and $100\,\%$ (i.e., an all-ones coded aperture), respectively. The CS^{2}LI with an .ªll-ones coded apertureïndicates that the optimal depth encoding approach is achieved by using a traditional grayscale sensor. Here, the depth recovery accuracy tends to be invariant to the amount of the filters that composed the CCA because the proposed CS^{2}LI system preserves with high quality the spatial features. These disjunctive results lead to selecting a CCA configuration that allows obtaining an optimal trade-off between spectral and depth recovery accuracy. Based on the Fig. 56 results, we found that the CCA configuration that promotes an optimal spectral/depth reconstruction trade-off is achieved with $\#_f = 6$ and $\kappa = 50\,\%$.

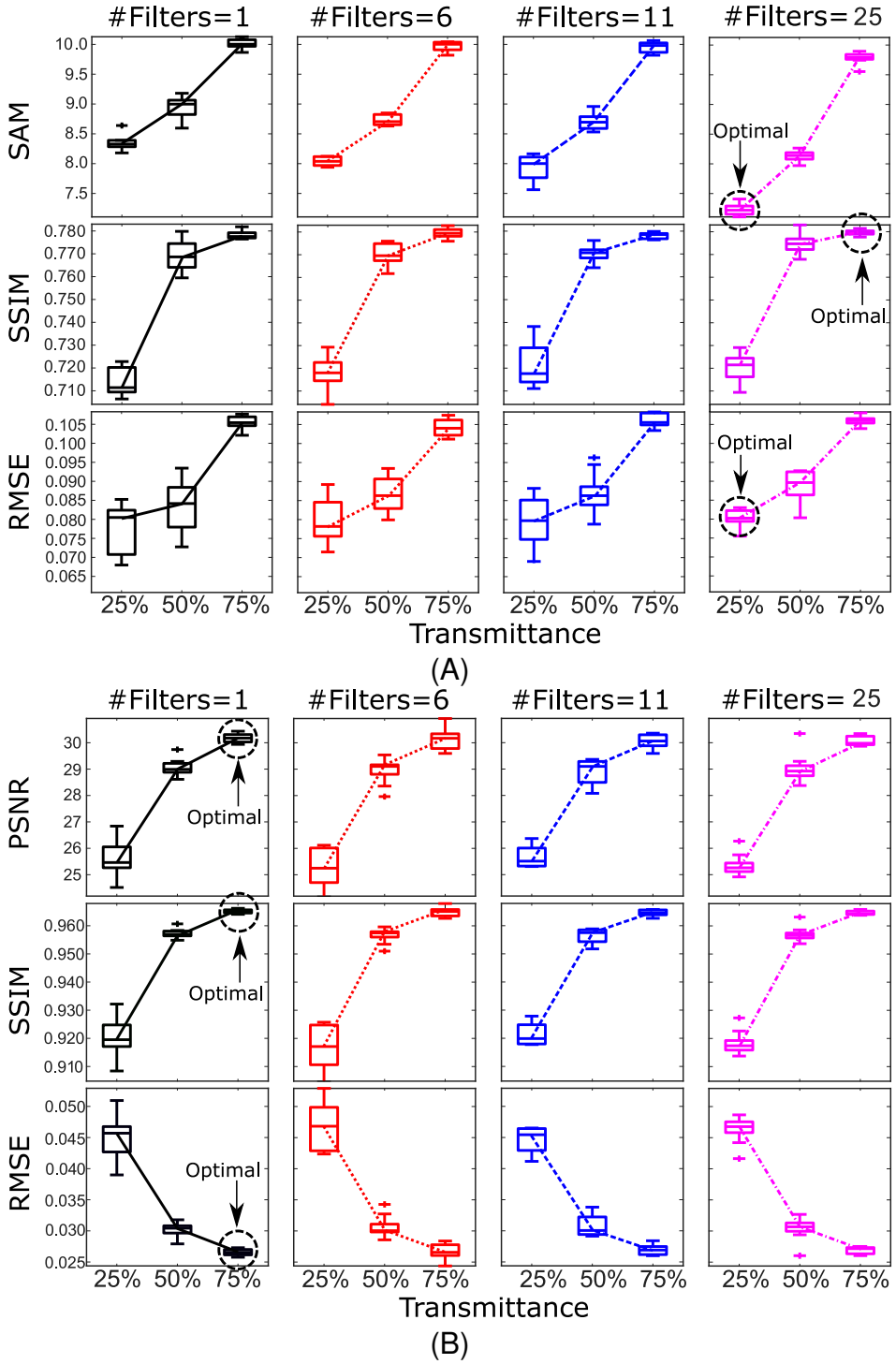


Figura 58. The CS^{2}LI's (A) spectral and (B) depth reconstruction analysis by varying the CCA construction in terms of $\#_f = N_\lambda$ and $\kappa = 25\,\%$.

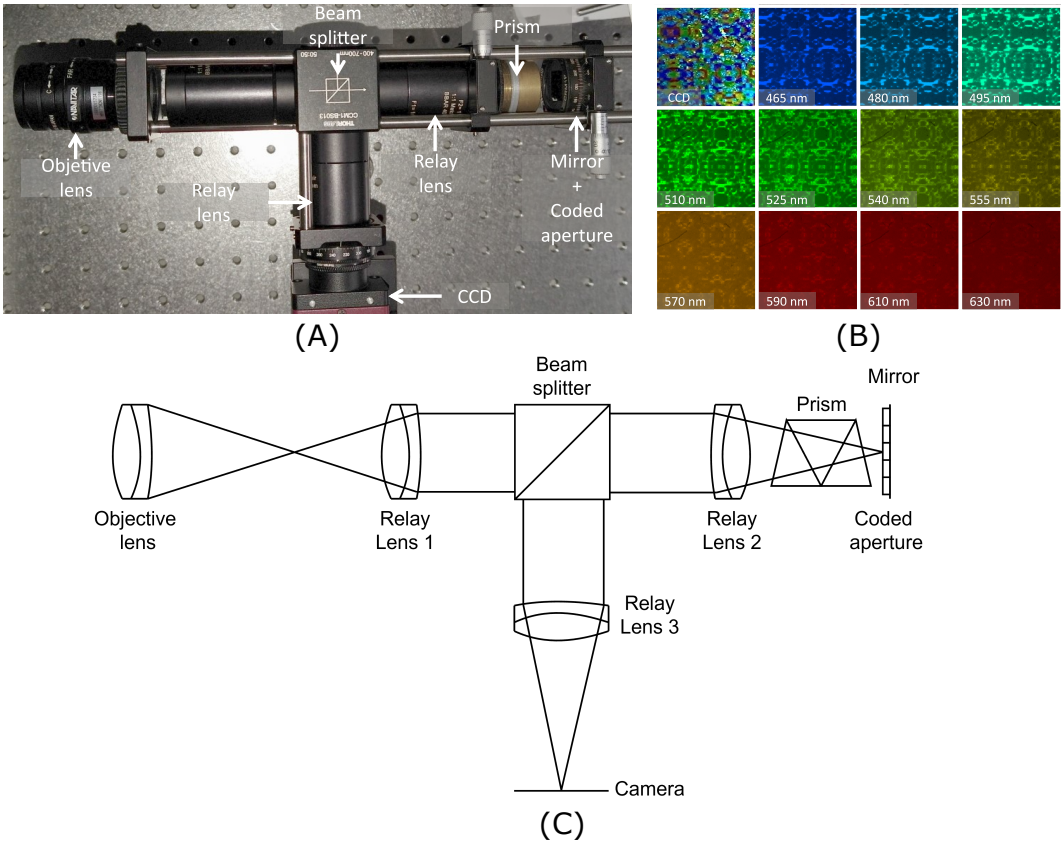


Figura 59. (A) Testbed implementation of the compressive spectral structured light imaging. (B) Colored-coded aperture spectral response. (C) Geometric sketch of the optical system.

6.4. Proof of concept experiments

We built a testbed in our laboratory to the proposed idea, through a proof-of-concept prototype, as shown in Fig. 59. This prototype uses a Navitar lens ($12mm$ fixed focal length, MVL12M23, $f/1.4$) as the objective lens to image the scene onto the image plane of a matched achromatic doublet pair (Thorlabs MAP105050 - A, $f_1 = 50mm, f_2 = 50mm$) that propagates the incoming wavefront through a beam splitter and onto a second matched achromatic doublet pair relay lens (Thorlabs MAP10100100-A, $f_1 = 100mm, f_2 = 100mm$). This second relay lens transmits the wavefront through a double Amici prism and onto the color-coded aperture which is printed on top of a broadband dielectric mirror (Thorlabs BB1-E02 - 400-700 nm). Taking advantage of the mirror surface, the now dispersed-and-coded wavefront is reflected back to the prism, which introduce a second dispersion effect. The resulting dispersed-coded-undispersed wavefront propagates through the beam-splitter until a third matched achromatic doublet pair relay lens (L3) (Thorlabs MAP10100100-A, $f_1 = 100mm, f_2 = 100mm$). Finally, the last relay lens focuses the dispersed-coded-dispersed wavefront onto the sensor (Stingray F-080B, $4.65\mu m$ pixel size). The experiments evaluate three target scenes named Kids [see Fig. 62], Tayrona [see Fig. 63], and Guatape [see Fig. 64], for which eight compressive projections were acquired with the testbed. The structured light fringes projected by the video-beam are generated by the equation

$$I^p_{i_1,i_2} = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi\nu(i_1 - i_2)}{L} - \frac{2\pi p}{K}\right), \tag{136}$$

where ν is the frequency of fringe patterns, $K \in \mathbb{N}$ is the total of pattern projections per frequency, $L \in \mathbb{N}$ is the total number of pixels in the transverse direction of the fringe pattern, and $p = \{0, ..., K-1\}$. In this regard, the parameters selected for the experiments are $K = 4$, $L = 1388$, and $\nu = \{14, 2\}$. Figure 60 illustrate the images of

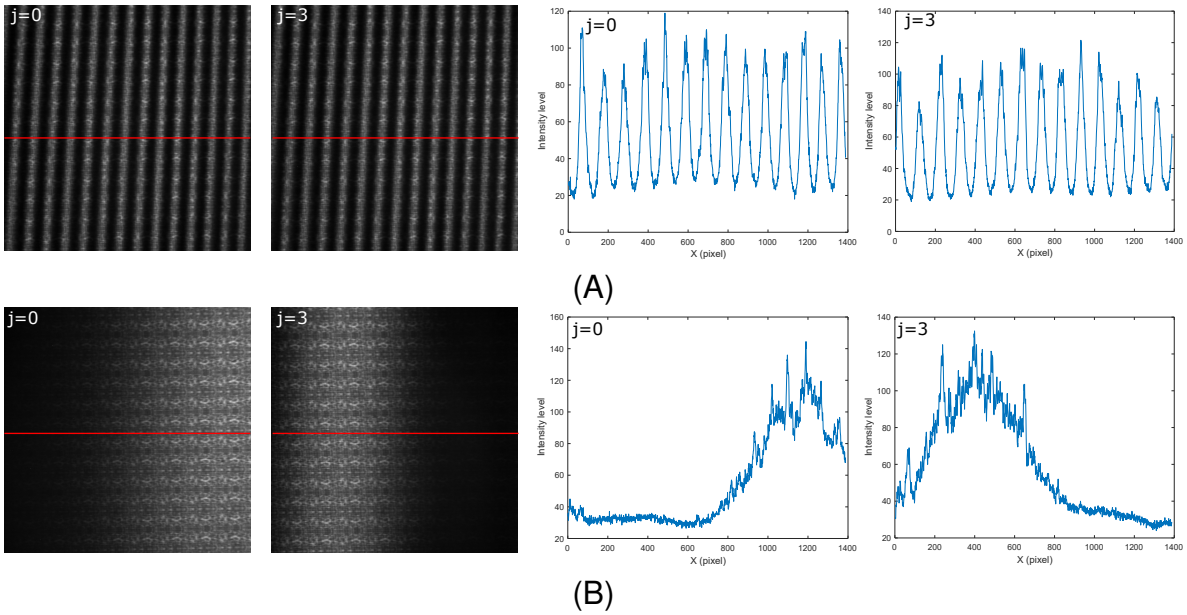


Figura 60. Images of experimental fringes captured with the CCD together with their respective transverse sections for $K = 4$, $L = 1388$ (A) $\nu = 14$ and (B) $\nu = 2$.

experimental fringes captured for the high (i.e., $\nu = 12$) and low frequency (i.e., $\nu = 12$). In these acquisitions can be noticed the color-coded aperture pattern. Finally, Fig. 61 shows the wrapped and unwrapped versions of the acquisitions illustrated in Fig. 60.

Figures 62-64(A) shows four out of the eight raw measurements. These measurements are given as input to the **Algorithm 6**, which returns the reconstructed spatial-spectral-depth information as output, as shown in Figs. 62-64(B)-(D). He-

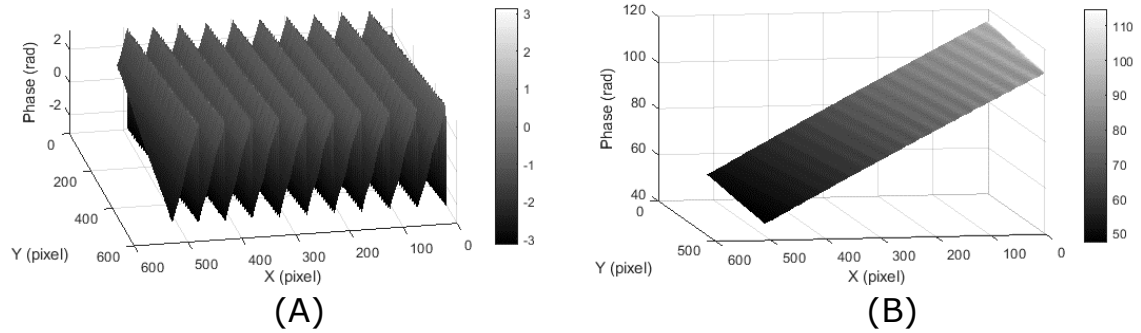


Figura 61. Discontinuous phase generated from the experimental fringes **I** (see Fig. 60) and (B) its respective absolute continuous phase.

re, the reconstructred phase information is illustrated in Figures 62-64(B). The RGB composite of the reconstruction is depicted in Fig. 62(C). Moreover, Fig. 62(D) shows the spatial reconstruction per band to evaluate the accuracy of the spectral reconstruction, 8 out of the 25 spectral bands are depicted. Moreover, Fig. 65 shows the spectral signature of six scene points measured with a spectrometer (Ocean Optics Flame S-VIS-NIR-ES spectrometer), assumed to be the ground truth (red solid line), and compared with the ones obtained with the CS^{2}LI (blue dashed line). In this figure it can be noted that the reconstructed spectral signatures resemble those obtained with the spectrometer, with SAM values smaller than $6.8°$.

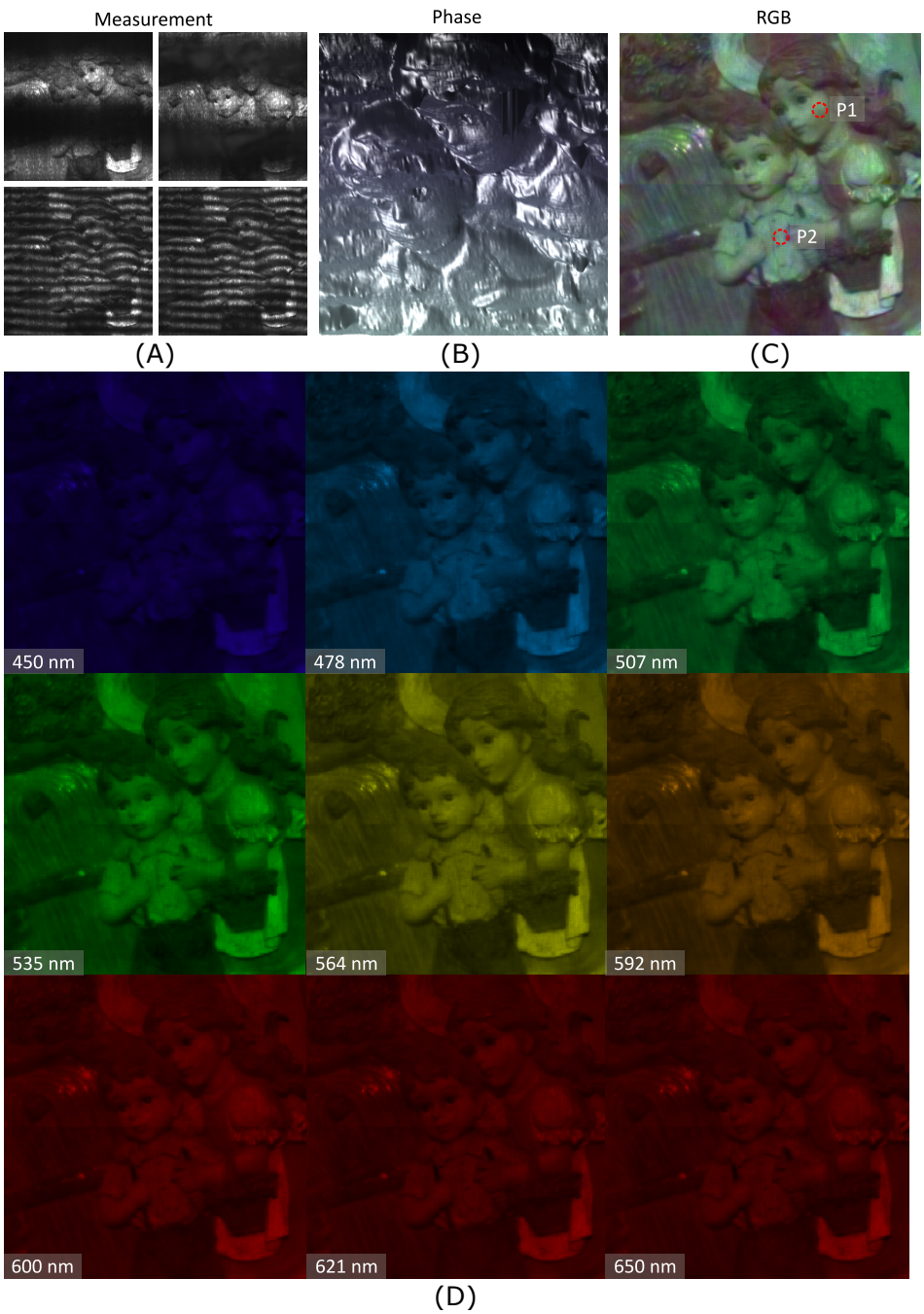


Figura 62. Testbed results. (A) Spectral and (B) depth reconstruction results obtained for the optical systems SCCI Illustration of 4 out of the 8 compressed measurements. (B) Phase estimation via Eq. (133) and PUMA **Algorithm**. (C) RGB composite of the reconstructions. (D) Illustration of 8 out of the 25 reconstructed spectral bands.

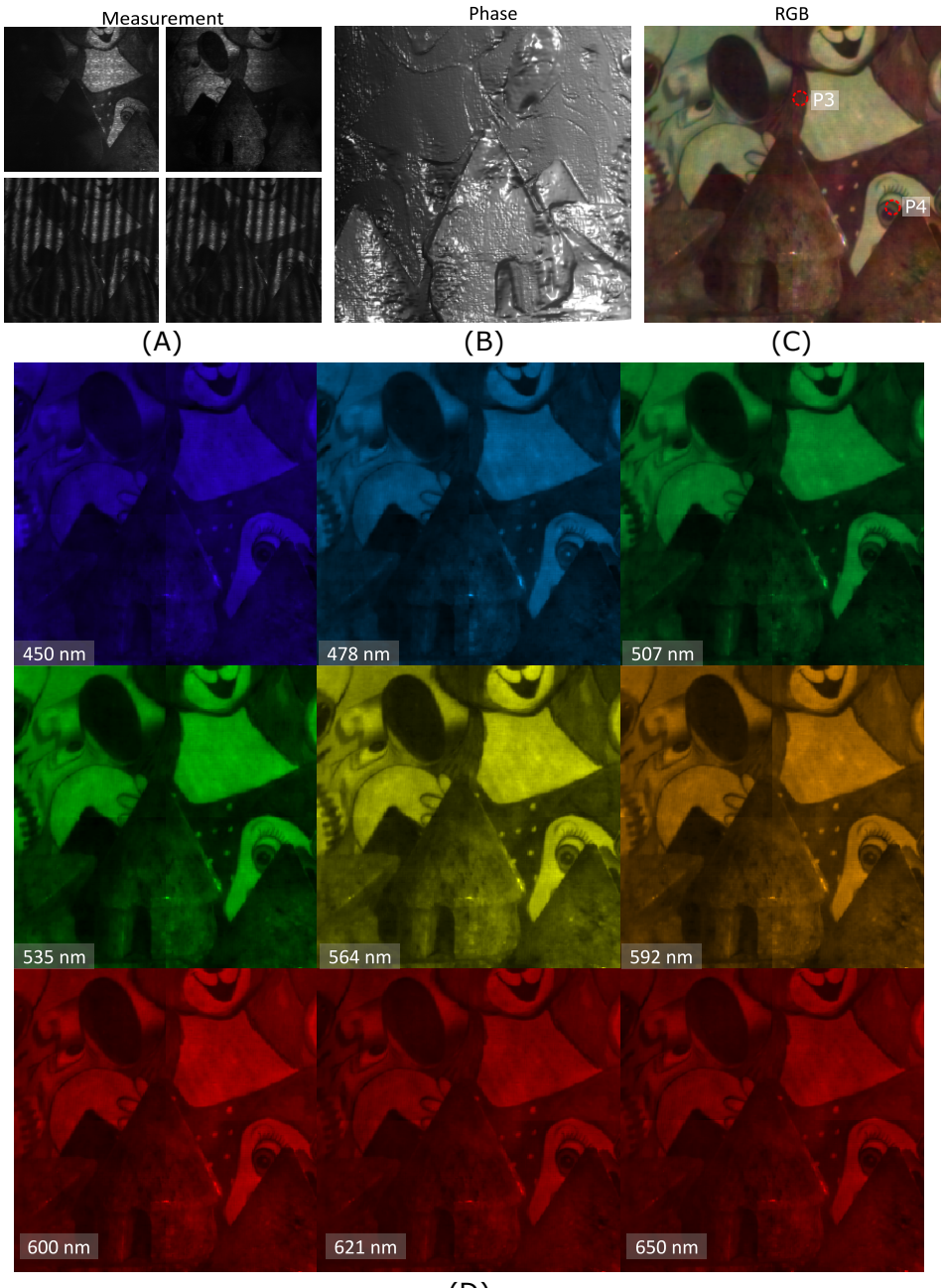


Figura 63. Testbed results. (A) Illustration of 4 out of the 8 compressed measurements. (B) Phase estimation via Eq. (133) and PUMA **Algorithm**. (C) RGB composite of the reconstructions. (D) Illustration of 8 out of the 25 reconstructed spectral bands.

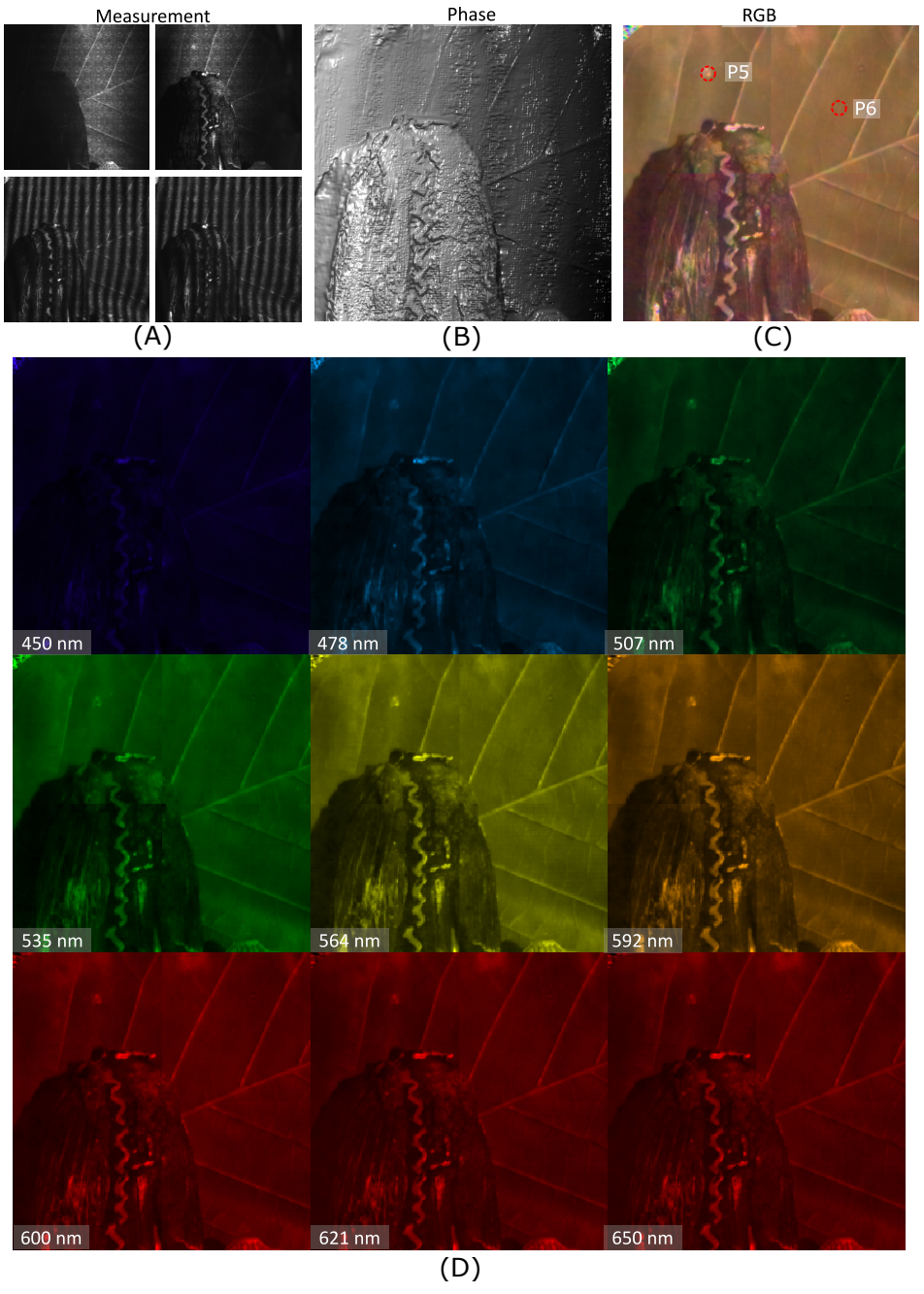


Figura 64. Testbed results. (A) Illustration of 4 out of the 8 compressed measurements. (B) Phase estimation via Eq. (133) and PUMA **Algorithm**. (C) RGB composite of the reconstructions. (D) Illustration of 8 out of the 25 reconstructed spectral bands.

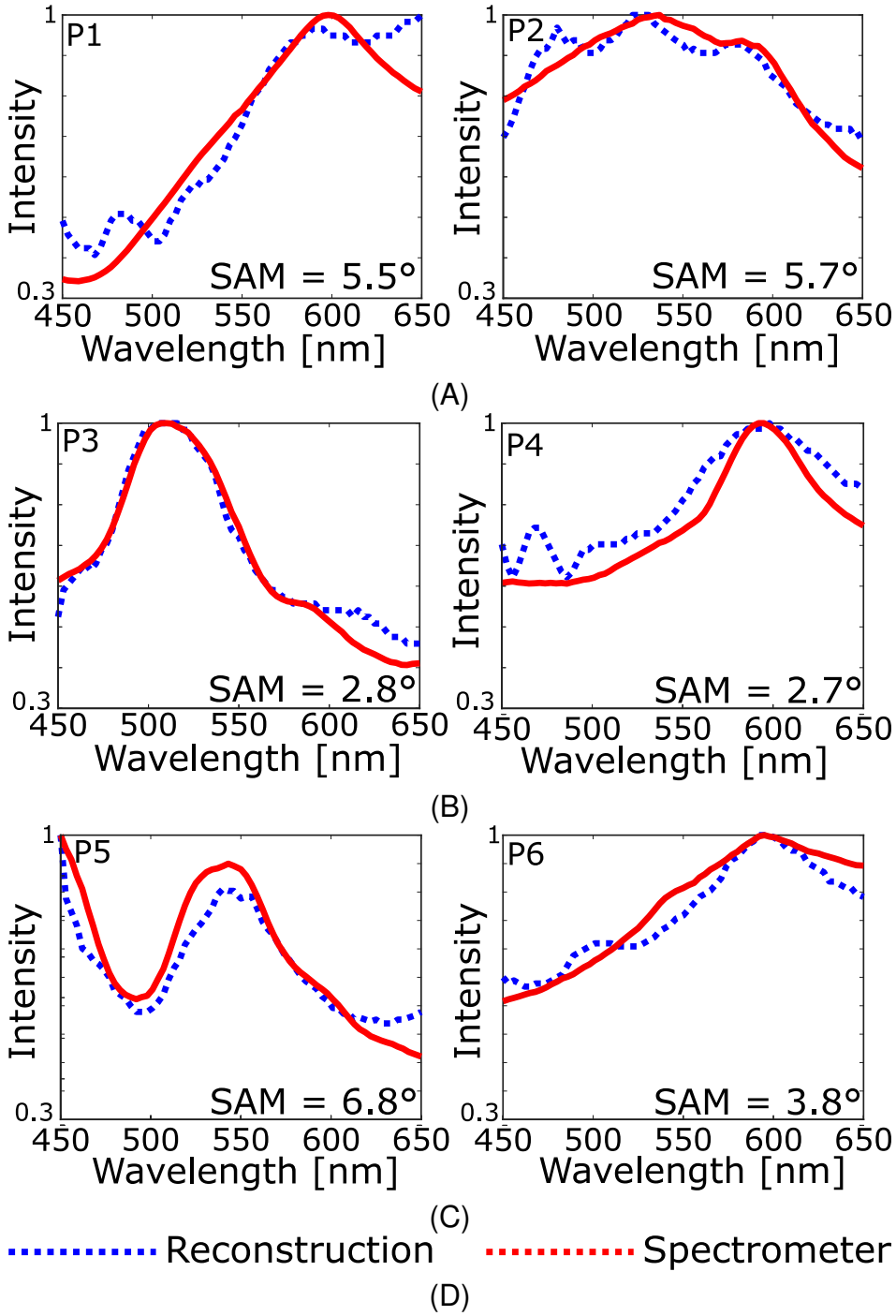


Figura 65. Spectral signatures of six spatial points from the (A) Kids (P1-P2), (B) Tayrona (P3-P4), and (C) Guatape (P5-P6) scenes.

6.5. Conclusions

This chapter proposed the CS^{2}LI system to acquire spectral depth images with a single sensor and active structured light illumination. CS^{2}LI was constructed following a dual-dispersion sensing geometry but relying on a single prism and a color-coded mirror. This dual-dispersion sensing geometry preserves the spatial features with high quality, which are critical for the SL-based CoSDI approaches. Moreover, inspired by the hierarchical resolution approaches and the ADMM paradigm, we proposed a low-complexity reconstruction algorithm that splits the spectral-depth reconstruction problem into two low-complex subproblems: spectral datacube reconstruction and grayscale structured light patterns estimation. A distinctive and attractive feature of the proposed architecture and reconstruction methodology is that unlike Heist y col., "5D hyperspectral imaging: fast and accurate measurement of surface shape and spectral characteristics using structured light"; Xueping Ni, Changying Li y Huanyu Jiang. "Development of a 3D Multispectral Imaging System using Structured Light". En: *2019 ASABE Annual International Meeting*. American Society of Agricultural y Biological Engineers. 2019, pág. 1; *Hyperspectral+ Depth Imaging Using Compressive Sensing and Structured Light*. Optical Society of America. 2018, págs. 1-2, our approach does not require multiple sensors, bulky configurations, and high-cost CFMD optical systems to obtain high-quality reconstructions.

The optimal CCA configuration regarding transmittance (referred to as κ) and filter composition (referred to as $\#_f$) was established by developing extensive simulations, resulting in the parameters $\kappa = 50\,\%$ and $\#_f = 6$. We validated the proposed CCA-based dual-dispersive encoding geometry by customizing a low-cost CCA using a FUJICHROME Velvia 50 transparency film. The customized CCA is attached to a dielectric mirror, resulting in a low-cost and efficient color-coded mirror. Finally, the CCA-based dual-dispersion encoding is achieved by coupling the customized CCA and the double Amici prism (see **Appendix 9**).

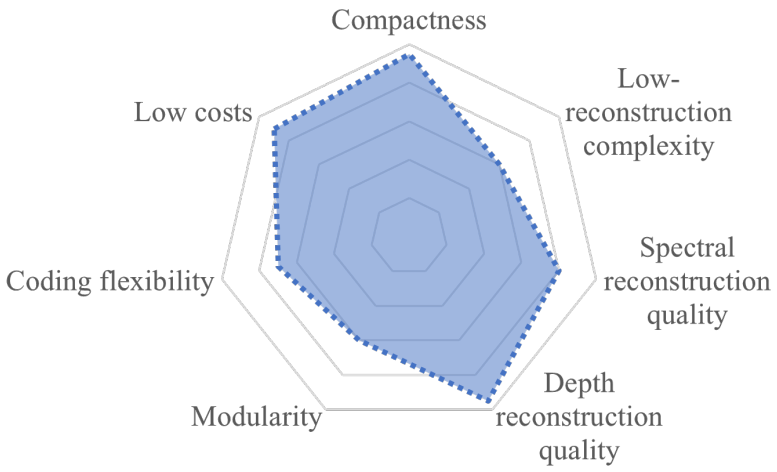


Figura 66. Chart graph for the CS^{2}LI system

As a chapter summary and to assess the CS^{2}LI's pros and cons, we construct a radar chart with seven relevant CoSDI features: compactness, low costs, low-recovery complexity, spectral recovery, depth recovery, modularity, and encoding flexibility. Here the bottom and the top contours represent the worst and best scores, respectively, i.e., the ideal CoSDI system would have to reach the upper shell in the seven features. Specifically, Fig. 66 illustrates the CS^{2}LI's radar chart, showing that CS^{2}LI is a low-cost and compact system that allows achieving high-quality information recovery at the expense of encoding flexibility and recovery complexity.

7. Mu-CoSDI: Multimodal Compressive Spectral Depth Imager

In the previous chapters, we introduced three compressive spectral depth imagers (CoSDI) that couple the coded aperture snapshot spectral imaging (CASSI) sensing geometry (introduced in **Chapter 3**) with three depth-sensing modalities: depth-from-aberration (DFA), light field (LF), and structured light (SL). Specifically, **Chapter 4** presented the coded aperture snapshot spectral imager via DFD (CASSI-DFA) based on a hybrid phase and amplitude modulation geometry. The CASSI-DFA has high coding flexibility –associated with the high-speed refresh rate of optoelectronic DM and DMD devices– making it a good alternative for applications where the acquisition of multiple measurements is feasible. **Chapter 5** presented a compressive color-coded compressive spectral light field imager (C^3SLFI) based on two-complementary color-coded apertures, a lenslet array, and a single sensor. The C^3SLFI allows accurate high spectral and depth information recovery by estimating ad-hoc representation bases via tensor theory. Finally, **Chapter 6** introduced the compressive spectral structured light imaging (CS^2LI) system to emulate a double-dispersive CASSI system from a double-Amici prism and a coded mirror. The CS^2LI offers a compact optical distribution that enables it to be coupled –in a modular way– to either a lenslet array (in front of the objective lens) or a light projector. The CS^2LI' modular structure has a high potential to become the baseline for a multimodal CoSDI system. Although the proposed CoSDI systems allow accurate spectral and depth recovery, the analytical-based reconstruction algorithms are overshadowed by the reconstruction speed up offered by the recently proposed convolutional neural network (CNN) approaches. For instance, CNN has been established as a crucial compressive spectral imaging constituent to enhance the reconstruction accuracy and alleviate the computational complexity Ziyi Meng y col. "Self-supervised neural networks for spectral snapshot compressive imaging". En: *Proceedings of the*

IEEE/CVF International Conference on Computer Vision. 2021, págs. 2622-2631.

The positive impact of CNN to alleviate the computational complexity and to improve the reconstruction quality along with the CS^2LI's modularity leads naturally to the question: ***Is it possible to develop a multimodal CoSDI system enhanced by the CNN paradigm without relying on expensive encoding optical devices and camera arranges?***

To answer this question, this chapter extends the CS^2LI sensing geometry to build a multimodular CoSDI system capable of working on either active or passive depth paradigms, e.g., SL or LF. Moreover, to enhance and speed up the spectral reconstruction, we developed an end-to-end convolutional neural network named deep compressive spectral net (D-CoSNet). Here, D-CoSNet is constructed following an alternating direction method of multipliers (ADMM) to transfer many advantageous mathematical properties embedded in the main CoSDI sensing operations (direct sensing, transposed, and inverse). We have coined the latter, multimodal compressive spectral depth imager â€“ Mu-CoSDI.

7.1. Continuous and discrete sensing model

Advanced image reconstruction methods that provide high reconstructed image quality have become indispensable in the compressive spectral depth imaging (CoSDI) framework. To date, analytical models are predominantly used in compressive spectral imagers (CSIs). These methods solve the inverse problem by using the prior knowledge of the sensing matrix and by leveraging the sparsity of high-dimensional images in certain domains (e.g., spatial, wavelet, discrete cosine transform, and gradient domains). However, they require precise prior knowledge of all sensing operations [i.e., sensing direct, transpose, and inverse] involved in the CSI systems. Moreover, their processing time increases exponentially in high-dimensional optimization problems, and the reconstructed image quality highly depends on the empirical tuning of parameters.

The increasing availability of vast amounts of image data and computational power has consolidated convolutional neural networks (CNNs) as one of the most potent and desired tools for CSI approaches *Reconnet: Non-iterative reconstruction of images from compressively sensed measurements*. 2016, págs. 449-458; Kyong Hwan Jin y col. “Deep convolutional neural network for inverse problems in imaging”. En: *IEEE Transactions on Image Processing* 26.9 (2017), págs. 4509-4522; Ulugbek S Kamilov y Hassan Mansour. “Learning optimal nonlinearities for iterative thresholding algorithms”. En: *IEEE Signal Processing Letters* 23.5 (2016), págs. 747-751; Ziyi Meng, Shirin Jalali y Xin Yuan. “Gap-net for snapshot compressive imaging”. En: *arXiv preprint arXiv:2012.08364* (2020); Zhuoyuan Wu, Jian Zhang y Chong Mou. “Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging”. En: *arXiv preprint arXiv:2109.06548* (2021); Meng y col., “Self-supervised neural networks for spectral snapshot compressive imaging”. CNN’s potential for compressive sensing was initially demonstrated to reconstruct images from measurements generated by a random Gaussian matrix Kulkarni y col., *Reconnet: Non-iterative recons-*

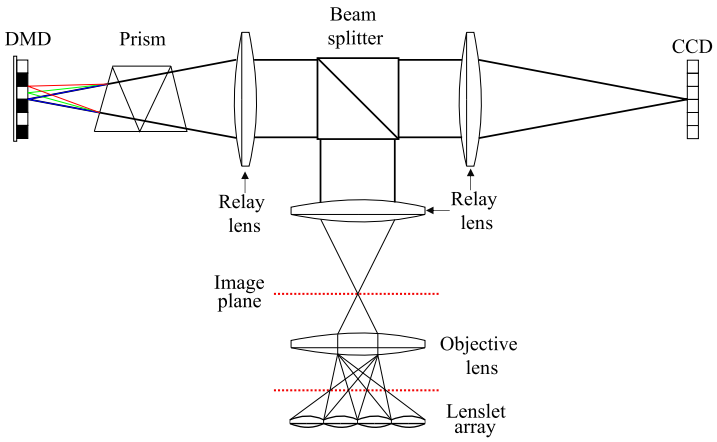
truction of images from compressively sensed measurements. Inspired by this early development, CNN paradigms were integrated into CSI's reconstruction by combining its forward model as an encoder, followed by convolutional layers as a fast approximated sparse decoder Jin y col., "Deep convolutional neural network for inverse problems in imaging". Despite high reconstructed image quality in simulation, these early CNN-based approaches produced a dramatically decreased reconstruction quality for real experiments. This limited performance was attributed to the fact that CNNs did not consider the sensing operators associated to the SCI models (e.g., direct, transpose, and inverse processes) that could work as prior information to improve the reconstruction results. In this path, subsequent works tackled this problem by taking advantage of the well-established link between analytical-modeling approaches and deep learning networks Kamilov y Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms", resulting in the development of the unrolling approach Vishal Monga, Yuelong Li y Yonina C Eldar. "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing". En: *IEEE Signal Processing Magazine* 38.2 (2021), págs. 18-44; Siming Zheng y col. "Deep plug-and-play priors for spectral snapshot compressive imaging". En: *Photonics Research* 9.2 (2021), B18-B29 and the plug-and-play (PnP) method Zheng y col., "Deep plug-and-play priors for spectral snapshot compressive imaging", both of which relied on iterative gradient descents models for the inferences. However, the unrolling approaches were highly prone to suffer vanishing gradient issues as the number of emulated iterations increased. The PnP approaches limited the CNN's inference potential by using them as simple denoisers. Thus, most existing CNN-based methods are merely used to replace analytical-modeling-based image reconstruction.

To overcome these limitations, we develop a convolutional neural network, termed deep compressive spectral net (D-CoSNet), for reconstructing 3D spectral datacubes from measurements generated by the Mu-CoSDI in either of its three modali-

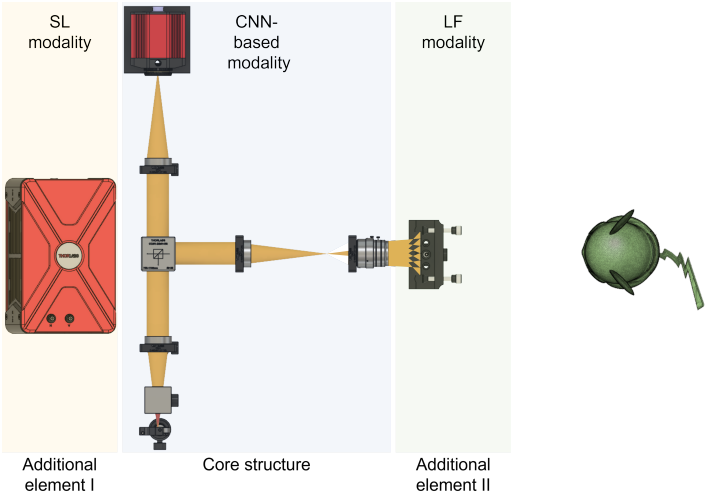
ties: light field (LF), structured light (SL), and pure-CNN-based depth approach. The Mu-CoSDI's sensing geometry is an extension of CS^{2}LI (see chapter 6), and their multimodal construction is illustrated in Fig. 79. The D-CoSNet is envisioned to benefit CoSDI in image reconstruction by transferring many advantageous properties embedded in certain mathematical models into custom layers. Towards this goal, the D-CoSNet comprises two main sequential stages: a deep-unfolding-based network to embody SCI's sensing model and a U-net Ronneberger, Fischer y Brox, "U-net: Convolutional networks for biomedical image segmentation" that works as a filter. The first stage is developed from the structure of an alternating direction method of multipliers (ADMM) Stephen Boyd y col. "Distributed optimization and statistical learning via the alternating direction method of multipliers". En: *Foundations and Trends® in Machine learning* 3.1 (2011), págs. 1-122, that integrates SCI's major operations with a mathematical coherence, emulated as three custom layers constructed by using the coded aperture and the shearing function.

7.1.1. Continuous sensing model For the D-CoSNet construction, we first require to establish a general sensing model for the proposed multimodal CoSDI, that encompasses the models described by Eqs. (56) (see Section 4), (**??**) (see Section 5) and (112) (see Section 6). In this modeling and for convenience, we referred to the light source object as the objective lens output wavefront, see 79 (A) represented with a dotted red line (image plane). This assumption allows us to develop a spectral CNN-based reconstruction methodology dockable to the Mu-CoSDI's multimodalities, e.g., LF, SL, and pure-depth approach.

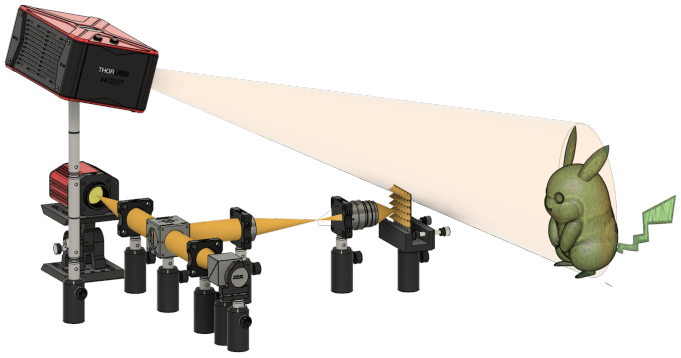
Let $f_o(x, y, \lambda, u, v)$ be the spatial spectral-angular scene in intensity, where (x, y) index the spatial coordinates, (u, v) index the view-angle dimension, and λ index the wavelength. The object's depth map is encoded by illuminating it with a structured light pattern, where the resulting pattern deformations are associated to the object shape. The pattern deformation can be modeled as a phase modulation, where



(A)



(B)



(C)

Figura 67. (A) Sketch of the proposed Mu-CoSDI architecture. (B)-(C) Upper and lateral sketch view where the distance configuration and the optical element labels are specified, respectively.

the object's shape acts as the phase mask. Based on Eq. (11), the object shape-modulated structured light can be modeled as

$$p(x, y, \lambda, u, v, t) = f_o(x, y, \lambda, u, v) \left[\cos\left(\varphi(x, y, u, v) + \omega_0 t\right) + b\right] + a(x, y, u, v), \quad (137)$$

where $\varphi(x, y, u, v)$ is the object phase/shape information, ω_0 represents the temporal-carrier, $a(x, y, u, v)$ is the background, and b is a balance constant that prevents $p(x, y, \lambda, u, v, t)$ achieve negative values with $b = 1$. First, an objective lens coupled with a lenslet array (LA) forms an image in an intermediate image plane, resulting in the wave field

$$f_1(x', y', \lambda, t) \iint \iint \gamma(\lambda) p(x, y, \lambda, u, v, t) h_{LA}(x - x', y - y', \lambda, u, v) dx dy du dv, \quad (138)$$

where $h_{LA}(\cdot)$ is the incoherent PSF of the objective lens + LA and $\gamma(\lambda)$ represent the spectral source density. This 4D to 2D parameterization (without considering the spectral dimension) was proposed in Adelson y Wang, "Single lens stereo with a plenoptic camera", in which the view angle dimension is collapsed into the spatial dimension. In this regard, the conventional light field systems, the previously proposed C^{3}SLFI (see Chapter 5), and consequently the proposed Mu-CoSDI system, aim to reconstruct the spectral information in this 2D projection. objective lens + LA forms image in the input plane of a 4f system, which propagates the incoming wavefront through a prism that introduces a wavelength-dependent horizontal shifting to form a dispersed image version over a coded aperture attached to a mirror (or DMD). The resulting dispersed-encoded wavefront can be seen in the detector plane (immediately before the sensor) as

$$f_2(x'', y'', \lambda, t) = k(x'', y'', \lambda) \iint h_2(x' - x'' - S_1(\lambda), y' - y'', \lambda) f_1(x', y', \lambda, t) dx' dy' \quad (139)$$

with

$$k(x'', y'', \lambda) = \frac{1}{2} \sum_{i_1', i_2, i_3} \text{rect} \left(\frac{x''}{\Delta_c} - i_1, \frac{y''}{\Delta_c} - i_2, \frac{\lambda}{\Delta_\lambda} - i_3 \right) C_{i_1, i_2, i_3}, \tag{140}$$

where $rect()\dot{}$ represents the rectangle function, $S_1(\lambda)$ is a wavelength-dependent function for the propagation from the relay lens to the mirror; Δ_c denotes the pixel size, Δ_λ denotes the spectral bandwidth, $C \in \mathbb{R}^(N_y \times (N_x + N_\lambda - 1) \times N_\lambda)$ is a color-coded aperture with $C_(i_1', i_2, i_3) \in 0, 1$, and $i_1' = 0, ..., (N_x + N_\lambda - 2)$, $i_2 = 0, , N_y - 1$ and $i_3 = 0, ..., N_\lambda - 1$ indexing the rows, columns, and spectral bands, respectively. Then, the color-coded mirror reflects-back the dispersed-modulated wavefront to the prism, which introduce a second spectral dispersion. The resulting dispersed-coded-dispersed wavefront is propagated and focused into a detector array. The sensing process can be expressed as

$$f_3(x''', y''', \lambda, t) = \iint f_2(x'', y'', \lambda, t) h_3(x'' - x''' + S_2(\lambda), y'' - y''', \lambda) dx'' dy'', \tag{141}$$

where h_3 represents the intensity PSF and $S_2(\lambda)$ is a wavelength-dependent function for the propagation from the mirror to the relay lens (i.e., backpropagating through the prism). Finally, the measurement at the $(i_1, i_2)^{th}$ pixel is represented by

$$g_{i_1, i_2} = \int_\Gamma \rho(\lambda) \iint f_3(x''', y''', \lambda) \text{rect} \left(\frac{x'''}{\Delta_d} - i_1, \frac{y'''}{\Delta_d} - i_2 \right) dx''' dy''' d\lambda, \tag{142}$$

where Λ represents the wavelength axis over the spectral range, Δ_p represents the pixel's side length, and $\rho(\lambda)$ represents the detector's normalized quantum efficiency. Note that the notion of a "multi-modality"sensing model in Eq. (159) also includes the three configurations illustrated in Fig. 79: (i) single view by relying just on the core structured; (ii) multi-view by relying on the core structure and the lenslet array (referred to as Additional element I in Fig. 79 (B); (iii) single view with structured

light by relying on the core structure and the video beam (referred to as Additional element II in Fig. 79 (B)).

7.1.2. Discrete sensing model Based on Eq. (**??**), the data acquisition of Mu-CoSDI consists of the following three operations. First, a 3D scene, denoted by **F**, is horizontally dispersed (denominated shearing operation) to be spatially modulated by a single 2D coded aperture. The resulting dispersed-modulated information is then undispersed to obtain a dispersed-coded-undispersed wavefront. Finally, via the operation of integration, a 2D detector integrates the encoded and sheared scene as a 2D image, commonly known as a compressed measurement, denoted by **G** Cao y col., “Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world”. The vectorized discrete forward model is expressed as

$$\mathbf{g} = \mathbf{\Phi f} + \boldsymbol{\epsilon}, \tag{143}$$

where $\mathbf{f} \in \mathbb{R}^{n\times 1}$ is the vector representation of a discrete version of the 3D scene $\mathbf{F} \in \mathbb{R}^{N_x N_u \times N_y N_v \times N_\lambda}$ with size $n = (N_x N_u)\cdot(N_y N_v)\cdot N_\lambda$; $N_x N_v$ and $N_y N_v$ represent the data lengths in the two spatial dimensions with $N_u \times N_v$ being the number of lenses in the lenslet array and $N_x \times N_y$ as the number of pixels along x and y contained in each view of each lens within the lenslet, and N_λ represents the data length in the spectral dimension. $\mathbf{g} \in \mathbb{R}^{m\times 1}$ is the vectorial version of the compressed measurement $\mathbf{G} \in \mathbb{R}^{N_x N_u \times N_y N_v}$ with $m = (N_x N_u)\cdot(N_y N_v)$; $\boldsymbol{\epsilon} \in \mathbb{R}^{m\times 1}$ represents the noise which is modeled with a colored Gaussian distribution; $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ is the CSI’s sensing matrix. The structure of $\mathbf{\Phi}$ accounts for the optical elements in the CSI system, and its entries are given by

$$\Phi_{i,j} = \begin{cases} c_k, & if\ i = \mathsf{mod}(j,m) \\ 0, & otherwise \end{cases}, \tag{144}$$

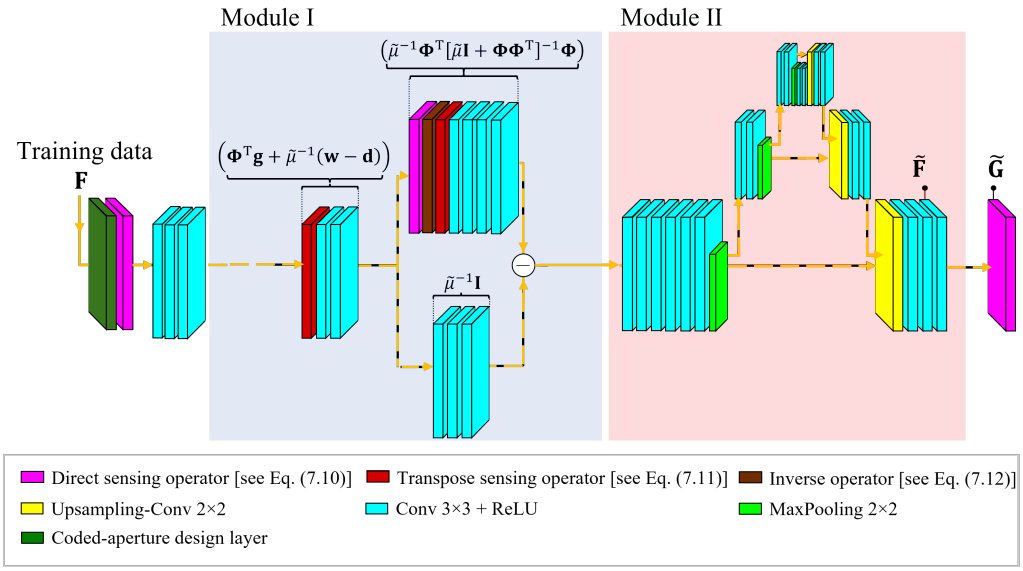


Figura 68. Schematic of D-CoSNet.

for $k = \text{mod}\,(j,m) + (N_x N_u)\Gamma\left(\lfloor \frac{j}{m} \rfloor\right)$, where $\Gamma(\cdot) : \mathbb{W} \to \mathbb{W}$ is the function that models the shearing operation, c_k is the value at the kth position of $\mathbf{c} \in \mathbb{R}^{(N_x N_u)\cdot(N_y N_v+\gamma\cdot(N_\lambda-1))}$, which is the vectorial version of the coded aperture $\mathbf{C}$ with $\gamma \in \mathbb{W}$ being the shearing magnitude factor. In the scope of this work, the coded aperture is set to be binary (i.e., $c_k \in \{0,1\}$) and the shearing magnitude to be an integer value (i.e., $\gamma \in \{0, N_y\}$).

7.2. Reconstruction algorithm

In the proposed D-CoSNet, the image reconstruction relies on the joint effort of two sequential stages: a set of parallel layers emulating the closed-form solution of CoSDI's inverse problem modeled by the ADMM [referred to as Module I in Fig. 80], and a U-net structure [referred to as Module II in Fig. 80]. By regularizing the compressed measurement, this connection increases the fidelity of the estimated three-dimensional datacube. Meanwhile, the parallel layers and the U-net manifest the â€œsplitting and optimizationâ€ approach embedded in the ADMM.

The ADMM-based inverse problem is incorporated by using the Eckstein and Bertsekas lemma Jonathan Eckstein y Dimitri P Bertsekas. “On the Douglasâ€”Rachford

splitting method and the proximal point algorithm for maximal monotone operators". En: *Mathematical Programming* 55.1-3 (1992), págs. 293-318, which is expressed as

$$\underset{\mathbf{f},\mathbf{w}}{\arg\min} \quad \tfrac{\mu}{2}\|\mathbf{w}-\mathbf{f}-\mathbf{d}\|_2^2 + P_1(\mathbf{f}) + P_2(\mathbf{w}), \tag{145}$$

where $P_1(\cdot):\mathbb{R}^n \to \bar{\mathbb{R}}$, $P_2(\cdot):\mathbb{R}^q \to \bar{\mathbb{R}}$ with $q \in \mathbb{N}$ as an arbitrary constant, $\mathbf{w} \in \mathbb{R}^{q\times 1}$ is a splitting variable that serves as the argument of $P_2(\cdot)$ with $\mathbf{w} = \mathbf{If}$, $\mu \geq 0$ is called the augmented lagrangian (AL) penalty parameter, and $\mathbf{d} \in \mathbb{R}^{n\times 1}$ is the Lagrange multiplier vector. To integrate this ADMM structure into the D-CoSNet, $P_1(\mathbf{f}) = \|\mathbf{\Phi f} - \mathbf{g}\|_2^2$ is established with $\mathbf{\Phi}$ generated via Eq. (161), and $P_2(\mathbf{w})$ is an auxiliary function where its structure is not pertinent to be defined for the PnP-ADMM-based approaches Xin Yuan y col. "Plug-and-play algorithms for large-scale snapshot compressive imaging". En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, págs. 1447-1457. Then, Eq. (162) is decoupled into two analytical inverse models

$$\mathbf{f} = \underset{\mathbf{f}}{\arg\min}\, \|\mathbf{\Phi f} - \mathbf{g}\|_2^2 + \frac{\mu}{2}\|\mathbf{w}-\mathbf{f}-\mathbf{d}\|_2^2, \tag{146}$$

and

$$\mathbf{w} = \underset{\mathbf{w}}{\arg\min}\, P_2(\mathbf{w}) + \frac{\mu}{2}\|\mathbf{w}-\mathbf{f}-\mathbf{d}\|_2^2, \tag{147}$$

Note that in Eq. (163), the analytical inverse model of **f** refers to a quadratic problem, while the analytical model in Eq. (164) refers to a denoising problem. To solve the first inverse model, in Eq. (163), we exploit the Sherman Woodbury Morrison (SWM) matrix inversion lemma William W Hager. "Updating the inverse of a matrix". En: *SIAM Review* 31.2 (1989), págs. 221-239 and the full-column rank properties to

obtain the closed-form solution that simultaneously involves the sensing process, the sensing matrix conditionality, and the gradient solution, which is given by:

$$\mathbf{f} = \tilde{\mu}^{-1}\left[\mathbf{I} - \mathbf{\Phi}^T\left(\tilde{\mu}\mathbf{I} + \mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}\right] \cdot \left[\mathbf{\Phi}^T\mathbf{g} + \tilde{\mu}^{-1}\left(\mathbf{w} - \mathbf{d}\right)\right], \tag{148}$$

where $\mathbf{\Phi}\mathbf{\Phi}^T \in \mathbb{R}^{m\times m}$ represents a matrix product resulting in a diagonal matrix, $\mathbf{I}$ the identity matrix with an arbitrary size, and $\tilde{\mu} = \mu/2$.

To construct the D-CoSNet using Eq. (165), the first step is to define the operators of CoSDI's data acquisition using the 3D scene, the coded aperture, and shearing coefficients. First, for the direct sensing process (whose layer is shown in magenta in Fig. 80), the operator is expressed by

$$\mathcal{G}(\mathbf{F}, \mathbf{C}) = \sum_{i_l=0}^{N_\lambda - 1} \mathbf{F}_{:,:,i_l} \circ \mathbf{C}_{:,\Gamma(i_l):N_yN_v+\Gamma(i_l)}, \tag{149}$$

where $\mathcal{G}(\cdot) : \mathbb{R}^{N_xN_u\times N_yN_v\times N_\lambda} \rightarrow \mathbb{R}^{N_xN_u\times N_yN_v}$ is the direct sensing operator. Then, the transpose sensing operator (whose layer is shown in red in Fig. 80) is defined as

$$\mathcal{F}(\mathbf{G}, \mathbf{C}) = \mathbf{G} \circ \mathbf{C}_{:,\Gamma(i_l):N_yN_v+\Gamma(i_l)}, \tag{150}$$

where $\mathcal{F}(\cdot) : \mathbb{R}^{(N_xN_u)\times(N_yN_v)} \rightarrow \mathbb{R}^{(N_xN_u)\times(N_yN_v)\times N_\lambda}$ is an operator that returns a datacube from a 2D compressed measurement. Finally, the inverse operator (whose layer is shown in brown in Fig. 80) is defined as

$$\mathcal{I}(\mathbf{G}, \mathbf{C}) = \mathbf{G} \circ \left[\left(\sum_{i_l=0}^{N_\lambda - 1} \mathbf{C}_{:,\Gamma(i_l):N_yN_v+\Gamma(i_l)}\right)^{\circ 2} + \tilde{\mu}\mathbf{I}\right]^{\circ -1}, \tag{151}$$

where $\mathcal{I}(\cdot) : \mathbb{R}^{N_xN_v\times N_yN_u} \rightarrow \mathbb{R}^{N_xN_v\times N_yN_u}$, $(\cdot)^{\circ 2}$ and $(\cdot)^{\circ -1}$ represent the Hadamard quadratic power and the Hadamard inverse operation, respectively.

Following the definition of these three SCI operators, the next step is to model the

SWM matrix approach in Eq. (163). Toward this goal, Eq. (163) is split into two main equations $\mathbf{\Phi}^T\mathbf{g} + \tilde{\mu}^{-1}(\mathbf{w} - \mathbf{d})$ and $\tilde{\mu}^{-1}\mathbf{I} - \tilde{\mu}^{-1}\mathbf{\Phi}^T\left(\tilde{\mu}\mathbf{I} + \mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}$. In the D-CoSNet, the first equation is reflected as a transpose operator coupled to two 2D convolutional layers (shown as cyan in Fig. 80), each of which are coupled with a ReLU activation function (referred to hereafter as a conventional+ReLU layer). These layers are located at the beginning of Module I, as shown in Fig. 80. Subsequently, the second equation is represented by two parallel arms. The upper arm, corresponding to $\tilde{\mu}^{-1}\mathbf{\Phi}^T\left(\tilde{\mu}\mathbf{I} + \mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}$, is composed of the direct sensing operator as a first layer followed by an inverse and transpose sensing operator along with four 2D convolutional+ReLU layers, resulting in a block composed of seven layers. The bottom arm, correspond to $\tilde{\mu}^{-1}\mathbf{I}$, has three 2D convolutional+ReLU layers, resulting in a block composed of three layers. The outputs of both arms are added and given as the input to the U-net [referred to as the Module II in Fig. 80] in the D-CoSNet that reflects the Eq. (164). Finally, the output from the U-Net block is the reconstructed datacube (denoted by $\tilde{\mathbf{F}}$). The loss function $\mathcal{L}(\cdot)$ used to learn the weights of the proposed end-to-end network is established as

$$\mathcal{L}(\mathbf{F}, \mathbf{G}) = \ell_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_1(\mathbf{G}, \tilde{\mathbf{G}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}}), \tag{152}$$

where $(\tilde{\mathbf{F}}, \tilde{\mathbf{G}})$ are the D-CoSNet outputs, l_1 is the l_1-norm, and l_{SSIM} represents the structural similarity index measure metric.

7.3. Simulation results

We used two popular databases for simulations: "Sparse Recovery of Hyperspectral Signal from Natural RGB Images"Boaz Arad y Ohad Ben-Shahar. “Sparse recovery of hyperspectral signal from natural RGB images”. En: *European Conference on Computer Vision*. Springer. 2016, págs. 19-34, and Çhallenge on Spectral Re-

construction from RGB Images"Boaz Arad y col. “Ntire 2020 challenge on spectral reconstruction from an rgb image”. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, págs. 446-447. All simulations were implemented in Tensorflow and trained on Google Colaboratory with a Tesla P100-PCI-E GPU using the ADAM optimizer Diederik P Kingma y Jimmy Ba. “Adam: A method for stochastic optimization”. En: *arXiv preprint arXiv:1412.6980* (2014). To train the coded aperture and the shearing weights, we randomly selected and cropped 450 datacubes with $N_xN_u \times N_yN_v = 256 \times 256$ and $N_\lambda = 25$. The same coded aperture was used to reconstruct the underlying datacube in the three Mu-CoSDI' depth-alternatives. The depth estimation is carried out via a third-party specialized depth algorithm –LF Jeon y col., “Accurate depth map estimation from a lenslet light field camera”, pure-CNN Ranftl y col., “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer”, and SL Yin y col., “Temporal phase unwrapping using deep learning”– adapted to receive as input the D-CoSNet output.

We conducted comprehensive ablation studies on how the terms $l_1(\mathbf{F}, \tilde{\mathbf{F}})$, $l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$, and $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ in the loss function (i.e., Eq. (169)) would affect the performance of D-CoSNet in reconstruction. The investigation of $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ was inspired by Hang Zhao y col. “Loss functions for image restoration with neural networks”. En: *IEEE Transactions on Computational Imaging* 3.1 (2016), págs. 47-57. Table 3 shows that implementing the $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$,which integrates the robust training capability of $l_1(\mathbf{F}, \tilde{\mathbf{F}})$ and the ability to preserve the detailed features using $l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$, allows the D-CoSNet to exceed reconstruction performance in comparison to when these two terms are considered individually by 1.15 dB and 0.63 dB, respectively (Table 3). Specifically, this study revealed that $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ shows a better reconstruction performance than when these two terms are considered individually, because $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ exploits l_1's robust training capability

Terms under analysis	PSNR (mean $\pm$ standard deviation) [dB]
$l_1(\mathbf{F}, \tilde{\mathbf{F}})$	25.64 $\pm$ 2.83
$l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$	26.16 $\pm$ 2.66
$l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$	26.79 $\pm$ 2.42

Tabla 2. Ablation analysis of various terms in the loss function

of datasets with outliers (i.e., l_1 does not over-penalize large errors) and the l_{SSIM}'s ability to preserve local structures and contrasts in high-frequency regions. These results are matched with the claims in Zhao y col., "Loss functions for image restoration with neural networks".

Finally, D-CoSNet shows a reconstruction speedup of ~ 600x against the proposed PnP-ADMM-based algorithm (see Section 6).

7.4. Proof of concept experiments

To demonstrate the Mu-CoSDI's multimodal capability, we conducted experimental validations using three representative depth modalities: LF, SL, and pure-CNN-based depth, as shown in Fig. 81. Specifically, the Mu-CoSDI testbed is inspired by the CS^{2}LI imager and constructed following a modular approach, i.e., capable of being coupled with either a lenslet array (to perform light field imaging) or a light projector (to conduct structured light-based acquisition.). Note that we replaced the DFD approach with the CNN-based depth approach to demonstrate the Mu-CoSDI's capabilities to be used in outdoor environments for applications that do not require high depth accuracy. The D-CoSNet returns spectral data cubes with their spectral signatures without illumination corrections, i.e., in intensity. Therefore, an additional offline reconstruction step is the spectral reflectance estimation, see Supplemental material A.

The Mu-CoSDI uses –for the LF scenario, as shown in Fig. 81(A)– a lenslets array (46$\times$46 mm, 4$\times$ 3 mm lenslets) coupled to a Navitar lens ($12mm$ FixedFocal Length, MVL12M23 - $12mm$ EFL, $f/1.4$). These two coupled lenses are the objective lens

that images the scene into the input plane of a Fizeau-based 4f geometry composed of a 50:50 matched achromatic doublet pair (L1)(Thorlabs MAP105050-A - 1:1), a polarizing beam splitter (BS)(Thorlabs CCM1-PBS251, 30 mm Cage Cube-Mounted), an $\lambda/4$ achromatic waveplate (Edmund #6-558, 25.4 mm. 465-610 nm), a 100:100 matched achromatic doublet pair (L2)(Thorlabs MAP10100100-A - 1:1), and mounted achromatic doublets lens (L3)(Thorlabs AC254-100-A-ML). L1 and L2 are located on the same propagation axis (i.e., the BM's transmission path), and L3 is located on the BM's opposite reflection path–following an L1-L2 propagation direction. L2 propagates the incoming wavefront through a customized double Amici prism (see Appendix 9) to image onto a coded mirror (CM). The CM is constructed using a broadband dielectric mirror (Thorlabs BB1-E02, 1" diameter) attached to a customized mask (FUJICHROME Velvia 50 transparency film). Taking advantage of the mirror surface, the now dispersed-modulated wavefront is reflected back through the prism until the L2 lens, where the prism undoes the undispersion effect. The resulting dispersed-coded-dispersed wavefront propagates through the BS until the L3, which finally images thewavefront onto the sensor (Stingray F-080B, $4.65\mu m$ pixel size). For the SL and CNN-based depth modalities, the Mu-CoSDI configuration is the same described previously but without the lenslet array, as shown in Fig. 81(B).
The experiments evaluate seven target scenes: three for SL, three for pure-CNN, and one for SL (to complement the results shown in Section 6). In the LF modality, the scenes were named Penguin, Astronaut, and Church. The scenes were named: Penguin, Astronaut, and Church [LF modality]; Students, Creator, and Palms [Pure-CNN modality]; Leaf [SL modality]. The Mu-CoSDI's measurements were acquired using the same customized binary-coded aperture generated from a random distribution with a transmittance level of 50 %. The resulting CA has a pixel size of $61.8\mu m \times 61.8\mu m$, as shown in Fig. 82. The compressed measurement exhibited 256×256 binned CCD pixels ($4.65\mu m \times 4.65\mu m$ pixel size, 12×12 binning) in size. The

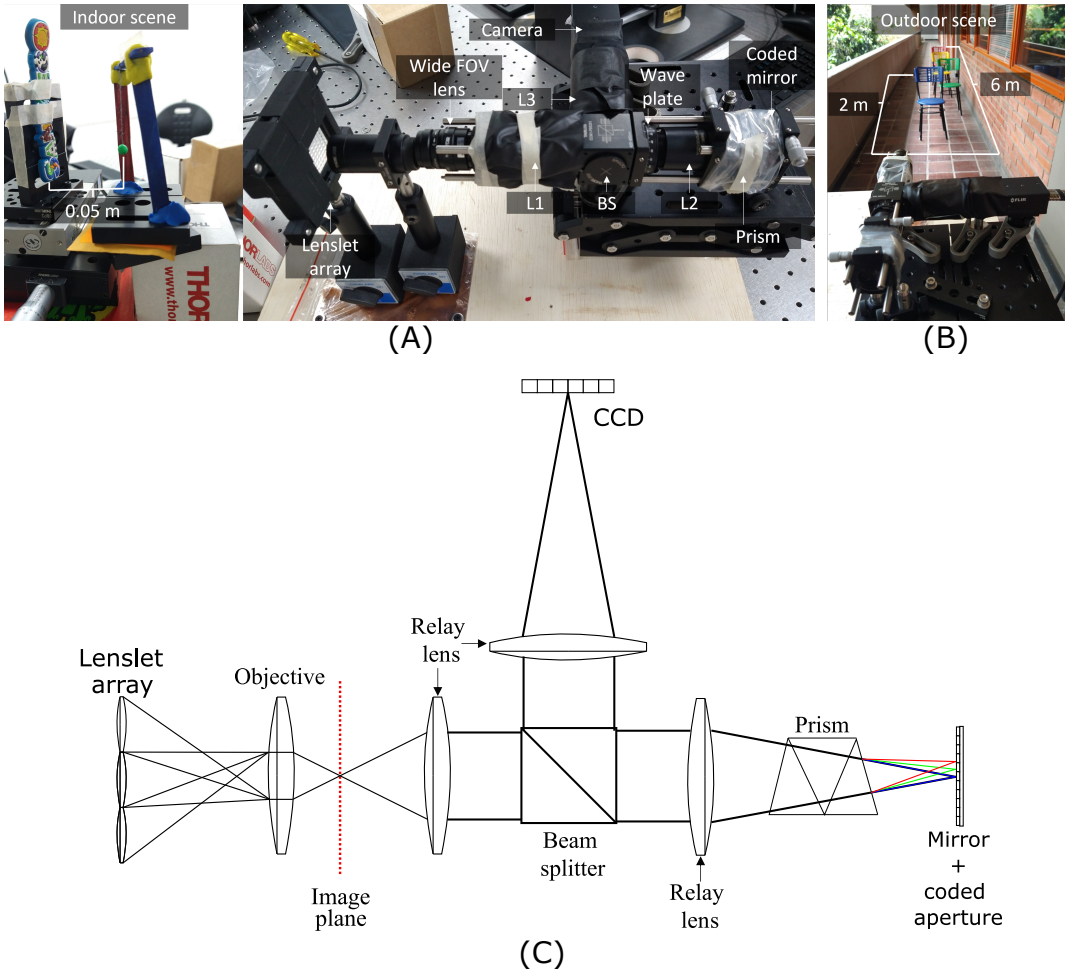


Figura 69. Mu-CoSDI's validation in indoor and outdoor scenarios. Mu-CoSDI configuration for (A) LF modality and (B) CNN-based depth modalities. (C) Geometric sketch of the optical system. Note that the systems in (A) and (B) share the same optical configuration except for the lenslet array.

D-CoSNet recovered a hyperspectral datacube of $N_x \times N_y \times N_L = 256 \times 256 \times 25$ in size. The Mu-CoSDI's LF modality was configured with a field-of-view of $2cm \times 2cm$ for scenes with a depth range of $5cm$ in an indoor environment. Here Penguin and Church targets are composed of two depth-layers and Astronaut is composed of three depth-layers. Insets (A) and (B) in Figs. 83-85 depict the CoSDI's measurements generated with a 5×3 lenslet array and the RGB composite of the reconstruction (obtained via D-CoSNet), respectively. Inset (C) in Fig. 85 shows the depth estimation

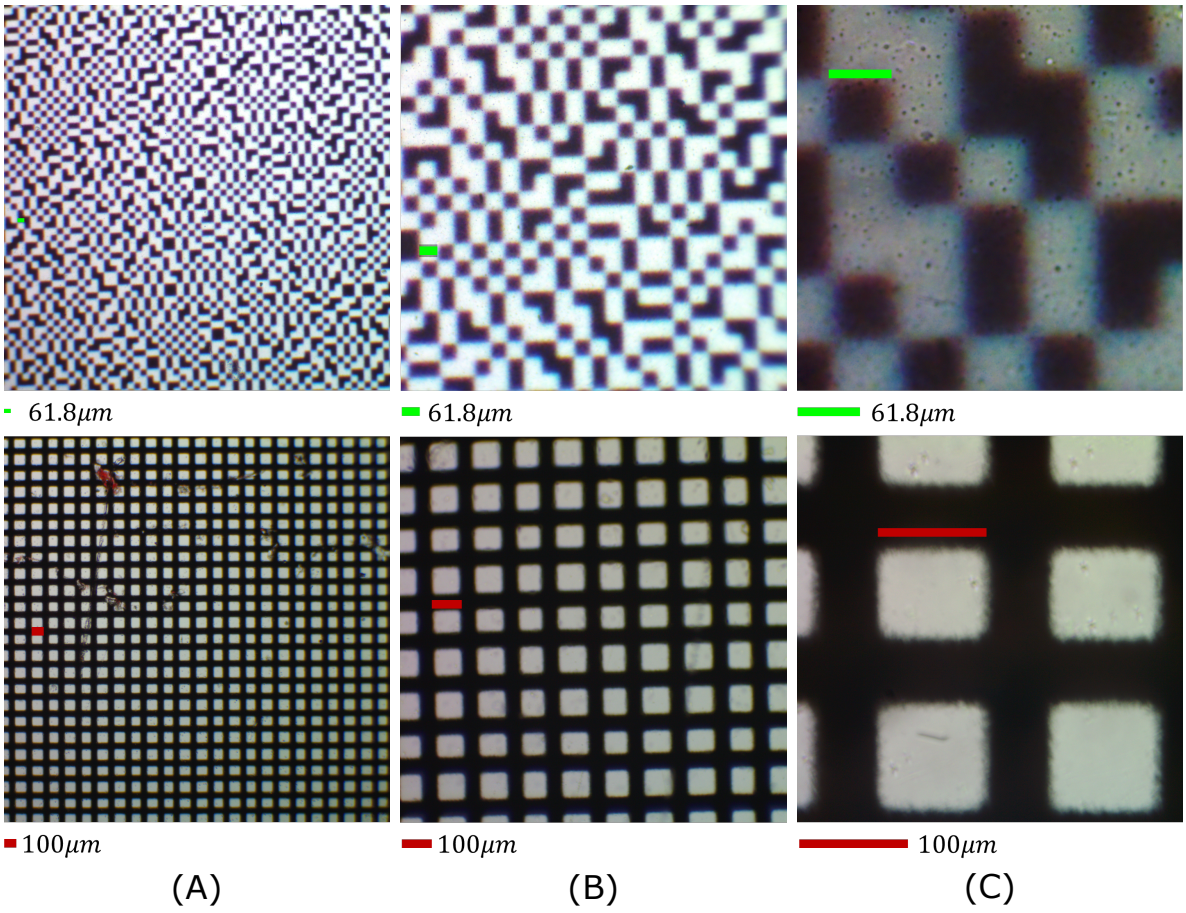


Figura 70. (First row) Customized coded aperture characterization using a (A) 4x, (B) 10x, and (C) 40x microscope objective lens. (Second row) Microscope micrometer calibration grid shaped (DIV=$100\mu m \times 100\mu m$).

obtained from the third-party algorithm Jeon y col., “Accurate depth map estimation from a lenslet light field camera” that receives the resulting D-CoSNet reconstruction as input. Note that the depth map estimations are alongside a cropped and zoomed version of Figures 83-85(B), where we can appreciate the parallax effect induced by the lenslet array (highlighted by dotted red circles) of the RGB composite. Finally, inset (D) in Figs. 83 to 85 depicts twelve representatives recovered spectral bands out of the 25. These results show that Mu-CoSDI has a high potential for endoscopy Jianyu Lin y col. “Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks”. En: *Medical*

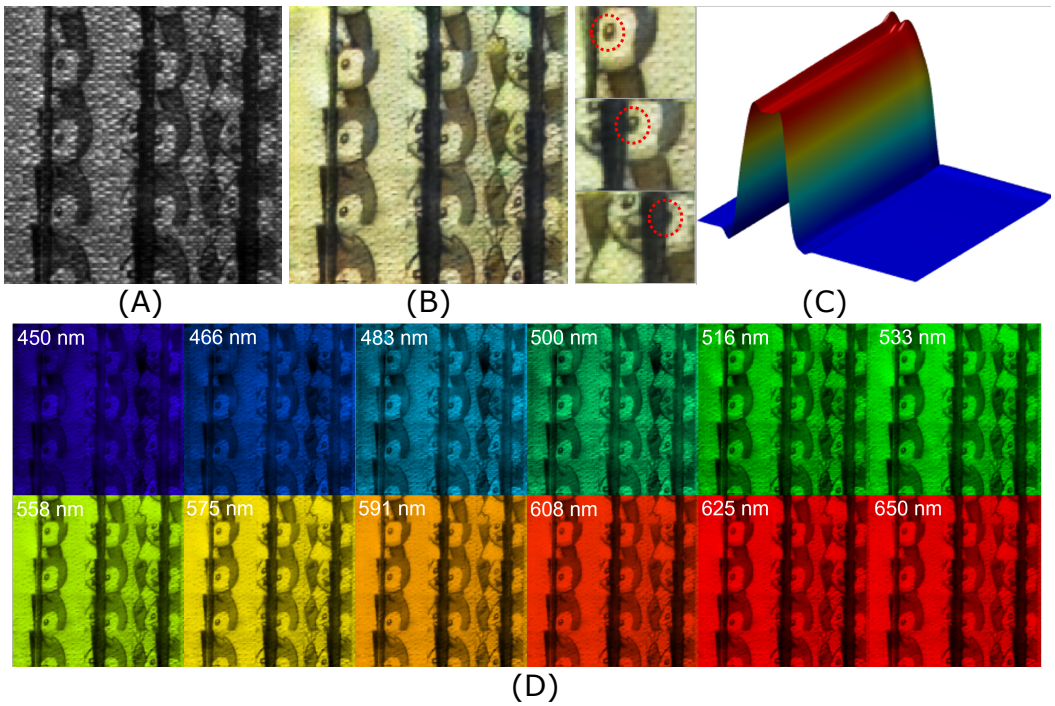


Figura 71. D-CoSNet reconstruction for Penguin scene in the LF modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Parallax illustration along with the resulting depth map estimation via Hae-Gon Jeon y col. “Accurate depth map estimation from a lenslet light field camera”. En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555. (D) Twelve selected wavelength.

Image Analysis 48 (2018), págs. 162-176 applications. The Mu-CoSDI's SL modality was configured with a wide field-of-view of $\sim 10cm \times 10cm$ for scenes with a depth range of $\sim 0.5cm$ in an indoor environment. Here Leaf target is composed of a continuous 3D surface representing the depth map. This scene was included to complement the results shown in Section 6. Figure 86(A) depicts the RGB composite of the reconstruction obtained via D-CoSNet from eight complementary structured light projections: four patterns per frequency. The structured light fringes projected by the video-beam are generated by the equation

$$I^{j}_{i_1,i_2} = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi\nu(i_1 - i_2)}{L} - \frac{2\pi j}{K}\right), \tag{153}$$

where ν is the frequency of fringe patterns, $K \in \mathbb{N}$ is the total of pattern projections per frequency, $L \in \mathbb{N}$ is the total number of pixels in the transverse direction of the fringe pattern, and $j = \{0, ..., K-1\}$. In this regard, the parameters selected for the experiments are $K = 4$, $L = 1388$, and $\nu = \{14, 8\}$. Due to the D-CosNet speeding up the reconstruction process, for convenience, we reconstruct a spectral data cube for each measurement. Then, we estimate estimated the spectral datacube and depth information via Eqs. (119) and (120), respectively. Specifically, Fig. 86(B) shows the depth estimation –in a 3D mapping– obtained from the third-party algorithm Yin y col., "Temporal phase unwrapping using deep learning". Finally, Figs. 86(C) depicts twelve representatives recovered spectral bands. These results echo the claims established in Section 6. Finally, the Mu-CoSDI's pure-CNN modality was configured with a wide field-of-view of $\sim 500cm \times 500cm$ for scenes with a depth range of $\sim 600cm$ in an outdoor environment. Here Students, Creator, and Palms targets are composed of three depth-layers. Figures 87-89(A) and (B) depict the CoSDI's measurements and the RGB composite of the reconstruction (obtained via D-CoSNet), respectively. Figures 89(C) shows the depth estimation –in a 2D and 3D mapping– obtained from the third-party algorithm Ranftl y col., "Towards Robust

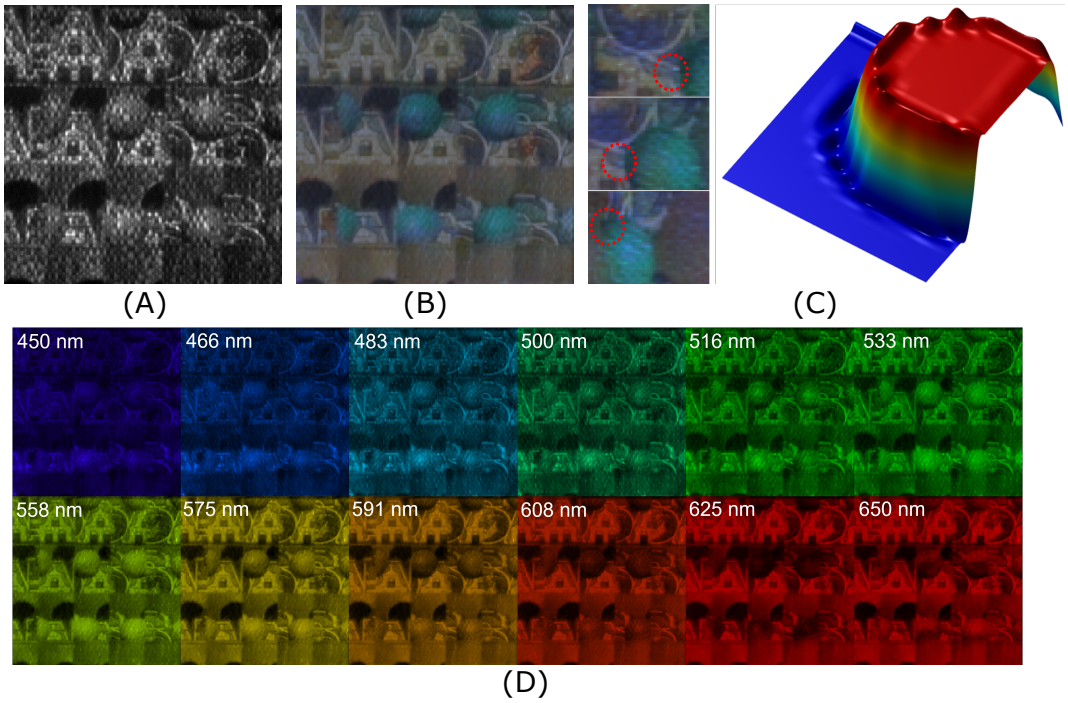


Figura 72. D-CoSNet reconstruction for Church scene in the LF modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Parallax illustration along with the resulting depth map estimation via Hae-Gon Jeon y col. “Accurate depth map estimation from a lenslet light field camera”. En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555. (D) Twelve selected wavelength.

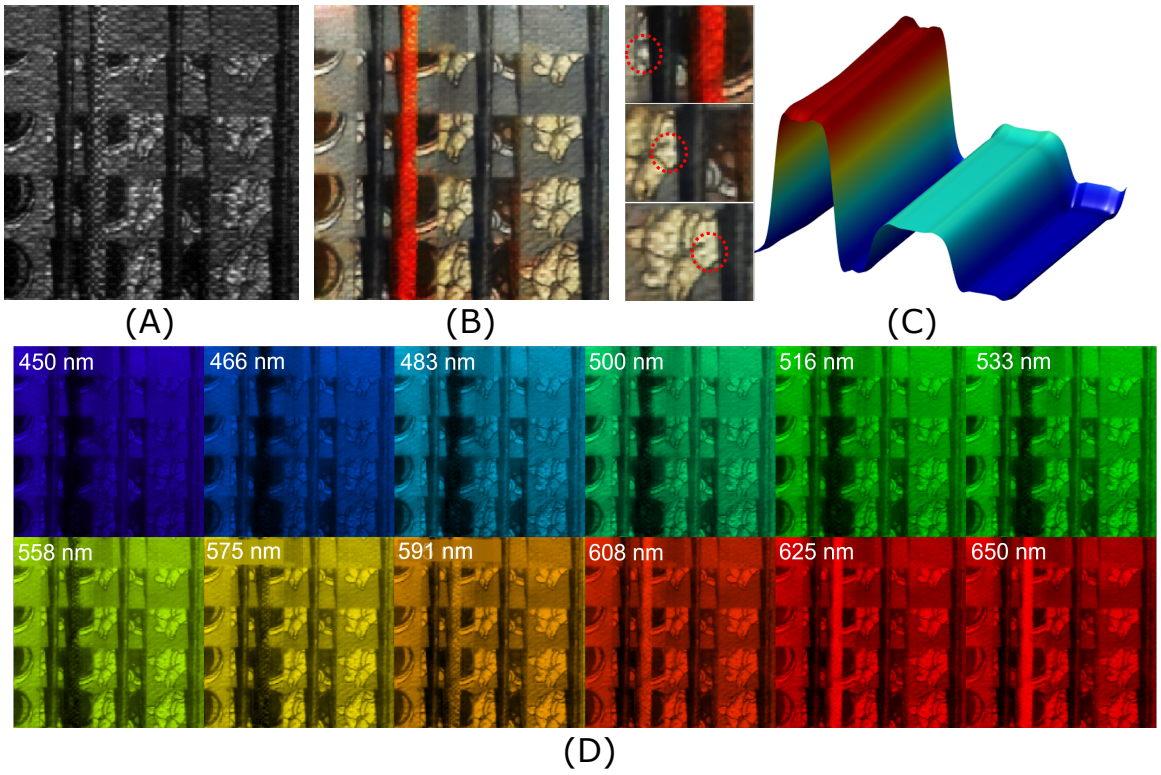


Figura 73. D-CoSNet reconstruction for Astronaut scene in the LF modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Parallax illustration along with the resulting depth map estimation via Hae-Gon Jeon y col. “Accurate depth map estimation from a lenslet light field camera”. En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555. (D) Illustration of twelve selected wavelength.

Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer" that receives the resulting D-CoSNet reconstruction as inputs. These depth-map results support the Mu-CoSDI's claims about the preservation of the spatial features [i.e., depth-dependent geometrical and achromatic aberrations] with good quality. Finally, Figs. 87-89(C) depicts twelve representatives recovered spectral bands. These results show that Mu-CoSDI has a high potential for outdoor applications such as tracking Yun Chen y col. "Remote sensing for vegetation monitoring in carbon capture storage regions: A review". En: *Applied Energy* 240 (2019), págs. 312-326, surveillance Mingde Yao y col. "Spectral-depth imaging with deep learning based reconstruction". En: *Optics Express* 27.26 (2019), págs. 38312-38325, and precision agriculture Uferah Shafi y col. "Precision agriculture techniques and practices: From considerations to applications". En: *Sensors* 19.17 (2019), pág. 3796.

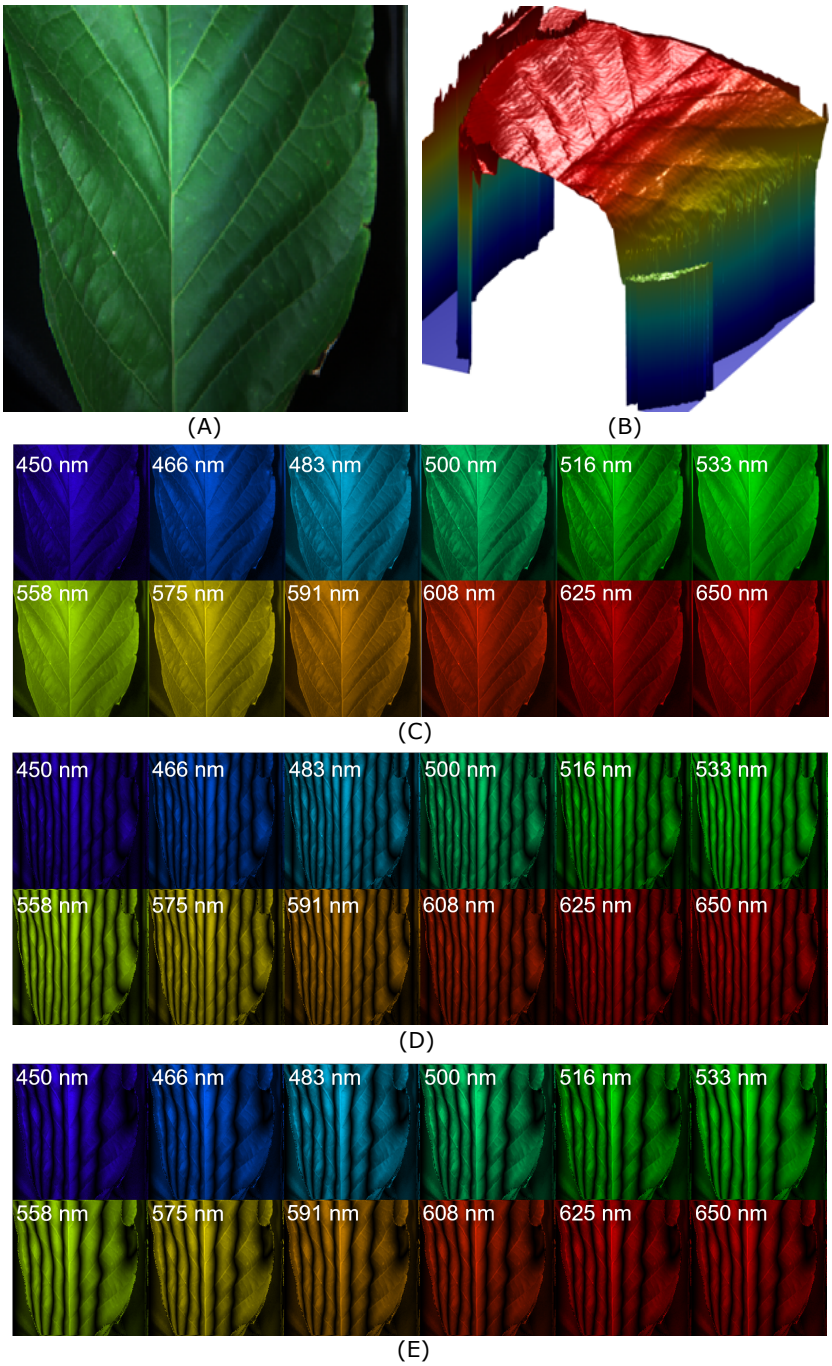


Figura 74. (A) RGB composite image from reconstructed datacube. (B) Depth map estimation from D-CoSNet output. D-CoSNet reconstruction for (C) full datacube (see Eq. (121)). (D)-(E) D-CoSNet reconstruction for the scene illuminated with a high ($\nu = 14$) and low frequency ($\nu = 8$) light patterns (see Eq. (170)), respectively.

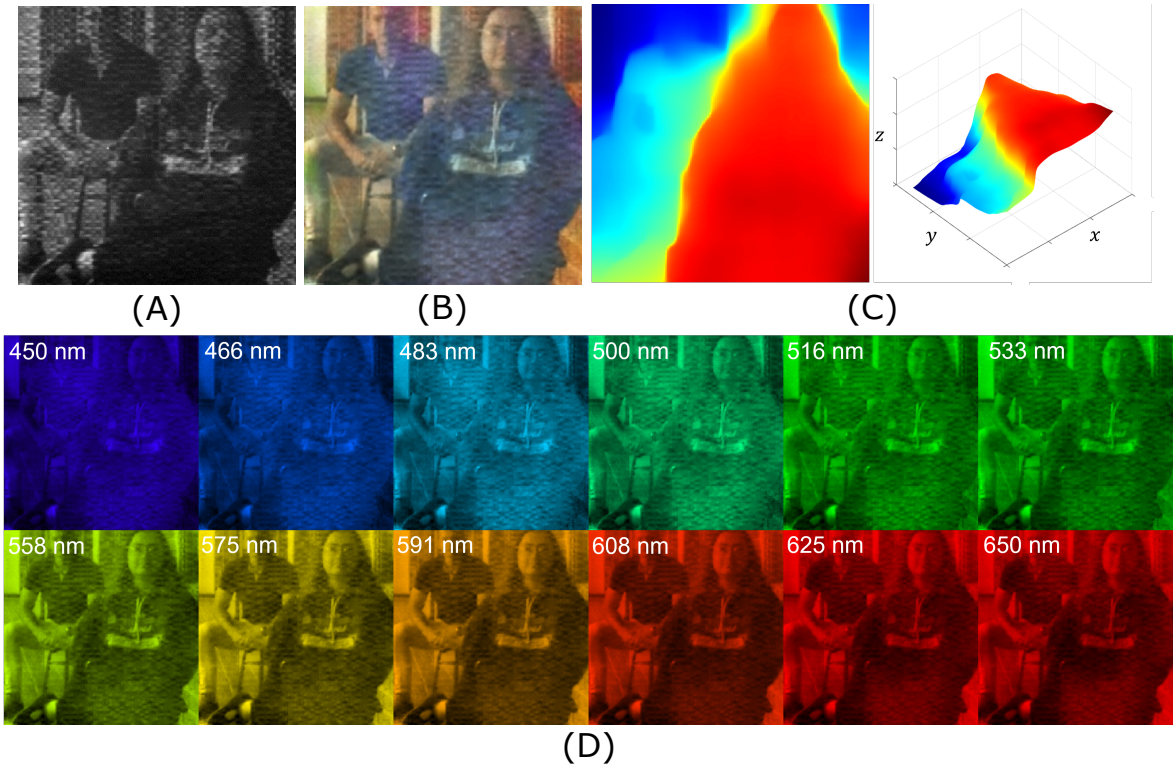


Figura 75. D-CoSNet reconstruction for Students scene in the pure-CNN modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Depth map estimation. (D) Illustration of twelve selected wavelength.

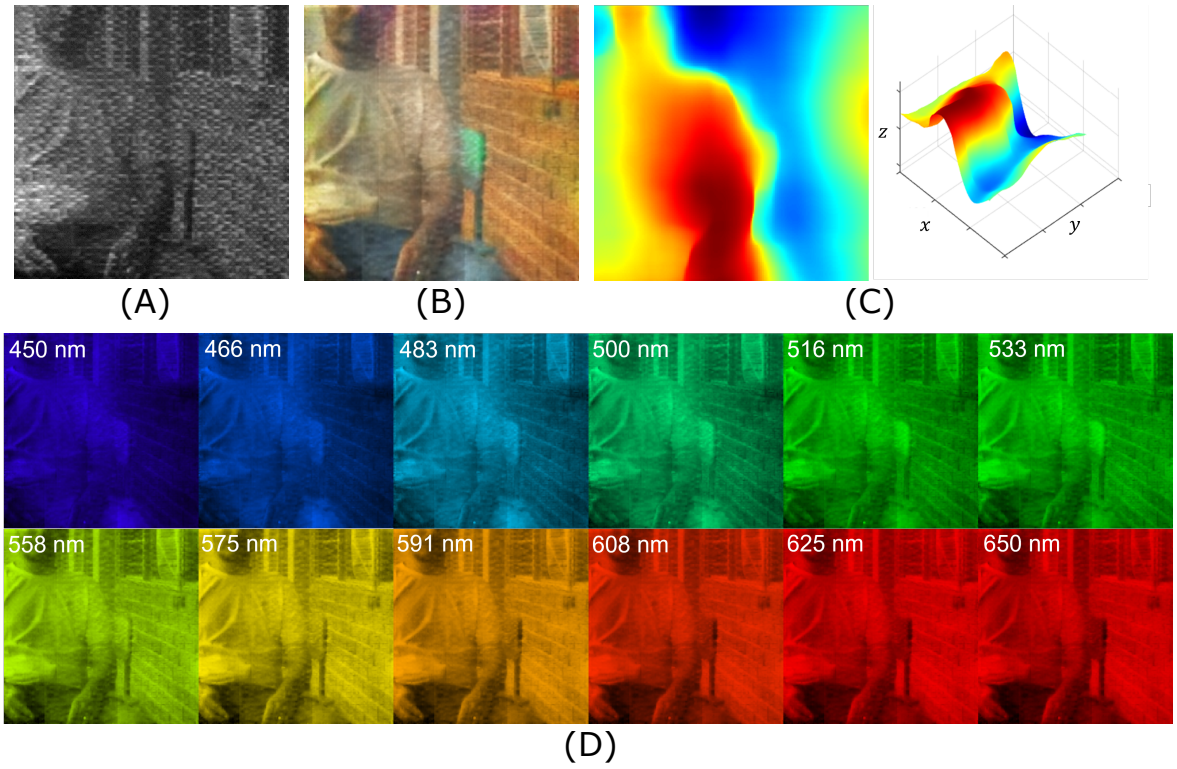


Figura 76. D-CoSNet reconstruction for Creator scene in the pure-CNN modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Depth map estimation. (D) Illustration of twelve selected wavelength.

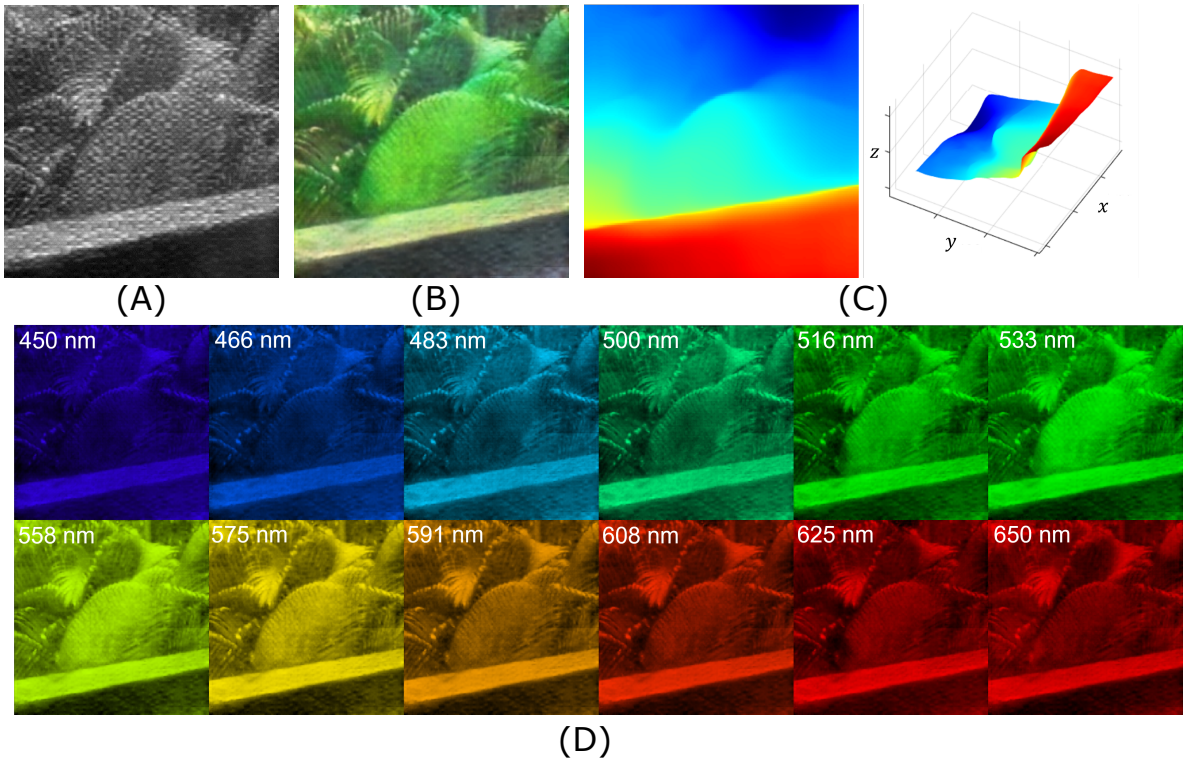


Figura 77. D-CoSNet reconstruction for Palms scene in the pure-CNN modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Depth map estimation. (D) Illustration of twelve selected wavelength.

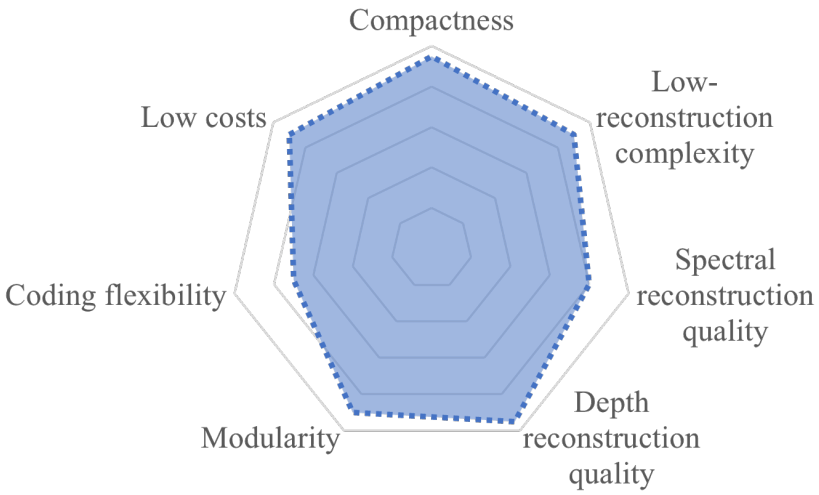


Figura 78. Chart graph for the Mu-CoSDI system

7.5. Conclusions

This chapter extended the CS^{2}LI sensing geometry to propose the Mu-CoSDI system capable of acquiring spectral depth images with a single sensor in three representatives depth modalities: SL, LF, and pure-CNN-based approach. Moreover, we have developed the D-CoSNet that enhances and speeds up the spectral reconstruction. Finally, experiments in indoor and outdoor environments were performed to verify the feasibility of the Mu-CoSDI. In general, the Mu-CoSDI is envisioned as a feasible system to being implemented either on active or passive illumination geometries that simultaneously implement multiples depth modalities, such as SL+LF, LF with a pure CNN-based approach, SL with a pure CNN-based approach, and the three modalities at the same time. D-CoSNet's architecture is highly suitable for solving the inverse problem in Mu-CoSDI. First, the ADMM-based inverse problem is implemented into the first three blocks in the D-CoSNet (see Fig. 80) to promote a higher resonance between the D-CoSNet's weights and the CoSDI's sensing geometry. In

particular, a set of two parallel layers are built to model the direct sensing, transpose process, and inverse operations involved in SCI's inverse problems (see Eq. (166)-(168)). Moreover, the matrix inversion is solved by taking advantage of the SWM formula and the SCI's diagonal matrix property (see Eq. (166)). Finally, the SCI's direct sensing operator is added at the end of D-CoSNet to regularize the estimated datacube. This attractive feature results in a more robust CNN that speeds up the datacube reconstruction.

As a chapter summary and to assess the Mu-CoSDI's pros and cons, we construct a radar chart with seven relevant CoSDI features: compactness, low costs, low-recovery complexity, spectral recovery, depth recovery, modularity, and encoding flexibility. Here the bottom and the top shell represent the worst and best scores, respectively, i.e., the ideal CoSDI system would have to reach the upper contours in the seven features. Specifically, Fig. 90 illustrates the Mu-CoSDI's radar chart, showing that the Mu-CoSDI achieves an optimal balance against the previously proposed CASSDI-DFA, C^3LFI, and CS^2LI systems.

8. Mu-CoSDI: Multimodal Compressive Spectral Depth Imager

In the previous chapters, we introduced three compressive spectral depth imagers (CoSDI) that couple the coded aperture snapshot spectral imaging (CASSI) sensing geometry (introduced in **Chapter 3**) with three depth-sensing modalities: depth-from-aberration (DFA), light field (LF), and structured light (SL). Specifically, **Chapter 4** presented the coded aperture snapshot spectral imager via DFD (CASSI-DFA) based on a hybrid phase and amplitude modulation geometry. The CASSI-DFA has high coding flexibility –associated with the high-speed refresh rate of optoelectronic DM and DMD devices– making it a good alternative for applications where the acquisition of multiple measurements is feasible. **Chapter 5** presented a compressive color-coded compressive spectral light field imager (C^3SLFI) based on two-complementary color-coded apertures, a lenslet array, and a single sensor. The C^3SLFI allows accurate high spectral and depth information recovery by estimating ad-hoc representation bases via tensor theory. Finally, **Chapter 6** introduced the compressive spectral structured light imaging (CS^2LI) system to emulate a double-dispersive CASSI system from a double-Amici prism and a coded mirror. The CS^2LI offers a compact optical distribution that enables it to be coupled –in a modular way– to either a lenslet array (in front of the objective lens) or a light projector. The CS^2LI' modular structure has a high potential to become the baseline for a multimodal CoSDI system. Although the proposed CoSDI systems allow accurate spectral and depth recovery, the analytical-based reconstruction algorithms are overshadowed by the reconstruction speed up offered by the recently proposed convolutional neural network (CNN) approaches. For instance, CNN has been established as a crucial compressive spectral imaging constituent to enhance the reconstruction accuracy and alleviate the computational complexity Meng y col., “Self-supervised neural networks for spectral snapshot compressive imaging”.

The positive impact of CNN to alleviate the computational complexity and to improve the reconstruction quality along with the CS^{2}LI's modularity leads naturally to the question: ***Is it possible to develop a multimodal CoSDI system enhanced by the CNN paradigm without relying on expensive encoding optical devices and camera arranges?***

To answer this question, this chapter extends the CS^{2}LI sensing geometry to build a multimodular CoSDI system capable of working on either active or passive depth paradigms, e.g., SL or LF. Moreover, to enhance and speed up the spectral reconstruction, we developed an end-to-end convolutional neural network named deep compressive spectral net (D-CoSNet). Here, D-CoSNet is constructed following an alternating direction method of multipliers (ADMM) to transfer many advantageous mathematical properties embedded in the main CoSDI sensing operations (direct sensing, transposed, and inverse). We have coined the latter, multimodal compressive spectral depth imager â€“ Mu-CoSDI.

8.1. Continuous and discrete sensing model

Advanced image reconstruction methods that provide high reconstructed image quality have become indispensable in the compressive spectral depth imaging (CoSDI) framework. To date, analytical models are predominantly used in compressive spectral imagers (CSIs). These methods solve the inverse problem by using the prior knowledge of the sensing matrix and by leveraging the sparsity of high-dimensional images in certain domains (e.g., spatial, wavelet, discrete cosine transform, and gradient domains). However, they require precise prior knowledge of all sensing operations [i.e., sensing direct, transpose, and inverse] involved in the CSI systems. Moreover, their processing time increases exponentially in high-dimensional optimization problems, and the reconstructed image quality highly depends on the empirical tuning of parameters.

The increasing availability of vast amounts of image data and computational power has consolidated convolutional neural networks (CNNs) as one of the most potent and desired tools for CSI approaches Kulkarni y col., *Reconnet: Non-iterative reconstruction of images from compressively sensed measurements*; Jin y col., “Deep convolutional neural network for inverse problems in imaging”; Kamilov y Mansour, “Learning optimal nonlinearities for iterative thresholding algorithms”; Meng, Jalali y Yuan, “Gap-net for snapshot compressive imaging”; Wu, Zhang y Mou, “Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging”; Meng y col., “Self-supervised neural networks for spectral snapshot compressive imaging”. CNN’s potential for compressive sensing was initially demonstrated to reconstruct images from measurements generated by a random Gaussian matrix Kulkarni y col., *Reconnet: Non-iterative reconstruction of images from compressively sensed measurements*. Inspired by this early development, CNN paradigms were integrated into CSI’s reconstruction by combining its forward model as an encoder, followed by convolutional layers as a fast approximated sparse decoder Jin y col., “Deep convo-

lutional neural network for inverse problems in imaging". Despite high reconstructed image quality in simulation, these early CNN-based approaches produced a dramatically decreased reconstruction quality for real experiments. This limited performance was attributed to the fact that CNNs did not consider the sensing operators associated to the SCI models (e.g., direct, transpose, and inverse processes) that could work as prior information to improve the reconstruction results. In this path, subsequent works tackled this problem by taking advantage of the well-established link between analytical-modeling approaches and deep learning networks Kamilov y Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms", resulting in the development of the unrolling approach Monga, Li y Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing"; Zheng y col., "Deep plug-and-play priors for spectral snapshot compressive imaging" and the plug-and-play (PnP) method Zheng y col., "Deep plug-and-play priors for spectral snapshot compressive imaging", both of which relied on iterative gradient descents models for the inferences. However, the unrolling approaches were highly prone to suffer vanishing gradient issues as the number of emulated iterations increased. The PnP approaches limited the CNN's inference potential by using them as simple denoisers. Thus, most existing CNN-based methods are merely used to replace analytical-modeling-based image reconstruction.

To overcome these limitations, we develop a convolutional neural network, termed deep compressive spectral net (D-CoSNet), for reconstructing 3D spectral datacubes from measurements generated by the Mu-CoSDI in either of its three modalities: light field (LF), structured light (SL), and pure-CNN-based depth approach. The Mu-CoSDI's sensing geometry is an extension of CS^2LI (see chapter 6), and their multimodal construction is illustrated in Fig. 79. The D-CoSNet is envisioned to benefit CoSDI in image reconstruction by transferring many advantageous properties embedded in certain mathematical models into custom layers. Towards this goal, the

D-CoSNet comprises two main sequential stages: a deep-unfolding-based network to embody SCI's sensing model and a U-net Ronneberger, Fischer y Brox, "U-net: Convolutional networks for biomedical image segmentation" that works as a filter. The first stage is developed from the structure of an alternating direction method of multipliers (ADMM) Boyd y col., "Distributed optimization and statistical learning via the alternating direction method of multipliers", that integrates SCI's major operations with a mathematical coherence, emulated as three custom layers constructed by using the coded aperture and the shearing function.

8.1.1. Continuous sensing model For the D-CoSNet construction, we first require to establish a general sensing model for the proposed multimodal CoSDI, that encompasses the models described by Eqs. (56) (see Section 4), (**??**) (see Section 5) and (112) (see Section 6). In this modeling and for convenience, we referred to the light source object as the objective lens output wavefront, see 79 (A) represented with a dotted red line (image plane). This assumption allows us to develop a spectral CNN-based reconstruction methodology dockable to the Mu-CoSDI's multimodalities, e.g., LF, SL, and pure-depth approach.

Let $f_o(x, y, \lambda, u, v)$ be the spatial spectral-angular scene in intensity, where (x, y) index the spatial coordinates, (u, v) index the view-angle dimension, and λ index the wavelength. The object's depth map is encoded by illuminating it with a structured light pattern, where the resulting pattern deformations are associated to the object shape. The pattern deformation can be modeled as a phase modulation, where the object's shape acts as the phase mask. Based on Eq. (11), the object shape-modulated structured light can be modeled as

$$p(x, y, \lambda, u, v, t) = f_o(x, y, \lambda, u, v) \left[\cos\left(\varphi(x, y, u, v) + \omega_0 t\right) + b\right] + a(x, y, u, v), \quad (154)$$

where $\varphi(x, y, u, v)$ is the object phase/shape information, ω_0 represents the temporal-

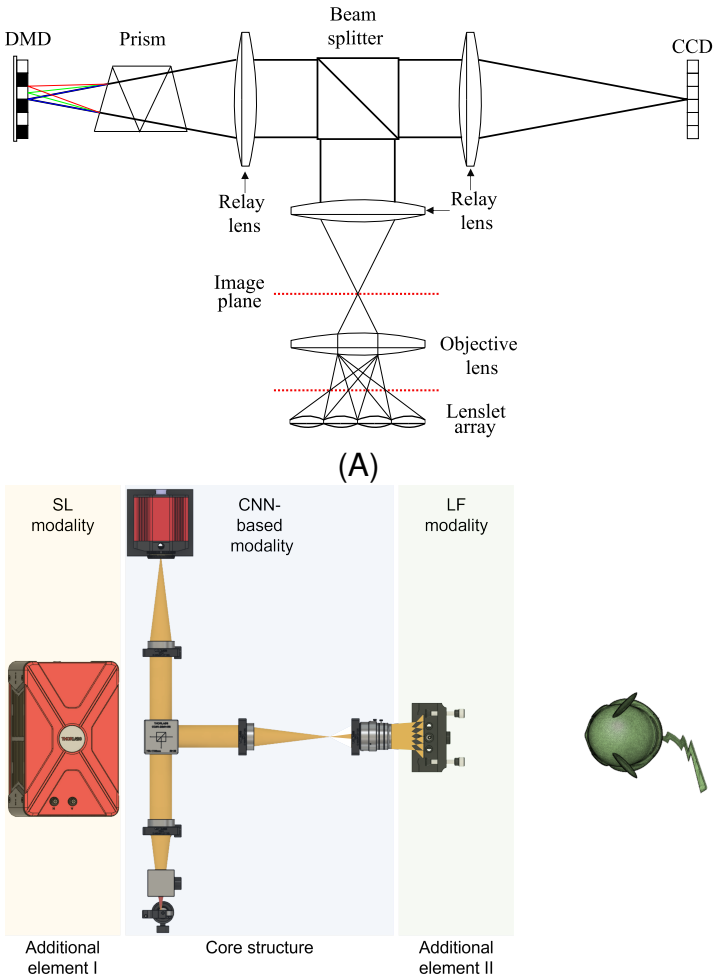


Figura 79. (A) Sketch of the proposed Mu-CoSDI architecture. (B)-(C) Upper and lateral sketch view where the distance configuration and the optical element labels are specified, respectively.

carrier, $a(x, y, u, v)$ is the background, and b is a balance constant that prevents $p(x, y, \lambda, u, v, t)$ achieve negative values with $b = 1$. First, an objective lens coupled with a lenslet array (LA) forms an image in an intermediate image plane, resulting in the wave field

$$f_1(x', y', \lambda, t) \iint \iint \gamma(\lambda) p(x, y, \lambda, u, v, t) h_{LA}(x - x', y - y', \lambda, u, v) dx dy du dv, \quad (155)$$

where $h_{LA}(\cdot)$ is the incoherent PSF of the objective lens + LA and $\gamma(\lambda)$ represent the spectral source density. This 4D to 2D parameterization (without considering the spectral dimension) was proposed in Adelson y Wang, "Single lens stereo with a plenoptic camera", in which the view angle dimension is collapsed into the spatial dimension. In this regard, the conventional light field systems, the previously proposed C^{3}SLFI (see Chapter 5), and consequently the proposed Mu-CoSDI system, aim to reconstruct the spectral information in this 2D projection. objective lens + LA forms image in the input plane of a 4f system, which propagates the incoming wavefront through a prism that introduces a wavelength-dependent horizontal shifting to form a dispersed image version over a coded aperture attached to a mirror (or DMD). The resulting dispersed-encoded wavefront can be seen in the detector plane (immediately before the sensor) as

$$f_2(x'', y'', \lambda, t) = k(x'', y'', \lambda) \iint h_2(x' - x'' - S_1(\lambda), y' - y'', \lambda) f_1(x', y', \lambda, t) dx' dy' \quad (156)$$

with

$$k(x'', y'', \lambda) = \frac{1}{2} \sum_{i_1', i_2, i_3} \text{rect} \left(\frac{x''}{\Delta_c} - i_1, \frac{y''}{\Delta_c} - i_2, \frac{\lambda}{\Delta_\lambda} - i_3 \right) C_{i_1, i_2, i_3}, \quad (157)$$

where $rect(\dot{})$ represents the rectangle function, $S_1(\lambda)$ is a wavelength-dependent function for the propagation from the relay lens to the mirror; Δ_c denotes the pixel size,

Δ_λ denotes the spectral bandwidth, $C \in \mathbb{R}^(N_y \times (N_x + N_\lambda - 1) \times N_\lambda)$ is a color-coded aperture with $C_(i'_1, i_2, i_3) \in 0, 1$, and $i'_1 = 0, ..., (N_x + N_\lambda - 2)$, $i_2 = 0, , N_y - 1$ and $i_3 = 0, ..., N_\lambda - 1$ indexing the rows, columns, and spectral bands, respectively. Then, the color-coded mirror reflects-back the dispersed-modulated wavefront to the prism, which introduce a second spectral dispersion. The resulting dispersed-coded-dispersed wavefront is propagated and focused into a detector array. The sensing process can be expressed as

$$f_3(x''', y''', \lambda, t) = \iint f_2(x'', y'', \lambda, t) h_3(x'' - x''' + S_2(\lambda), y'' - y''', \lambda) dx'' dy'', \quad (158)$$

where h_3 represents the intensity PSF and $S_2(\lambda)$ is a wavelength-dependent function for the propagation from the mirror to the relay lens (i.e., backpropagating through the prism). Finally, the measurement at the $(i_1, i_2)^{th}$ pixel is represented by

$$g_{i_1,i_2} = \int_\Gamma \rho(\lambda) \iint f_3(x''', y''', \lambda) \text{rect}\left(\frac{x'''}{\Delta_d} - i_1, \frac{y'''}{\Delta_d} - i_2\right) dx''' dy''' d\lambda, \quad (159)$$

where Λ represents the wavelength axis over the spectral range, Δ_p represents the pixel's side length, and $\rho(\lambda)$ represents the detector's normalized quantum efficiency. Note that the notion of a "multi-modality"sensing model in Eq. (159) also includes the three configurations illustrated in Fig. 79: (i) single view by relying just on the core structured; (ii) multi-view by relying on the core structure and the lenslet array (referred to as Additional element I in Fig. 79 (B); (iii) single view with structured light by relying on the core structure and the video beam (referred to as Additional element II in Fig. 79 (B)).

8.1.2. Discrete sensing model Based on Eq. (**??**), the data acquisition of Mu-CoSDI consists of the following three operations. First, a 3D scene, denoted by **F**, is horizontally dispersed (denominated shearing operation) to be spatially modulated

by a single 2D coded aperture. The resulting dispersed-modulated information is then undispersed to obtain a dispersed-coded-undispersed wavefront. Finally, via the operation of integration, a 2D detector integrates the encoded and sheared scene as a 2D image, commonly known as a compressed measurement, denoted by **G** Cao y col., “Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world”. The vectorized discrete forward model is expressed as

$$\mathbf{g} = \mathbf{\Phi f} + \boldsymbol{\epsilon}, \tag{160}$$

where $\mathbf{f} \in \mathbb{R}^{n\times 1}$ is the vector representation of a discrete version of the 3D scene $\mathbf{F} \in \mathbb{R}^{N_x N_u \times N_y N_v \times N_\lambda}$ with size $n = (N_x N_u)\cdot(N_y N_v)\cdot N_\lambda$; $N_x N_v$ and $N_y N_v$ represent the data lengths in the two spatial dimensions with $N_u \times N_v$ being the number of lenses in the lenslet array and $N_x \times N_y$ as the number of pixels along x and y contained in each view of each lens within the lenslet, and N_λ represents the data length in the spectral dimension. $\mathbf{g} \in \mathbb{R}^{m\times 1}$ is the vectorial version of the compressed measurement $\mathbf{G} \in \mathbb{R}^{N_x N_u \times N_y N_v}$ with $m = (N_x N_u)\cdot(N_y N_v)$; $\boldsymbol{\epsilon} \in \mathbb{R}^{m\times 1}$ represents the noise which is modeled with a colored Gaussian distribution; $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ is the CSI's sensing matrix. The structure of Φ accounts for the optical elements in the CSI system, and its entries are given by

$$\Phi_{i,j} = \begin{cases} c_k, & if\ i = \text{mod}(j,m) \\ 0, & otherwise \end{cases}, \tag{161}$$

for $k = \text{mod}\,(j,m) + (N_x N_u)\Gamma\left(\lfloor \frac{j}{m} \rfloor\right)$, where $\Gamma(\cdot) : \mathbb{W} \to \mathbb{W}$ is the function that models the shearing operation, c_k is the value at the kth position of $\mathbf{c} \in \mathbb{R}^{(N_x N_u)\cdot(N_y N_v + \gamma\cdot(N_\lambda - 1))}$, which is the vectorial version of the coded aperture **C** with $\gamma \in \mathbb{W}$ being the shearing magnitude factor. In the scope of this work, the coded aperture is set to be binary (i.e., $c_k \in \{0,1\}$) and the shearing magnitude to be an integer value (i.e., $\gamma \in \{0, N_y\}$).

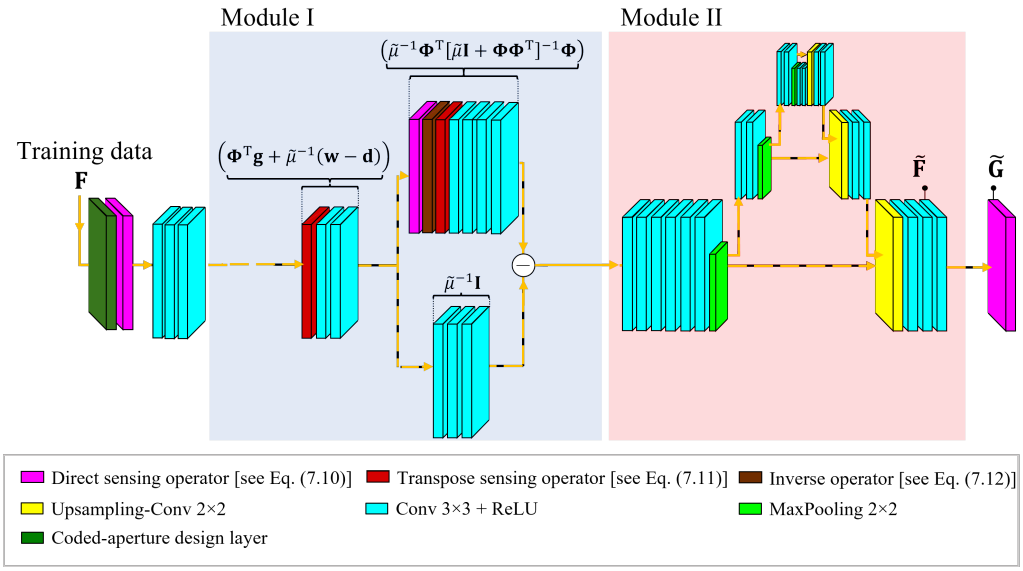


Figura 80. Schematic of D-CoSNet.

8.2. Reconstruction algorithm

In the proposed D-CoSNet, the image reconstruction relies on the joint effort of two sequential stages: a set of parallel layers emulating the closed-form solution of CoSDI's inverse problem modeled by the ADMM [referred to as Module I in Fig. 80], and a U-net structure [referred to as Module II in Fig. 80]. By regularizing the compressed measurement, this connection increases the fidelity of the estimated three-dimensional datacube. Meanwhile, the parallel layers and the U-net manifest the â€œsplitting and optimizationâ€ approach embedded in the ADMM.

The ADMM-based inverse problem is incorporated by using the Eckstein and Bertsekas lemma Eckstein y Bertsekas, “On the Douglasâ€”Rachford splitting method and the proximal point algorithm for maximal monotone operators”, which is expressed as

$$\underset{\mathbf{f},\mathbf{w}}{\arg\min} \quad \tfrac{\mu}{2}\|\mathbf{w}-\mathbf{f}-\mathbf{d}\|_2^2 + P_1(\mathbf{f}) + P_2(\mathbf{w}), \tag{162}$$

where $P_1(\cdot):\mathbb{R}^n \to \bar{\mathbb{R}}$, $P_2(\cdot):\mathbb{R}^q \to \bar{\mathbb{R}}$ with $q \in \mathbb{N}$ as an arbitrary constant, $\mathbf{w} \in \mathbb{R}^{q\times 1}$ is

a splitting variable that serves as the argument of $P_2(\cdot)$ with $\mathbf{w} = \mathbf{If}$, $\mu \geq 0$ is called the augmented lagrangian (AL) penalty parameter, and $\mathbf{d} \in \mathbb{R}^{n\times 1}$ is the Lagrange multiplier vector. To integrate this ADMM structure into the D-CoSNet, $P_1(\mathbf{f}) = \|\mathbf{\Phi f} - \mathbf{g}\|_2^2$ is established with $\mathbf{\Phi}$ generated via Eq. (161), and $P_2(\mathbf{w})$ is an auxiliary function where its structure is not pertinent to be defined for the PnP-ADMM-based approaches Yuan y col., "Plug-and-play algorithms for large-scale snapshot compressive imaging". Then, Eq. (162) is decoupled into two analytical inverse models

$$\mathbf{f} = \arg\min_{\mathbf{f}} \|\mathbf{\Phi f} - \mathbf{g}\|_2^2 + \frac{\mu}{2}\|\mathbf{w} - \mathbf{f} - \mathbf{d}\|_2^2, \tag{163}$$

and

$$\mathbf{w} = \arg\min_{\mathbf{w}} P_2(\mathbf{w}) + \frac{\mu}{2}\|\mathbf{w} - \mathbf{f} - \mathbf{d}\|_2^2, \tag{164}$$

Note that in Eq. (163), the analytical inverse model of **f** refers to a quadratic problem, while the analytical model in Eq. (164) refers to a denoising problem. To solve the first inverse model, in Eq. (163), we exploit the Sherman Woodbury Morrison (SWM) matrix inversion lemma Hager, "Updating the inverse of a matrix" and the full-column rank properties to obtain the closed-form solution that simultaneously involves the sensing process, the sensing matrix conditionality, and the gradient solution, which is given by:

$$\mathbf{f} = \tilde{\mu}^{-1}\left[\mathbf{I} - \mathbf{\Phi}^T\left(\tilde{\mu}\mathbf{I} + \mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}\right] \cdot \left[\mathbf{\Phi}^T\mathbf{g} + \tilde{\mu}^{-1}\left(\mathbf{w} - \mathbf{d}\right)\right], \tag{165}$$

where $\mathbf{\Phi}\mathbf{\Phi}^T \in \mathbb{R}^{m\times m}$ represents a matrix product resulting in a diagonal matrix, **I** the identity matrix with an arbitrary size, and $\tilde{\mu} = \mu/2$.

To construct the D-CoSNet using Eq. (165), the first step is to define the operators of CoSDI's data acquisition using the 3D scene, the coded aperture, and shearing

coefficients. First, for the direct sensing process (whose layer is shown in magenta in Fig. 80), the operator is expressed by

$$\mathcal{G}(\mathbf{F},\mathbf{C}) = \sum_{i_l=0}^{N_\lambda-1} \mathbf{F}_{:,:,i_l} \circ \mathbf{C}_{:,\Gamma(i_l):N_yN_v+\Gamma(i_l)}, \tag{166}$$

where $\mathcal{G}(\cdot) : \mathbb{R}^{N_xN_u\times N_yN_v\times N_\lambda} \rightarrow \mathbb{R}^{N_xN_u\times N_yN_v}$ is the direct sensing operator. Then, the transpose sensing operator (whose layer is shown in red in Fig. 80) is defined as

$$\mathcal{F}(\mathbf{G},\mathbf{C}) = \mathbf{G} \circ \mathbf{C}_{:,\Gamma(i_l):N_yN_v+\Gamma(i_l)}, \tag{167}$$

where $\mathcal{F}(\cdot) : \mathbb{R}^{(N_xN_u)\times(N_yN_v)} \rightarrow \mathbb{R}^{(N_xN_u)\times(N_yN_v)\times N_\lambda}$ is an operator that returns a datacube from a 2D compressed measurement. Finally, the inverse operator (whose layer is shown in brown in Fig. 80) is defined as

$$\mathcal{I}(\mathbf{G},\mathbf{C}) = \mathbf{G} \circ \left[\left(\sum_{i_l=0}^{N_\lambda-1} \mathbf{C}_{:,\Gamma(i_l):N_yN_v+\Gamma(i_l)} \right)^{\circ 2} + \tilde{\mu}\mathbf{I} \right]^{\circ -1}, \tag{168}$$

where $\mathcal{I}(\cdot) : \mathbb{R}^{N_xN_v\times N_yN_u} \rightarrow \mathbb{R}^{N_xN_v\times N_yN_u}$, $(\cdot)^{\circ 2}$ and $(\cdot)^{\circ -1}$ represent the Hadamard quadratic power and the Hadamard inverse operation, respectively.

Following the definition of these three SCI operators, the next step is to model the SWM matrix approach in Eq. (163). Toward this goal, Eq. (163) is split into two main equations $\mathbf{\Phi}^T\mathbf{g} + \tilde{\mu}^{-1}(\mathbf{w}-\mathbf{d})$ and $\tilde{\mu}^{-1}\mathbf{I} - \tilde{\mu}^{-1}\mathbf{\Phi}^T\left(\tilde{\mu}\mathbf{I}+\mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}$. In the D-CoSNet, the first equation is reflected as a transpose operator coupled to two 2D convolutional layers (shown as cyan in Fig. 80), each of which are coupled with a ReLU activation function (referred to hereafter as a conventional+ReLU layer). These layers are located at the beginning of Module I, as shown in Fig. 80. Subsequently, the second equation is represented by two parallel arms. The upper arm, corresponding to $\tilde{\mu}^{-1}\mathbf{\Phi}^T\left(\tilde{\mu}\mathbf{I}+\mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1}\mathbf{\Phi}$, is composed of the direct sensing operator as

a first layer followed by an inverse and transpose sensing operator along with four 2D convolutional+ReLU layers, resulting in a block composed of seven layers. The bottom arm, correspond to $\tilde{\mu}^{-1}\mathbf{I}$, has three 2D convolutional+ReLU layers, resulting in a block composed of three layers. The outputs of both arms are added and given as the input to the U-net [referred to as the Module II in Fig. 80] in the D-CoSNet that reflects the Eq. (164). Finally, the output from the U-Net block is the reconstructed datacube (denoted by $\tilde{\mathbf{F}}$). The loss function $\mathcal{L}(\cdot)$ used to learn the weights of the proposed end-to-end network is established as

$$\mathcal{L}(\mathbf{F},\mathbf{G}) = \ell_1(\mathbf{F},\tilde{\mathbf{F}}) + l_1(\mathbf{G},\tilde{\mathbf{G}}) + l_{SSIM}(\mathbf{F},\tilde{\mathbf{F}}), \tag{169}$$

where $(\tilde{\mathbf{F}},\tilde{\mathbf{G}})$ are the D-CoSNet outputs, l_1 is the l_1-norm, and l_{SSIM} represents the structural similarity index measure metric.

8.3. Simulation results

We used two popular databases for simulations: "Sparse Recovery of Hyperspectral Signal from Natural RGB Images"Arad y Ben-Shahar, “Sparse recovery of hyperspectral signal from natural RGB images”, and Çhallenge on Spectral Reconstruction from RGB Images"Arad y col., “Ntire 2020 challenge on spectral reconstruction from an rgb image”. All simulations were implemented in Tensorflow and trained on Google Colaboratory with a Tesla P100-PCI-E GPU using the ADAM optimizer Kingma y Ba, “Adam: A method for stochastic optimization”. To train the coded aperture and the shearing weights, we randomly selected and cropped 450 datacubes with $N_xN_u \times N_yN_v = 256 \times 256$ and $N_\lambda = 25$. The same coded aperture was used to reconstruct the underlying datacube in the three Mu-CoSDI’ depth-alternatives. The depth estimation is carried out via a third-party specialized depth algorithm –LF Jeon y col., “Accurate depth map estimation from a lenslet light field camera”, pure-

Terms under analysis	PSNR (mean $\pm$ standard deviation) [dB]
$l_1(\mathbf{F}, \tilde{\mathbf{F}})$	25.64 $\pm$ 2.83
$l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$	26.16 $\pm$ 2.66
$l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$	26.79 $\pm$ 2.42

Tabla 3. Ablation analysis of various terms in the loss function

CNN Ranftl y col., "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer", and SL Yin y col., "Temporal phase unwrapping using deep learning"– adapted to receive as input the D-CoSNet output.

We conducted comprehensive ablation studies on how the terms $l_1(\mathbf{F}, \tilde{\mathbf{F}})$, $l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$, and $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ in the loss function (i.e., Eq. (169)) would affect the performance of D-CoSNet in reconstruction. The investigation of $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ was inspired by Zhao y col., "Loss functions for image restoration with neural networks". Table 3 shows that implementing the $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$,which integrates the robust training capability of $l_1(\mathbf{F}, \tilde{\mathbf{F}})$ and the ability to preserve the detailed features using $l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$, allows the D-CoSNet to exceed reconstruction performance in comparison to when these two terms are considered individually by 1.15 dB and 0.63 dB, respectively (Table 3). Specifically, this study revealed that $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ shows a better reconstruction performance than when these two terms are considered individually, because $l_1(\mathbf{F}, \tilde{\mathbf{F}}) + l_{SSIM}(\mathbf{F}, \tilde{\mathbf{F}})$ exploits l_1's robust training capability of datasets with outliers (i.e., l_1 does not over-penalize large errors) and the l_{SSIM}'s ability to preserve local structures and contrasts in high-frequency regions. These results are matched with the claims in Zhao y col., "Loss functions for image restoration with neural networks".

Finally, D-CoSNet shows a reconstruction speedup of ~ 600x against the proposed PnP-ADMM-based algorithm (see Section 6).

8.4. Proof of concept experiments

To demonstrate the Mu-CoSDI's multimodal capability, we conducted experimental validations using three representative depth modalities: LF, SL, and pure-CNN-based depth, as shown in Fig. 81. Specifically, the Mu-CoSDI testbed is inspired by the CS^{2}LI imager and constructed following a modular approach, i.e., capable of being coupled with either a lenslet array (to perform light field imaging) or a light projector (to conduct structured light-based acquisition.). Note that we replaced the DFD approach with the CNN-based depth approach to demonstrate the Mu-CoSDI's capabilities to be used in outdoor environments for applications that do not require high depth accuracy. The D-CoSNet returns spectral data cubes with their spectral signatures without illumination corrections, i.e., in intensity. Therefore, an additional offline reconstruction step is the spectral reflectance estimation, see Supplemental material A.

The Mu-CoSDI uses –for the LF scenario, as shown in Fig. 81(A)– a lenslets array (46$\times$46 mm, 4$\times$ 3 mm lenslets) coupled to a Navitar lens ($12mm$ FixedFocal Length, MVL12M23 - $12mm$ EFL, $f/1.4$). These two coupled lenses are the objective lens that images the scene into the input plane of a Fizeau-based 4f geometry composed of a 50:50 matched achromatic doublet pair (L1)(Thorlabs MAP105050-A - 1:1), a polarizing beam splitter (BS)(Thorlabs CCM1-PBS251, 30 mm Cage Cube-Mounted), an $\lambda/4$ achromatic waveplate (Edmund #6-558, 25.4 mm. 465-610 nm), a 100:100 matched achromatic doublet pair (L2)(Thorlabs MAP10100100-A - 1:1), and mounted achromatic doublets lens (L3)(Thorlabs AC254-100-A-ML). L1 and L2 are located on the same propagation axis (i.e., the BM's transmission path), and L3 is located on the BM's opposite reflection path–following an L1-L2 propagation direction. L2 propagates the incoming wavefront through a customized double Amici prism (see Appendix 9) to image onto a coded mirror (CM). The CM is constructed using a broadband dielectric mirror (Thorlabs BB1-E02, 1" diameter) attached to a

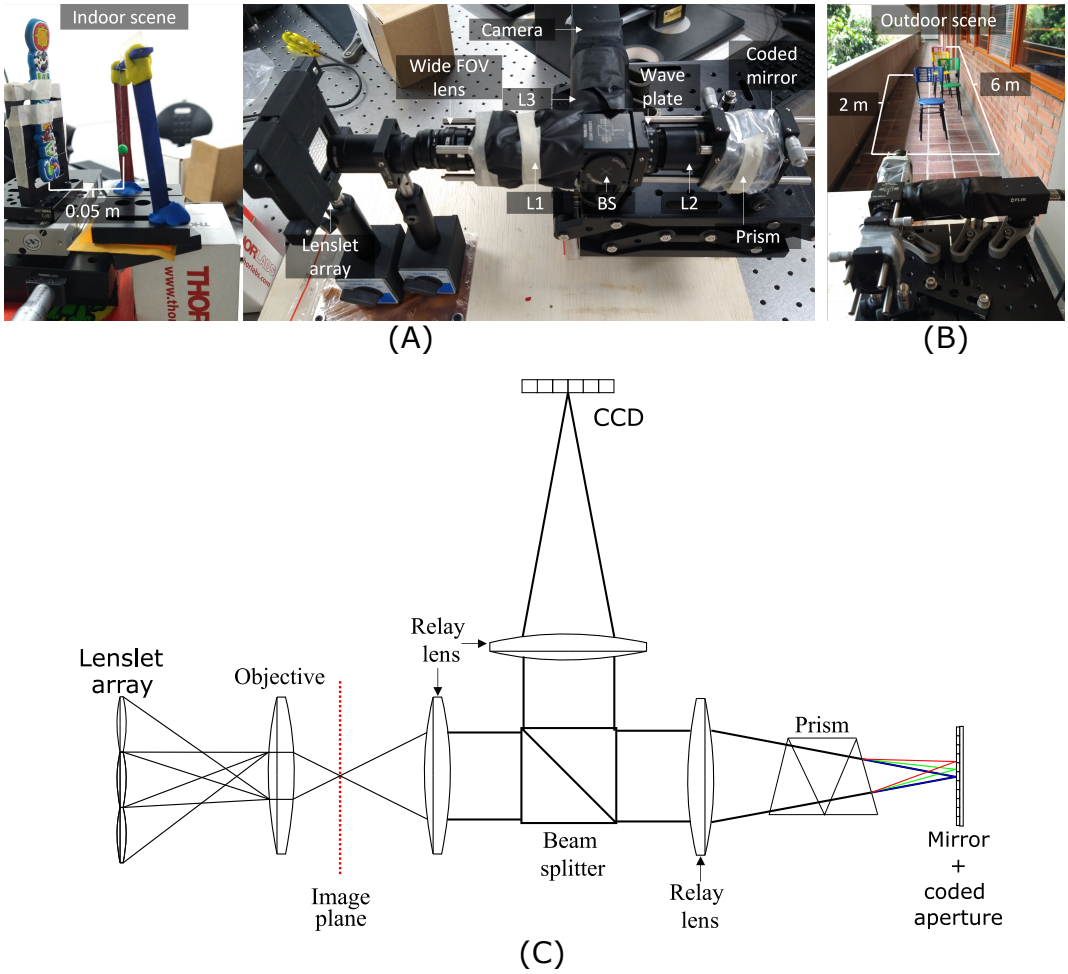


Figura 81. Mu-CoSDI's validation in indoor and outdoor scenarios. Mu-CoSDI configuration for (A) LF modality and (B) CNN-based depth modalities. (C) Geometric sketch of the optical system. Note that the systems in (A) and (B) share the same optical configuration except for the lenslet array.

customized mask (FUJICHROME Velvia 50 transparency film). Taking advantage of the mirror surface, the now dispersed-modulated wavefront is reflected back through the prism until the L2 lens, where the prism undoes the undispersion effect. The resulting dispersed-coded-dispersed wavefront propagates through the BS until the L3, which finally images thewavefront onto the sensor (Stingray F-080B, $4.65\mu m$ pixel size). For the SL and CNN-based depth modalities, the Mu-CoSDI configuration is the same described previously but without the lenslet array, as shown in Fig. 81(B).

The experiments evaluate seven target scenes: three for SL, three for pure-CNN, and one for SL (to complement the results shown in Section 6). In the LF modality, the scenes were named Penguin, Astronaut, and Church. The scenes were named: Penguin, Astronaut, and Church [LF modality]; Students, Creator, and Palms [Pure-CNN modality]; Leaf [SL modality]. The Mu-CoSDI's measurements were acquired using the same customized binary-coded aperture generated from a random distribution with a transmittance level of 50 %. The resulting CA has a pixel size of $61.8\mu m \times 61.8\mu m$, as shown in Fig. 82. The compressed measurement exhibited 256×256 binned CCD pixels ($4.65\mu m \times 4.65\mu m$ pixel size, 12×12 binning) in size. The D-CoSNet recovered a hyperspectral datacube of $N_x \times N_y \times N_L = 256 \times 256 \times 25$ in size. The Mu-CoSDI's LF modality was configured with a field-of-view of $2cm \times 2cm$ for scenes with a depth range of $5cm$ in an indoor environment. Here Penguin and Church targets are composed of two depth-layers and Astronaut is composed of three depth-layers. Insets (A) and (B) in Figs. 83-85 depict the CoSDI's measurements generated with a 5×3 lenslet array and the RGB composite of the reconstruction (obtained via D-CoSNet), respectively. Inset (C) in Fig. 85 shows the depth estimation obtained from the third-party algorithm Jeon y col., "Accurate depth map estimation from a lenslet light field camera" that receives the resulting D-CoSNet reconstruction as input. Note that the depth map estimations are alongside a cropped and zoomed version of Figures 83-85(B), where we can appreciate the parallax effect induced by the lenslet array (highlighted by dotted red circles) of the RGB composite. Finally, inset (D) in Figs. 83 to 85 depicts twelve representatives recovered spectral bands out of the 25. These results show that Mu-CoSDI has a high potential for endoscopy Lin y col., "Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks" applications. The Mu-CoSDI's SL modality was configured with a wide field-of-view of $\sim 10cm \times 10cm$ for scenes with a depth range of $\sim 0.5cm$ in an indoor environment. Here Leaf target

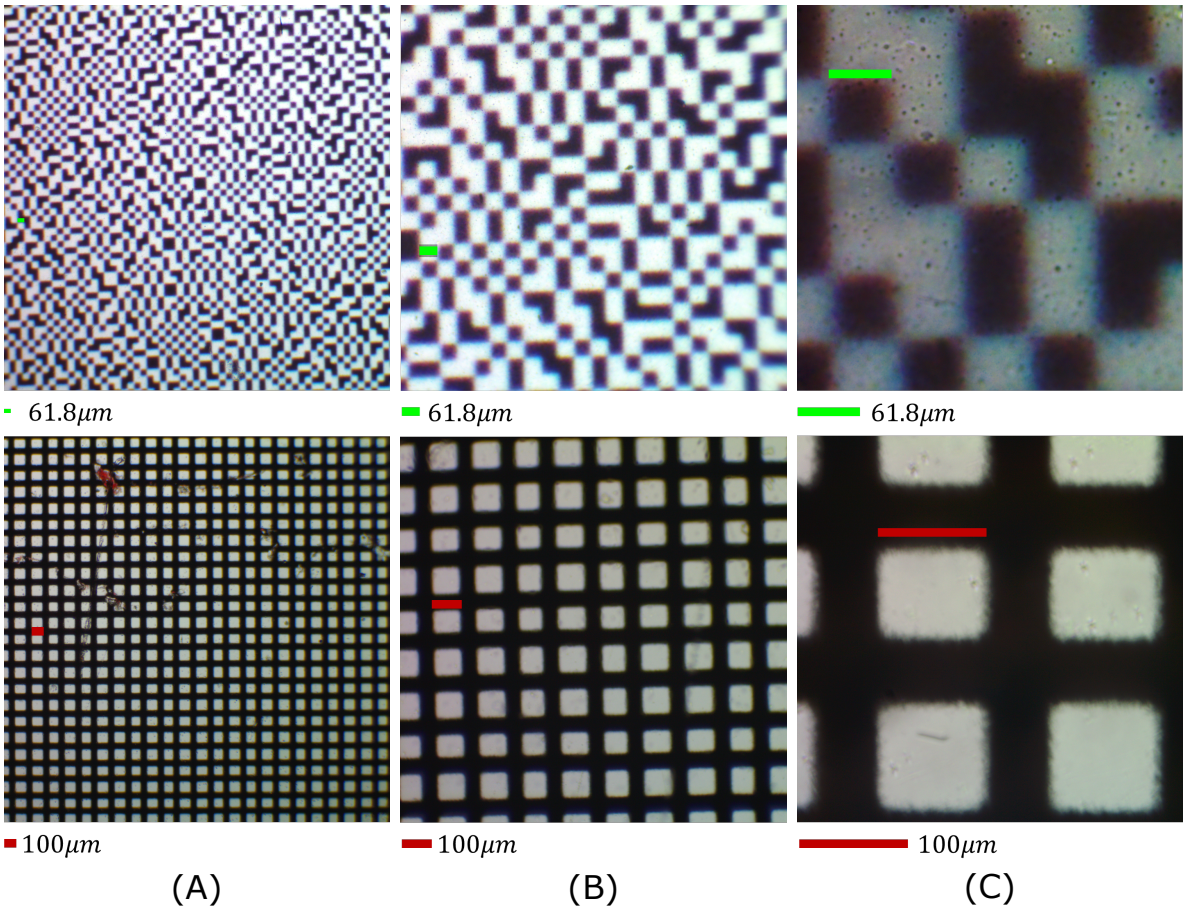


Figura 82. (First row) Customized coded aperture characterization using a (A) 4x, (B) 10x, and (C) 40x microscope objective lens. (Second row) Microscope micrometer calibration grid shaped (DIV=$100\mu m \times 100\mu m$).

is composed of a continuous 3D surface representing the depth map. This scene was included to complement the results shown in Section 6. Figure 86(A) depicts the RGB composite of the reconstruction obtained via D-CoSNet from eight complementary structured light projections: four patterns per frequency. The structured light fringes projected by the video-beam are generated by the equation

$$I^{j}_{i_1,i_2} = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi\nu(i_1 - i_2)}{L} - \frac{2\pi j}{K}\right), \tag{170}$$

where ν is the frequency of fringe patterns, $K \in \mathbb{N}$ is the total of pattern projections per frequency, $L \in \mathbb{N}$ is the total number of pixels in the transverse direction

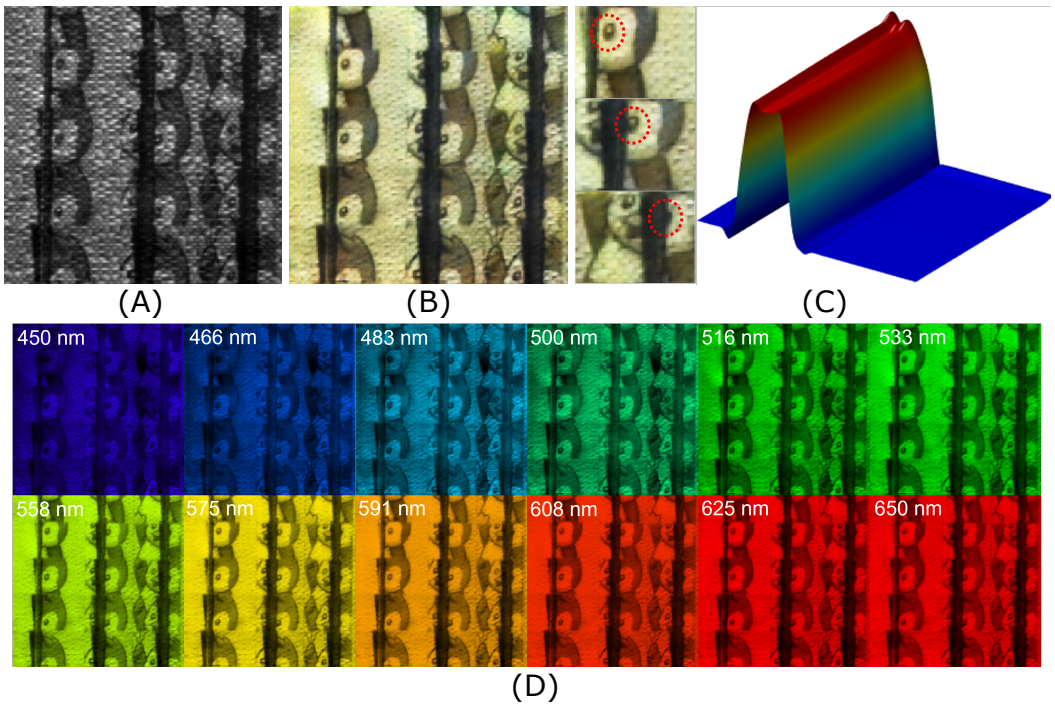


Figura 83. D-CoSNet reconstruction for Penguin scene in the LF modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Parallax illustration along with the resulting depth map estimation via Hae-Gon Jeon y col. “Accurate depth map estimation from a lenslet light field camera”. En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555. (D) Twelve selected wavelength.

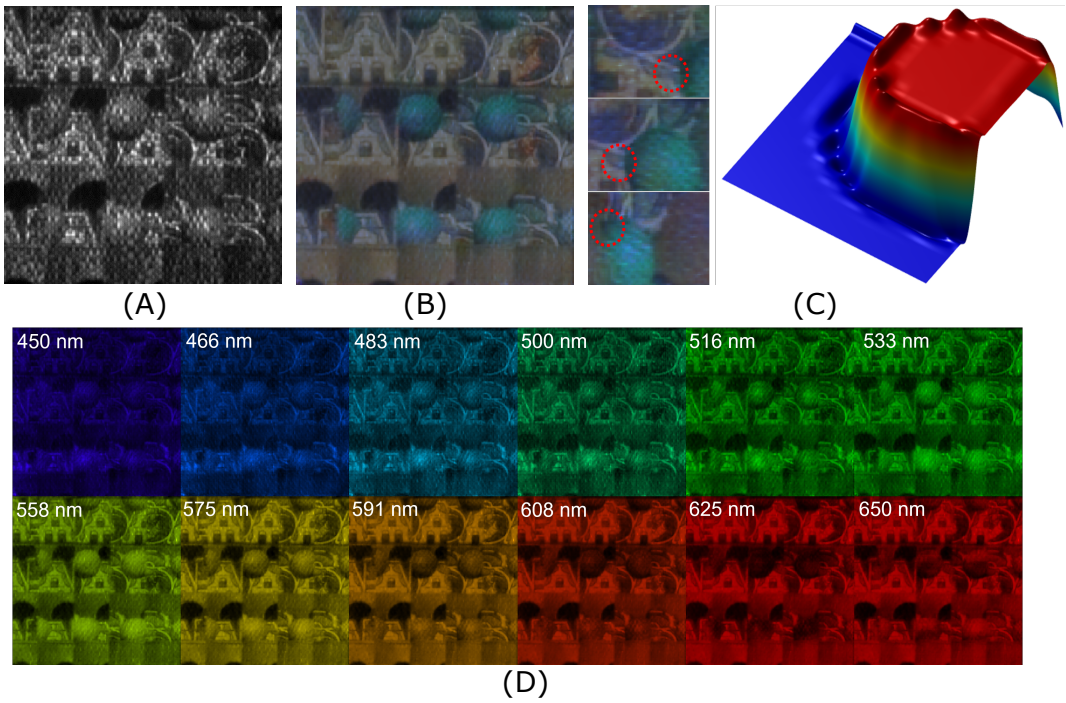


Figura 84. D-CoSNet reconstruction for Church scene in the LF modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Parallax illustration along with the resulting depth map estimation via Hae-Gon Jeon y col. "Accurate depth map estimation from a lenslet light field camera". En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555. (D) Twelve selected wavelength.

of the fringe pattern, and $j = \{0, ..., K-1\}$. In this regard, the parameters selected for the experiments are $K = 4$, $L = 1388$, and $\nu = \{14, 8\}$. Due to the D-CosNet speeding up the reconstruction process, for convenience, we reconstruct a spectral data cube for each measurement. Then, we estimate estimated the spectral data-cube and depth information via Eqs. (119) and (120), respectively. Specifically, Fig. 86(B) shows the depth estimation –in a 3D mapping– obtained from the third-party algorithm Yin y col., "Temporal phase unwrapping using deep learning". Finally, Figs. 86(C) depicts twelve representatives recovered spectral bands. These results echo the claims established in Section 6. Finally, the Mu-CoSDI's pure-CNN modality was configured with a wide field-of-view of $\sim 500cm \times 500cm$ for scenes with a depth range of $\sim 600cm$ in an outdoor environment. Here Students, Creator, and Palms

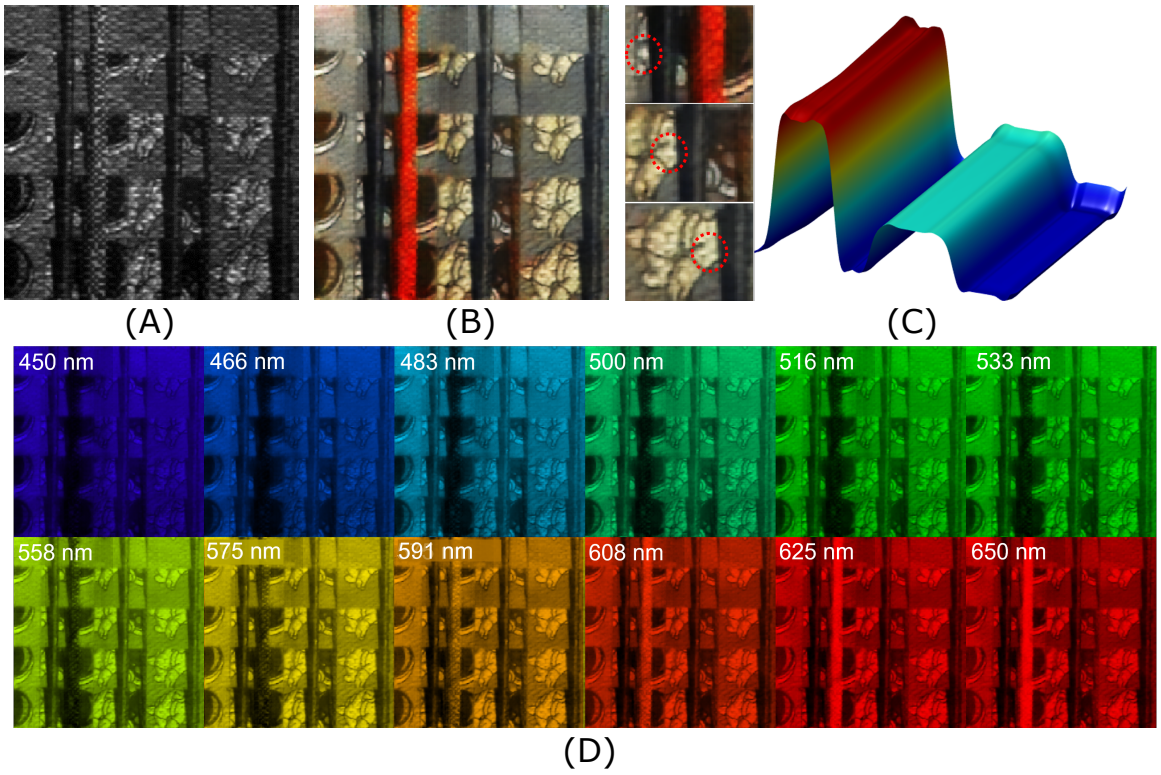


Figura 85. D-CoSNet reconstruction for Astronaut scene in the LF modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Parallax illustration along with the resulting depth map estimation via Hae-Gon Jeon y col. “Accurate depth map estimation from a lenslet light field camera”. En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555. (D) Illustration of twelve selected wavelength.

targets are composed of three depth-layers. Figures 87-89(A) and (B) depict the CoSDI's measurements and the RGB composite of the reconstruction (obtained via D-CoSNet), respectively. Figures 89(C) shows the depth estimation –in a 2D and 3D mapping– obtained from the third-party algorithm Ranftl y col., "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer" that receives the resulting D-CoSNet reconstruction as inputs. These depth-map results support the Mu-CoSDI's claims about the preservation of the spatial features [i.e., depth-dependent geometrical and achromatic aberrations] with good quality. Finally, Figs. 87-89(C) depicts twelve representatives recovered spectral bands. These results show that Mu-CoSDI has a high potential for outdoor applications such as tracking Chen y col., "Remote sensing for vegetation monitoring in carbon capture storage regions: A review", surveillance Yao y col., "Spectral-depth imaging with deep learning based reconstruction", and precision agriculture Shafi y col., "Precision agriculture techniques and practices: From considerations to applications".

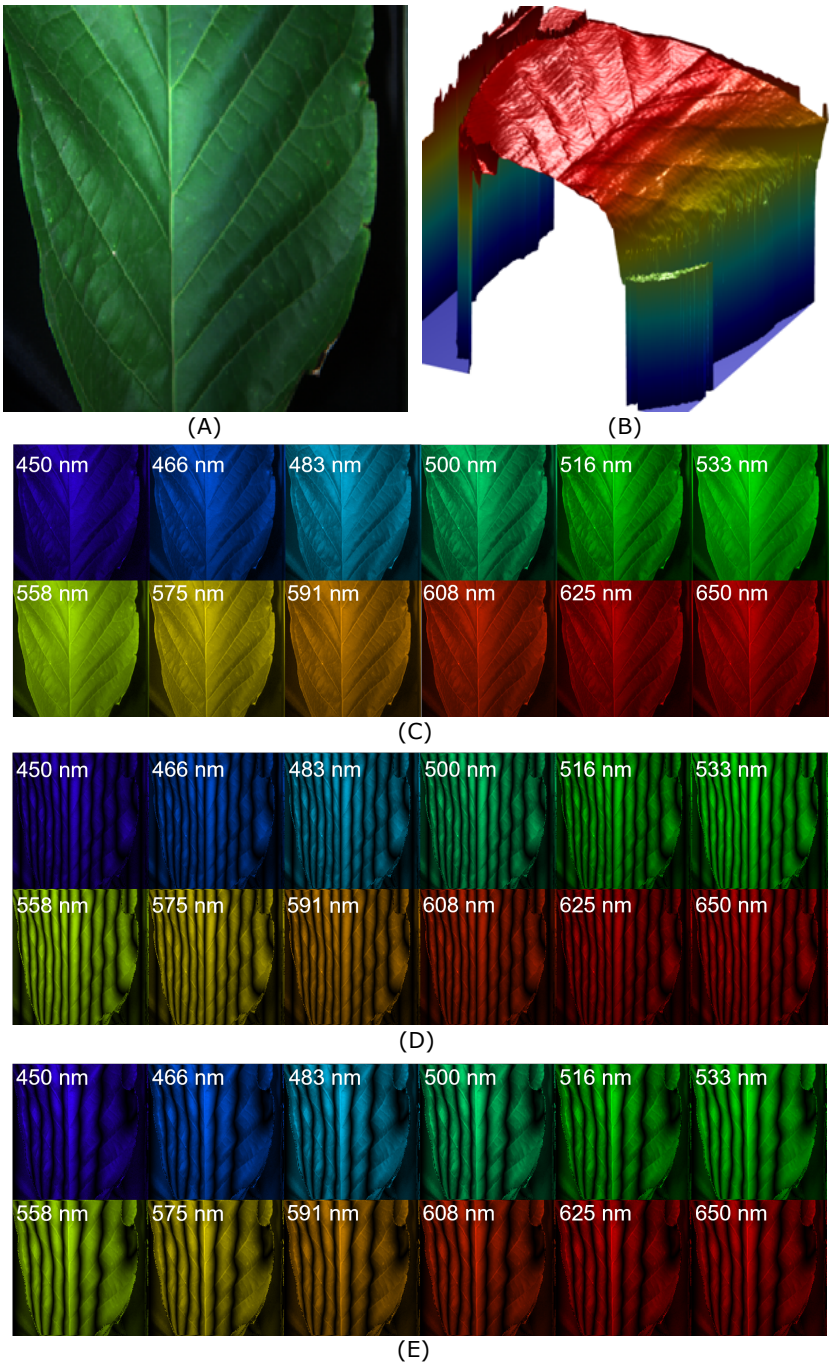


Figura 86. (A) RGB composite image from reconstructed datacube. (B) Depth map estimation from D-CoSNet output. D-CoSNet reconstruction for (C) full datacube (see Eq. (121)). (D)-(E) D-CoSNet reconstruction for the scene illuminated with a high ($\nu = 14$) and low frequency ($\nu = 8$) light patterns (see Eq. (170)), respectively.

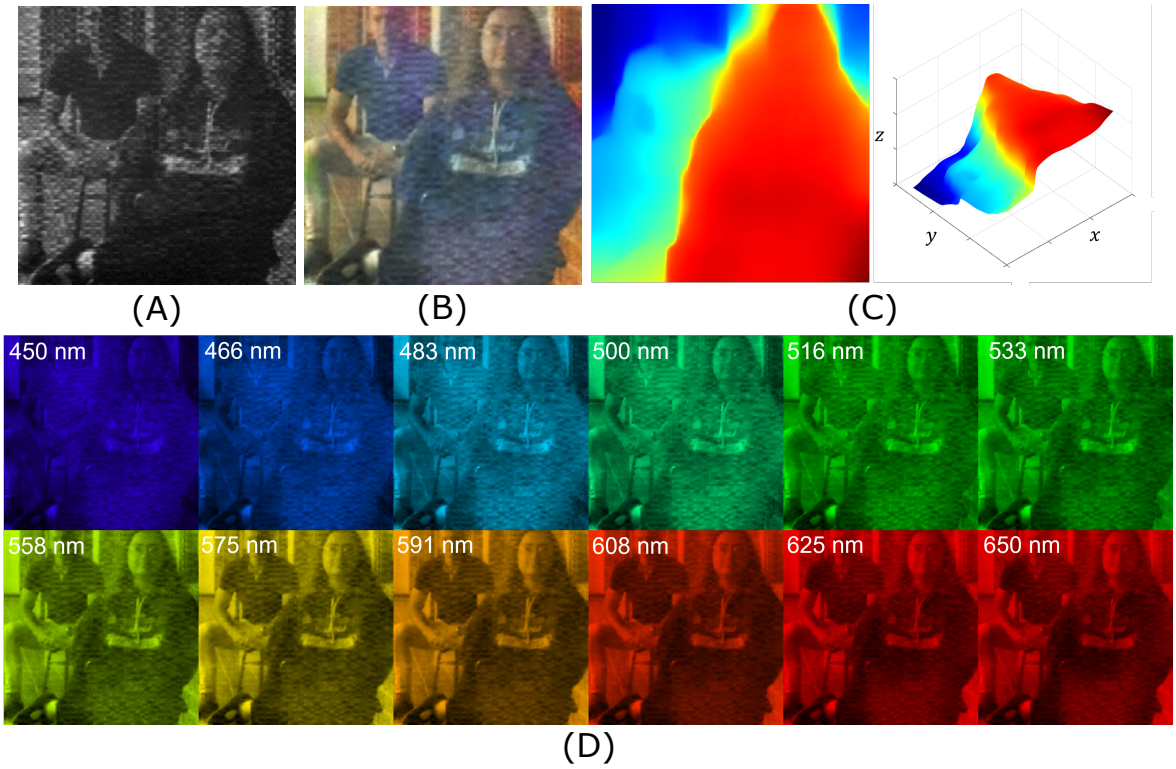


Figura 87. D-CoSNet reconstruction for Students scene in the pure-CNN modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Depth map estimation. (D) Illustration of twelve selected wavelength.

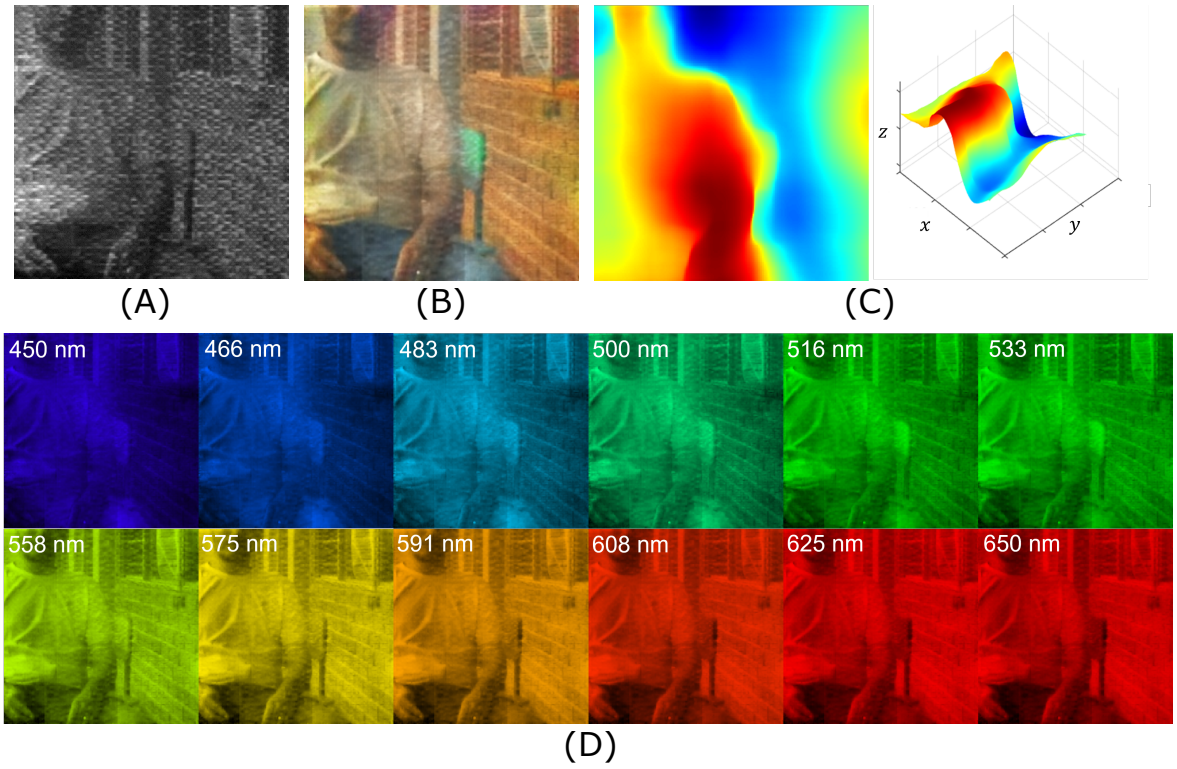


Figura 88. D-CoSNet reconstruction for Creator scene in the pure-CNN modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Depth map estimation. (D) Illustration of twelve selected wavelength.

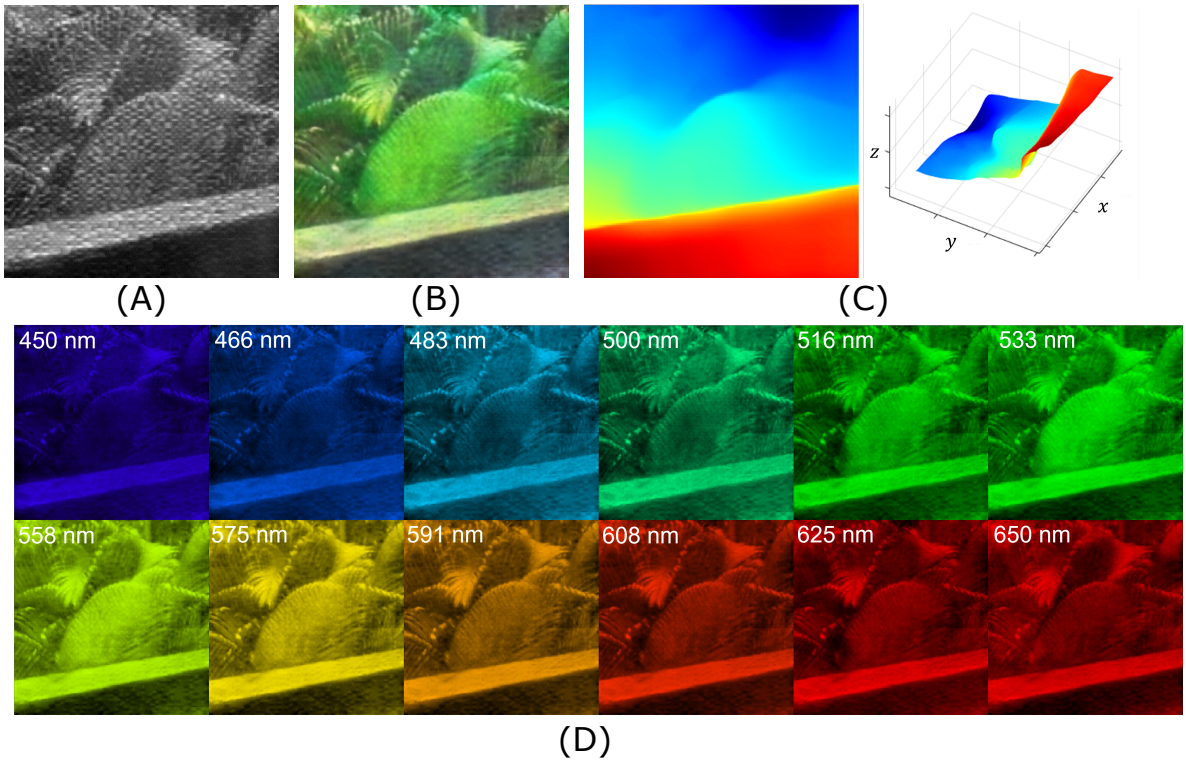


Figura 89. D-CoSNet reconstruction for Palms scene in the pure-CNN modality. (A) Compressed measurement. (B) RGB composite of the reconstruction. (C) Depth map estimation. (D) Illustration of twelve selected wavelength.

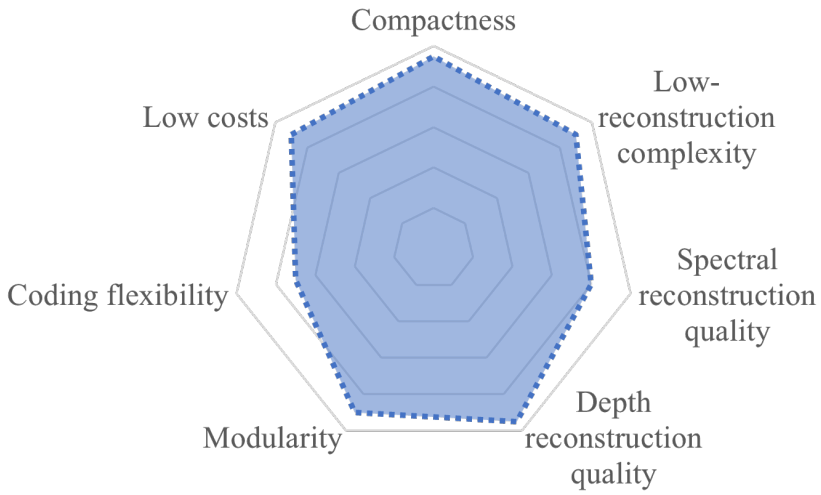


Figura 90. Chart graph for the Mu-CoSDI system

8.5. Conclusions

This chapter extended the CS^{2}LI sensing geometry to propose the Mu-CoSDI system capable of acquiring spectral depth images with a single sensor in three representatives depth modalities: SL, LF, and pure-CNN-based approach. Moreover, we have developed the D-CoSNet that enhances and speeds up the spectral reconstruction. Finally, experiments in indoor and outdoor environments were performed to verify the feasibility of the Mu-CoSDI. In general, the Mu-CoSDI is envisioned as a feasible system to being implemented either on active or passive illumination geometries that simultaneously implement multiples depth modalities, such as SL+LF, LF with a pure CNN-based approach, SL with a pure CNN-based approach, and the three modalities at the same time. D-CoSNet's architecture is highly suitable for solving the inverse problem in Mu-CoSDI. First, the ADMM-based inverse problem is implemented into the first three blocks in the D-CoSNet (see Fig. 80) to promote a higher resonance between the D-CoSNet's weights and the CoSDI's sensing geometry. In

particular, a set of two parallel layers are built to model the direct sensing, transpose process, and inverse operations involved in SCI's inverse problems (see Eq. (166)-(168)). Moreover, the matrix inversion is solved by taking advantage of the SWM formula and the SCI's diagonal matrix property (see Eq. (166)). Finally, the SCI's direct sensing operator is added at the end of D-CoSNet to regularize the estimated datacube. This attractive feature results in a more robust CNN that speeds up the datacube reconstruction.

As a chapter summary and to assess the Mu-CoSDI's pros and cons, we construct a radar chart with seven relevant CoSDI features: compactness, low costs, low-recovery complexity, spectral recovery, depth recovery, modularity, and encoding flexibility. Here the bottom and the top shell represent the worst and best scores, respectively, i.e., the ideal CoSDI system would have to reach the upper contours in the seven features. Specifically, Fig. 90 illustrates the Mu-CoSDI's radar chart, showing that the Mu-CoSDI achieves an optimal balance against the previously proposed CASSDI-DFA, C^{3}LFI, and CS^{2}LI systems.

9. Conclusions and Future Work

The hypothesis of this dissertation was that: ***It is feasible to design a compressive optical-computational imaging system to acquire and reconstruct spatial, spectral, and depth information of a scene using light modulators as coding elements and a single 2D image sensor.***

To support this hypothesis, four CoSDI systems were proposed, built, and validated in three representative depth recovery modalities: depth-from-defocus, light field, and structured light. For the development of these imagers, seven representative features were taken into account: coding flexibility, low costs, compactness, low-reconstruction complexity, spectral reconstruction quality, depth reconstruction quality, and modularity, as shown in Fig. 91. Based on these seven features, the dissertation's main goal was to develop a compact multimodal CoSDI system capable of acquiring spectral-depth information in indoor and outdoor environments for passive and active illumination protocols. In summary, this goal was addressed as follows:

Chapter 4 introduced the CASSDI-DFA, which offers high coding flexibility for either the amplitude and the phase wavefront field at the expense of the modularity and low cots, as shown in Fig. 91 (A). The CASSDI-DFA is based on the depth-from-defocus framework and the PSF engineering to encode the spectral-depth information via dynamic phase and amplitude coding via a deformable mirror and a digital micromirror device. A sequential reconstruction methodology was proposed to enhance the spectral-depth information reconstruction composed of three based-ADMM optimization problems to recover the all-in-focus grayscale image, the grayscale focal stack image, and the spectral image. Moreover, a vanilla U-net was trained for depth recovery from the reconstructed grayscale focal stack image.

Chapter 5 presented the C^3SLFI system which achieves high-quality spectral, and depth information reconstruction quality at the expense of compactness, encoding

flexibility, and implementation costs, as shown in Fig. 91 (B). The C^{3}SLFI system builds on the ideas of complementary CCAs, lenslet array, and tensor models to encode– with two complementary CCA– multiple spectral views of the scene. Moreover, a reconstruction algorithm that exploits the inherent correlations of spectral light-filed images by treating them as tensors, was introduced. This algorithm integrated a 3D parameterization of the 5D spectral light field images to promote higher-spatial correlations, allowing improvements in the spectral image reconstruction quality.

Chapter 6 proposed the CS^{2}LI system, which is low-cost and compact that allows achieving high-quality spectral-depth information recovery at the expense of encoding flexibility and recovery complexity, as shown in Fig. 91 (C). The CS^{2}LI builds on the ideas of color-coded apertures, the dual-dispersive sensing geometry, and the Fizeau interferometer to emulate a double-dispersive CASSI system just using a single double-Amici prism and a coded mirror. Moreover, inspired by the hierarchical resolution approaches and the ADMM paradigm, a low-complexity reconstruction algorithm was integrated to split the spectral-depth reconstruction problem into two low-complex subproblems: spectral datacube reconstruction and grayscale structured light patterns estimation.

Chapter 8 proposed the Mu-CoSDI system that achieves an optimal balance against the previously proposed CASSDI-DFA, C^{3}LFI, and CS^{2}LI systems, as shown in Fig. 91 (D). The Mu-CoSDI sensing geometry extended the CS^{2}LI sensing geometry for acquiring spectral depth images by incorporating either a lenslet array, to conduct light field imaging, or a light projector, to conduct structed light imaging, in a multi-modularity way. The Mu-CoSDI's reconstruction performance was enhanced by developing the D-CoSNet, which enhances and speeds up the spectral reconstruction. The Mu-CoSDI's performance was validated in indoor and outdoor environments.

In perspective, there is still room for improvement when applying optimizations in the spectral and depth sensing protocols using the Mu-CoSDI's geometry as a baseline.

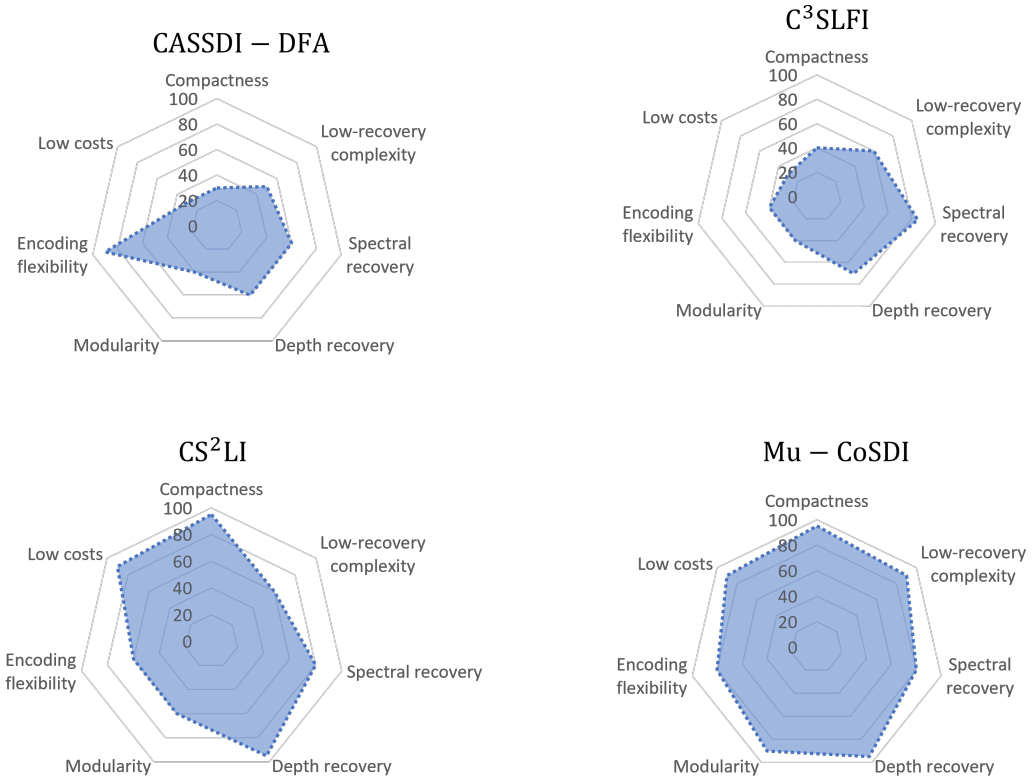


Figura 91. Chart graphs summarization of the four CoSDI systems proposed in this dissertation thesis.

Specifically, future work includes: the SL-based Mu-CoSDI configuration the design and use of a single structured light pattern to enable the acquisition of spectral-depth information in real-time time is still an open problem.

BIBLIOGRAPHY

Adelson, Edward H, James R Bergen y col. *The plenoptic function and the elements of early vision*. Vision y Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991 (vid. págs. 15, 46).

Adelson, Edward H y John Y. A. Wang. "Single lens stereo with a plenoptic camera". En: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 14.2 (1992), págs. 99-106 (vid. págs. 18, 50, 132, 197, 226).

Afonso, Manya V, José M Bioucas-Dias y Mário AT Figueiredo. "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems". En: *IEEE Transactions on Image Processing* 20.3 (2010), págs. 681-695 (vid. pág. 57).

— "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems". En: *IEEE Transactions on Image Processing* 20.3 (2011), págs. 681-695 (vid. pág. 143).

Akgün, Mehmet A, John H Garcelon y Raphael T Haftka. "Fast exact linear and non-linear structural reanalysis and the Sherman–Morrison–Woodbury formulas". En: *International Journal for Numerical Methods in Engineering* 50.7 (2001), págs. 1587-1606 (vid. pág. 293).

Arad, Boaz y Ohad Ben-Shahar. "Sparse Recovery of Hyperspectral Signal from Natural RGB Images". En: *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, págs. 19-34 (vid. pág. 77).

Arad, Boaz y Ohad Ben-Shahar. “Sparse recovery of hyperspectral signal from natural RGB images”. En: *European Conference on Computer Vision*. Springer. 2016, págs. 19-34 (vid. págs. 203, 232).

Arad, Boaz y col. “Ntire 2020 challenge on spectral reconstruction from an rgb image”. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, págs. 446-447 (vid. págs. 204, 232).

Arguello, Henry y Gonzalo R Arce. “Colored coded aperture design by concentration of measure in compressive spectral imaging”. En: *IEEE Transactions on Image Processing* 23.4 (2014), págs. 1896-1908 (vid. págs. 56, 58).

Arguello, Henry y col. “Shift-variant color-coded diffractive spectral imaging system”. En: *Optica* 8.11 (2021), págs. 1424-1434 (vid. pág. 120).

August, Yitzhak y col. “Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains”. En: *Applied Optics* 52.10 (2013), págs. D46-D54 (vid. pág. 56).

Baek, Seung-Hwan y col. “Compact single-shot hyperspectral imaging using a prism”. En: *ACM Transactions on Graphics (TOG)* 36.6 (2017), pág. 217 (vid. págs. 56, 58).

End-to-end hyperspectral-depth imaging with learned diffractive optics. IEEE/CVF International Conference on Computer Vision (ICCV), 2020. DOI: https://doi.org/10.1109/iccv48922.2021.00265 (vid. págs. 19, 98).

Baek, Seung-Hwan y col. “Single-shot Hyperspectral-Depth Imaging with Learned Diffractive Optics”. En: (2021), págs. 2651-2660 (vid. págs. 18, 98).

Bansal, N.P. y R.H. Doremus. *Handbook of Glass Properties*. Academic Press handbook series Handbook of glass properties. Elsevier Science, 2013 (vid. pág. 277).

Batlle, Joan, E Mouaddib y Joaquim Salvi. “Recent progress in coded structured light as a technique to solve the correspondence problem: a survey”. En: *Pattern Recognition* 31.7 (1998), págs. 963-982 (vid. pág. 17).

Bell, Howard E. “Gershgorin’s theorem and the zeros of polynomials”. En: *The American Mathematical Monthly* 72.3 (1965), págs. 292-295 (vid. pág. 78).

Boyd, Stephen y col. “Distributed optimization and statistical learning via the alternating direction method of multipliers”. En: *Foundations and Trends® in Machine learning* 3.1 (2011), págs. 1-122 (vid. págs. 195, 224).

Brady, David J. *Optical imaging and spectroscopy*. John Wiley & Sons, 2009 (vid. pág. 78).

Bruning, J Herriot y col. “Digital wavefront measuring interferometer for testing optical surfaces and lenses”. En: *Applied Optics* 13.11 (1974), págs. 2693-2703 (vid. pág. 42).

Candès, Emmanuel J y Michael B Wakin. “An introduction to compressive sampling”. En: *IEEE Signal Processing Magazine* 25.2 (2008), págs. 21-30 (vid. pág. 53).

Cao, Xun y col. “Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world”. En: *IEEE Signal Processing Magazine* 33.5 (2016), págs. 95-108 (vid. págs. 57, 62, 199, 228).

Castellanos-Gomez, Andres y col. “Spatially resolved optical absorption spectroscopy of single-and few-layer MoS2 by hyperspectral imaging”. En: *Nanotechnology* 27.11 (2016), págs. 115705-115710 (vid. págs. 16, 29).

Chan, Stanley H, Xiran Wang y Omar A Elgendy. “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications”. En: *IEEE Transactions on Computational Imaging* 3.1 (2016), págs. 84-98 (vid. págs. 169, 172).

Chang, Julie y Gordon Wetzstein. “Deep Optics for Monocular Depth Estimation and 3D Object Detection”. En: *arXiv preprint arXiv:1904.08601* (2019) (vid. págs. 16, 34, 36, 61, 105, 120).

Chen, Yun y col. “Remote sensing for vegetation monitoring in carbon capture storage regions: A review”. En: *Applied Energy* 240 (2019), págs. 312-326 (vid. págs. 213, 241).

Cheng, Yeou-Yen y James C Wyant. “Phase shifter calibration in phase-shifting interferometry”. En: *Applied Optics* 24.18 (1985), págs. 3049-3052 (vid. pág. 42).

Christy, Colin D. “Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy”. En: *Computers and electronics in agriculture* 61.1 (2008), págs. 10-19 (vid. pág. 271).

Cichocki, Andrzej y col. “Tensor decompositions for signal processing applications: From two-way to multiway component analysis”. En: *IEEE Signal Processing Magazine* 32.2 (2015), págs. 145-163 (vid. pág. 129).

Cichocki, Andrzej y col. “Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions”. En: *Foundations and Trends® in Machine Learning* 9.4-5 (2016), págs. 249-429 (vid. pág. 135).

Correa, Claudia V, Henry Arguello y Gonzalo R Arce. “Snapshot colored compressive spectral imager”. En: *JOSA A* 32.10 (2015), págs. 1754-1763 (vid. págs. 128, 179).

— “Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging”. En: *JOSA A* 33.12 (2016), págs. 2312-2322 (vid. pág. 140).

Cui, Yan y col. “3D shape scanning with a time-of-flight camera”. En: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. 2010, págs. 1173-1180 (vid. pág. 46).

Hyperspectral+ Depth Imaging Using Compressive Sensing and Structured Light. Optical Society of America. 2018, págs. 1-2 (vid. pág. 189).

Ding, Pengpeng y col. “Single-shot spectral-volumetric compressed ultrafast photography”. En: *Advanced Photonics* 3.4 (2021), págs. 1-6 (vid. pág. 98).

Duarte, Marco F y col. “Single-pixel imaging via compressive sampling”. En: *IEEE Signal Processing Magazine* 25.2 (2008), págs. 83-91 (vid. pág. 55).

Dun, Xiong y col. “Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging”. En: *Optica* 7.8 (2020), págs. 913-922 (vid. pág. 120).

Eckstein, Jonathan y Dimitri P. Bertsekas. “On the Douglasâ€”Rachford splitting method and the proximal point algorithm for maximal monotone operators.” En: *Mathematical Programming* 55.1-3 (1992), págs. 293-318 (vid. págs. 76, 297).

Eckstein, Jonathan y Dimitri P Bertsekas. “On the Douglasâ€”Rachford splitting method and the proximal point algorithm for maximal monotone operators”. En: *Mathematical Programming* 55.1-3 (1992), págs. 293-318 (vid. págs. 200, 229).

Favreau, Peter F y col. “Excitation-scanning hyperspectral imaging microscope”. En: *Journal of Biomedical Optics* 19.4 (2014), págs. 046010-046021 (vid. págs. 16, 29).

Foster, David H y col. “Frequency of metamerism in natural scenes”. En: *JOSA A* 23.10 (2006), págs. 2359-2372 (vid. pág. 76).

Foucart, Simon y Holger Rauhut. “An invitation to compressive sensing”. En: *A mathematical introduction to compressive sensing*. Springer, 2013, págs. 1-39 (vid. pág. 115).

Fu, Chen y col. “Compressive spectral polarization imaging by a pixelized polarizer and colored patterned detector”. En: *JOSA A* 32.11 (2015), págs. 2178-2188 (vid. pág. 179).

Galvis, Laura y col. “Coded aperture design in compressive spectral imaging based on side information”. En: *Applied Optics* 56.22 (2017), págs. 6332-6340 (vid. pág. 62).

Gehm, ME y col. “Single-shot compressive spectral imaging with a dual-disperser architecture”. En: *Optics express* 15.21 (2007), págs. 14013-14027 (vid. págs. 18, 56, 57, 146).

Georgiev, Todor y col. “Lytro camera technology: theory, algorithms, performance analysis”. En: 8667 (2013), 86671J-86681J (vid. págs. 18, 128).

Georgiev, Todor G y col. “Spatio-Angular Resolution Tradeoffs in Integral Photography.” En: *Proceedings of the 17th Eurographics Conference on Rendering Techniques*. Goslar, DEU: Eurographics Association, 2006, 263â€“272. DOI: 10.2312/EGWR/EGSR06/263-272 (vid. pág. 50).

Gilliam, Christopher, Pier-Luigi Dragotti y Mike Brookes. “On the spectrum of the plenoptic function”. En: *IEEE Transactions on Image Processing* 23.2 (2014), págs. 502-516 (vid. pág. 46).

Gómez-Chova, Luis y col. “Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images”. En: *Applied Optics* 47.28 (2008), F46-F60 (vid. págs. 16, 25).

Goodman, J.W. “Introduction to Fourier Optics”. En: McGraw-Hill physical and quantum electronics series (2005), págs. 1-491 (vid. págs. 33, 34).

Gortler, Steven J y col. “The lumigraph”. En: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM. 1996, págs. 43-54 (vid. pág. 46).

Greivenkamp, J.E. *Field Guide to Geometrical Optics*. Field Guides. Society of Photo Optical, 2004 (vid. pág. 274).

Gupta, Rajiv y Richard I Hartley. “Linear pushbroom cameras”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.9 (1997), págs. 963-975 (vid. págs. 15, 25).

Hager, William W. “Updating the inverse of a matrix”. En: *SIAM Review* 31.2 (1989), págs. 221-239 (vid. págs. 201, 230).

Halimi, Abderrahim y col. “Joint reconstruction of multitemporal or multispectral single-photon 3D LiDAR images”. En: *2019 Sensor Signal Processing for Defence Conference (SSPD)*. IEEE. 2019, págs. 1-5 (vid. pág. 15).

Heist, Stefan y col. “5D hyperspectral imaging: fast and accurate measurement of surface shape and spectral characteristics using structured light”. En: *Optics Express* 26.18 (2018), págs. 23366-23379 (vid. págs. 16, 161, 189).

Huntley, Jonathan M y Henrik O Saldner. “Shape measurement by temporal phase unwrapping: comparison of unwrapping algorithms”. En: *Measurement Science and Technology* 8.9 (1997), págs. 986-992 (vid. pág. 42).

Ioannidis, Vassilis N y col. “Coupled graphs and tensor factorization for recommender systems and community detection”. En: *IEEE Transactions on Knowledge and Data Engineering* 33.3 (2019), págs. 909-920 (vid. pág. 135).

Jeon, Daniel S y col. “Compact snapshot hyperspectral imaging with diffracted rotation”. En: *ACM Transactions on Graphics (TOG)* 38.4 (2019), págs. 1-13 (vid. págs. 18, 56, 58, 59, 98).

Jeon, Hae-Gon y col. “Accurate depth map estimation from a lenslet light field camera”. En: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2015, págs. 1547-1555 (vid. págs. 17, 152, 204, 208, 209, 211, 212, 232, 236, 238-240).

Jerez, Andrés, Miguel Márquez y Henry Arguello. “Adaptive coded aperture design for compressive computed tomography”. En: *Journal of Computational and Applied Mathematics* 384 (2020), pág. 113174 (vid. pág. 79).

Jin, Kyong Hwan y col. “Deep convolutional neural network for inverse problems in imaging”. En: *IEEE Transactions on Image Processing* 26.9 (2017), págs. 4509-4522 (vid. págs. 193, 194, 222).

Kamilov, Ulugbek S y Hassan Mansour. “Learning optimal nonlinearities for iterative thresholding algorithms”. En: *IEEE Signal Processing Letters* 23.5 (2016), págs. 747-751 (vid. págs. 193, 194, 222, 223).

Kingma, Diederik P y Jimmy Ba. “Adam: A method for stochastic optimization”. En: *arXiv preprint arXiv:1412.6980* (2014) (vid. págs. 204, 232).

Kolda, Tamara G y Brett W Bader. “Tensor decompositions and applications”. En: *SIAM review* 51.3 (2009), págs. 455-500 (vid. págs. 289, 297).

Krotkov, Eric. “Focusing”. En: *International Journal of Computer Vision* 1.3 (1988), págs. 223-237 (vid. pág. 18).

Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. 2016, págs. 449-458 (vid. págs. 193, 222).

Lai, Yingming y col. “Compressed ultrafast tomographic imaging by passive spatiotemporal projections”. En: *Optics Letters* 46.7 (2021), págs. 1788-1791 (vid. págs. 18, 98).

León-López, Kareth M, Laura V Galvis Carreno y Henry Arguello Fuentes. “Temporal colored coded aperture design in compressive spectral video sensing”. En: *IEEE Transactions on Image Processing* 28.1 (2018), págs. 253-264 (vid. págs. 137, 179).

Levin, Anat y col. “Image and depth from a conventional camera with a coded aperture”. En: *ACM Transactions On Graphics (TOG)* 26.3 (2007), págs. 1-9 (vid. pág. 59).

Levoy, Marc. “Light fields and computational imaging”. En: *Computer* 39.8 (2006), págs. 46-55 (vid. págs. 15, 46).

Levoy, Marc y Pat Hanrahan. “Light field rendering”. En: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, págs. 31-42 (vid. pág. 47).

Li, Chunyu y col. “Pro-cam ssfm: Projector-camera system for structure and spectral reflectance from motion”. En: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, págs. 2414-2423 (vid. pág. 161).

Lin, Jianyu y col. “Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks”. En: *Medical Image Analysis* 48 (2018), págs. 162-176 (vid. págs. 208, 236).

Lin, Xing y col. “Coded focal stack photography”. En: *IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2013, págs. 1-9 (vid. pág. 99).

Lin, Xing y col. “Dual-coded compressive hyperspectral imaging”. En: *Optics Letters* 39.7 (2014), págs. 2044-2047 (vid. págs. 99, 146).

Lin, Xing y col. “Spatial-spectral encoded compressive hyperspectral imaging”. En: *ACM Transactions on Graphics (TOG)* 33.6 (2014), pág. 233 (vid. págs. 146, 148).

Liu, Xianglei y col. “Fast wide-field upconversion luminescence lifetime thermometry enabled by single-shot compressed ultrahigh-speed imaging”. En: *Nature Communications* 12.1 (2021), págs. 1-9 (vid. pág. 98).

Llull, Patrick y col. “Image translation for single-shot focal tomography”. En: *Optica* 2.9 (2015), págs. 822-825 (vid. pág. 61).

Lumsdaine, Andrew y Todor Georgiev. “The focused plenoptic camera”. En: *2009 IEEE International Conference on Computational Photography (ICCP)*. 2009, págs. 1-8. DOI: 10.1109/ICCPHOT.2009.5559008 (vid. pág. 50).

Marquez, Miguel y Henry Arguello. “Coded aperture optimization for single pixel compressive computed tomography”. En: *Journal of Computational and Applied Mathematics* 348 (2019), págs. 58-69 (vid. pág. 79).

Marquez, Miguel, Hoover Rueda-Chacon y Henry Arguello. “Compressive Spectral Light Field Image Reconstruction via Online Tensor Representation”. En: *IEEE Transactions on Image Processing* 29 (2020), págs. 3558-3568 (vid. pág. 77).

Marquez, Miguel y col. “Compressive spectral imaging via deformable mirror and colored-mosaic detector”. En: *Optics Express* 27.13 (2019), págs. 17795-17808 (vid. págs. 56, 59, 100).

Martínez-Corral, Manuel y col. “Recent advances in the capture and display of macroscopic and microscopic 3-D scenes by integral imaging”. En: *Proceedings of the IEEE* 105.5 (2017), págs. 825-836 (vid. pág. 137).

Marwah, Kshitij y col. “Compressive light field photography using overcomplete dictionaries and optimized projections”. En: *ACM Transactions on Graphics (TOG)* 32.4 (2013), pág. 46 (vid. págs. 47, 151).

McMillan, Leonard y Gary Bishop. “Plenoptic modeling: An image-based rendering system”. En: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*. ACM. 1995, págs. 39-46 (vid. pág. 46).

Mejia, Yuri y Henry Arguello. “Binary codification design for compressive imaging by uniform sensing”. En: *IEEE Transactions on Image Processing* 27.12 (2018), págs. 5775-5786 (vid. págs. 79, 139).

Meng, Ziyi, Shirin Jalali y Xin Yuan. “Gap-net for snapshot compressive imaging”. En: *arXiv preprint arXiv:2012.08364* (2020) (vid. págs. 193, 222).

Meng, Ziyi y col. “Self-supervised neural networks for spectral snapshot compressive imaging”. En: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, págs. 2622-2631 (vid. págs. 191, 193, 220, 222).

Monga, Vishal, Yuelong Li y Yonina C Eldar. “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing”. En: *IEEE Signal Processing Magazine* 38.2 (2021), págs. 18-44 (vid. págs. 194, 223).

Morgan, CJ. “Least-squares estimation in phase-measurement interferometry”. En: *Optics Letters* 7.8 (1982), págs. 368-370 (vid. pág. 42).

Ng, R. y Stanford University. Computer Science Department. *Digital Light Field Photography*. Stanford University, 2006 (vid. págs. 48, 50).

Ng, Ren y col. “Light field photography with a hand-held plenoptic camera”. En: *Computer Science Technical Report CSTR* 2.11 (2005), págs. 1-11 (vid. pág. 46).

Ni, Xueping, Changying Li y Huanyu Jiang. “Development of a 3D Multispectral Imaging System using Structured Light”. En: *2019 ASABE Annual International Meeting*. American Society of Agricultural y Biological Engineers. 2019, pág. 1 (vid. pág. 189).

Noll, Robert J. “Zernike polynomials and atmospheric turbulence”. En: *JOSA A* 66.3 (1976), págs. 207-211 (vid. pág. 103).

Palangi, Hamid, Rabab Ward y Li Deng. “Distributed Compressive Sensing: A Deep Learning Approach”. En: *IEEE Transactions on Signal Processing* 64.17 (2016), págs. 4504-4518. DOI: 10.1109/TSP.2016.2557301 (vid. pág. 74).

Pearlman, Jay S y col. “Hyperion, a space-based imaging spectrometer”. En: *IEEE Transactions on Geoscience and Remote Sensing* 41.6 (2003), págs. 1160-1173 (vid. pág. 55).

Pellat-Finet, Pierre. *Optique de Fourier: Théorie métaxiale et fractionnaire*. Springer Science & Business Media, 2009 (vid. pág. 15).

Pérez-Cabré, Elisabet y María Sagrario Millán. “Liquid crystal spatial light modulator with optimized phase modulation ranges to display multiorder diffractive elements”. En: *Applied Sciences* 9.13 (2019), pág. 2592 (vid. pág. 99).

Porwik, Piotr y Agnieszka Lisowska. “The Haar-wavelet transform in digital image processing: its status and achievements”. En: *Machine Graphics and Vision* 13.1/2 (2004), págs. 79-98 (vid. pág. 139).

Ranftl, RenÃ© y col. “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.3 (2022), págs. 1623-1637. DOI: 10.1109/TPAMI.2020.3019967 (vid. págs. 16, 19, 36, 204, 210, 233, 241).

Rivenson, Yair y Adrian Stern. “Compressed imaging with a separable sensing operator”. En: *IEEE Signal Processing Letters* 16.6 (2009), págs. 449-452 (vid. pág. 136).

Ronneberger, Olaf, Philipp Fischer y Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. En: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, págs. 234-241 (vid. págs. 107, 124, 195, 224).

Rudin, Leonid I, Stanley Osher y Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. En: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), págs. 259-268 (vid. pág. 55).

Rueda, Hoover, Henry Arguello y Gonzalo R Arce. “DMD-based implementation of patterned optical filter arrays for compressive spectral imaging”. En: *JOSA A* 32.1 (2015), págs. 80-89 (vid. págs. 57, 154).

Rueda, Hoover y col. “Single aperture spectral+ ToF compressive camera: toward hyperspectral+ depth imagery”. En: *IEEE Journal of Selected Topics in Signal Processing* 11.7 (2017), págs. 992-1003 (vid. pág. 62).

Servin, M., J.A. Quiroga y M. Padilla. *Fringe Pattern Analysis for Optical Metrology: Theory, Algorithms, and Applications*. Wiley, 2014 (vid. págs. 38-40, 42).

Shafi, Uferah y col. “Precision agriculture techniques and practices: From considerations to applications”. En: *Sensors* 19.17 (2019), pág. 3796 (vid. págs. 213, 241).

Shaw, Gary A y Hsiaohua K Burke. “Spectral imaging for remote sensing”. En: *Lincoln laboratory journal* 14.1 (2003), págs. 3-28 (vid. pág. 46).

Shen, Huanfeng y col. “Adaptive norm selection for regularized image restoration and super-resolution”. En: *IEEE Transactions on Cybernetics* 46.6 (2015), págs. 1388-1399 (vid. pág. 108).

Sheoran, Gyanendra y col. “Swept-source digital holography to reconstruct tomographic images”. En: *Optics Letters* 34.12 (2009), págs. 1879-1881 (vid. págs. 16, 29).

Sidiropoulos, Nicholas D y col. “Tensor decomposition for signal processing and machine learning”. En: *IEEE Transactions on Signal Processing* 65.13 (2017), págs. 3551-3582 (vid. pág. 128).

Silberman, Nathan y col. “Indoor segmentation and support inference from rgbd images”. En: *European Conference on Computer Vision*. Springer. 2012, págs. 746-760 (vid. págs. 113, 175).

Subbarao, Murali y Gopal Surya. “Depth from defocus: A spatial domain approach”. En: *International Journal of Computer Vision* 13.3 (1994), págs. 271-294 (vid. pág. 16).

Sun, Baoqing y col. “3D computational imaging with single-pixel detectors”. En: *Science* 340.6134 (2013), págs. 844-847 (vid. pág. 99).

Sun, Ming-Jie y col. “Single-pixel three-dimensional imaging with time-based depth resolution”. En: *Nature Communications* 7 (2016), págs. 1-6. DOI: https://doi.org/10.1038/ncomms12010 (vid. pág. 60).

Tachella, Julián y col. “Bayesian 3D Reconstruction of Subsampled Multispectral Single-photon Lidar Signals”. En: *arXiv preprint arXiv:1904.02583* (2019) (vid. pág. 15).

Tahara, Tatsuki y col. “Multidimensional digital holographic microscopy based on computational coherent superposition for coherent and incoherent light sensing”.

En: *Holography, Diffractive Optics, and Applications X*. Vol. 11551. International Society for Optics y Photonics. 2020, pág. 115510M (vid. pág. 99).

Tang, Huixuan y col. “Depth from Defocus in the Wild”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 2740-2748 (vid. págs. 16, 34-36).

Torabzadeh, Mohammad y col. “Hyperspectral imaging in the spatial frequency domain with a supercontinuum source”. En: *Journal of Biomedical Optics* 24.7 (2019), págs. 071614-071623 (vid. págs. 16, 29).

Tsai, Chun-Wei y col. “Enhancing performance of LCoS-SLM as adaptive optics by using computer-generated holograms modulation software”. En: *Holography: Advances and Modern Trends V*. Vol. 10233. International Society for Optics y Photonics. 2017, pág. 102331D (vid. pág. 99).

Unger, Jonas y col. “Capturing and rendering with incident light fields”. En: *Proc. Eurograph. Workshop Rendering Techn.* 2003, 141â€“-149 (vid. pág. 50).

Valadao, Gonçalo y José Bioucas-Dias. “PUMA: phase unwrapping via max flows”. En: *Proceedings of Conference on Telecommunicationsâ€”ConfTele*. Citeseer. 2007, págs. 609-612 (vid. págs. 168, 175).

Wagadarikar, Ashwin y col. “Single disperser design for coded aperture snapshot spectral imaging”. En: *Applied optics* 47.10 (2008), B44-B51 (vid. págs. 18, 56-58, 62, 85, 104, 105, 128).

Wagadarikar, Ashwin A y col. “Video rate spectral imaging using a coded aperture snapshot spectral imager”. En: *Optics Express* 17.8 (2009), págs. 6368-6388 (vid. pág. 277).

Wang, Lizhi y col. “Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging”. En: *IEEE transactions on pattern analysis and machine intelligence* 39.10 (2016), págs. 2104-2111 (vid. págs. 78, 85).

Wang, Lizhi y col. “Hyperspectral image reconstruction using a deep spatial-spectral prior”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, págs. 8032-8041 (vid. pág. 86).

Wilburn, Bennett y col. “High performance imaging using large camera arrays”. En: 24.3 (2005), págs. 765-776 (vid. pág. 50).

Wu, Gaochang y col. “Light field image processing: An overview”. En: *IEEE Journal of Selected Topics in Signal Processing* 11.7 (2017), págs. 926-954 (vid. pág. 49).

Wu, Zhuoyuan, Jian Zhang y Chong Mou. “Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging”. En: *arXiv preprint arXiv:2109.06548* (2021) (vid. págs. 193, 222).

Xie, Qi y col. “Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.8 (2017), págs. 1888-1902 (vid. pág. 135).

Xiong, Zhiwei y col. “Snapshot hyperspectral light field imaging”. En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 3270-3278 (vid. págs. 62, 128, 147).

Yao, Jiali y col. “Total variation and block-matching 3d filtering-based image reconstruction for single-shot compressed ultrafast photography”. En: *Optics and Lasers in Engineering* 139 (2021), pág. 106475 (vid. pág. 99).

Yao, Mingde y col. “Spectral-depth imaging with deep learning based reconstruction”. En: *Optics Express* 27.26 (2019), págs. 38312-38325 (vid. págs. 213, 241).

Yao, Yunhua y col. “Single-Shot Real-Time Ultrafast Imaging of Femtosecond Laser Fabrication”. En: *ACS Photonics* 8.3 (2021), págs. 738-744 (vid. pág. 98).

Yasuma, Fumihito y col. “Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum”. En: *IEEE Transactions on Image Processing* 19.9 (2010), págs. 2241-2253 (vid. pág. 76).

Yin, Wei y col. “Temporal phase unwrapping using deep learning”. En: *Scientific reports* 9.1 (2019), págs. 1-12 (vid. págs. 43, 175, 204, 210, 233, 239).

Yuan, Xin y col. “Compressive hyperspectral imaging with side information”. En: *IEEE Journal of Selected Topics in Signal Processing* 9.6 (2015), págs. 964-976 (vid. pág. 62).

Yuan, Xin y col. “Efficient patch-based approach for compressive depth imaging”. En: *Applied Optics* 55.27 (2016), págs. 7556-7564 (vid. pág. 61).

Yuan, Xin y col. “Plug-and-play algorithms for large-scale snapshot compressive imaging”. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, págs. 1447-1457 (vid. págs. 201, 230).

Zhang, Yongbing y col. “Light-field depth estimation via epipolar plane image analysis and locally linear embedding”. En: *IEEE Transactions on Circuits and Systems for Video Technology* 27.4 (2016), págs. 739-747 (vid. págs. 47, 48).

Zhao, Hang y col. “Loss functions for image restoration with neural networks”. En: *IEEE Transactions on Computational Imaging* 3.1 (2016), págs. 47-57 (vid. págs. 204, 205, 233).

Zhao, Huijie y col. “A high throughput integrated hyperspectral imaging and 3D measurement system”. En: *Sensors* 18.4 (2018), pág. 1068 (vid. pág. 161).

Zhao, Yang y col. “Heterogeneous camera array for multispectral light field imaging”. En: *Optics express* 25.13 (2017), págs. 14008-14022 (vid. págs. 17, 48, 128).

Zhao, Yuzhi y col. “Hierarchical Regression Network for Spectral Reconstruction from RGB Images”. En: (2020), págs. 422-423 (vid. págs. 113, 175).

Zheng, Siming y col. “Deep plug-and-play priors for spectral snapshot compressive imaging”. En: *Photonics Research* 9.2 (2021), B18-B29 (vid. págs. 194, 223).

Zhu, Kang y col. “Hyperspectral Light Field Stereo Matching”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018) (vid. págs. 48, 128).

Zhu, Shuaishuai y col. “On the fundamental comparison between unfocused and focused light field cameras”. En: *Applied Optics* 57.1 (2018), A1-A11 (vid. pág. 51).

Zhu, WenLiang y col. “Architectural design of deep metallic sub-wavelength grating for practical holography display”. En: *AOPC 2017: Optoelectronics and Micro/Nano-Optics*. Vol. 10460. International Society for Optics y Photonics. 2017, 104601E (vid. pág. 99).

Appendix A. Light Correction

The response of the optical and electrical components that comprise the detection system is wavelength-dependent, and the optical system will detect different wavelengths of light with differing efficiency. The two primary contributions to this wavelength dependence are the illumination source and the detector quantum efficiency. In this regard, the light correction protocol used for this thesis is divided into seven main steps.

1. The characterization of the coded aperture using a monochromatic tunable lighting source. In this step, we seek to obtain the spectral response of the optical elements.

2. To construct the sensing matrix from the characterized coded apertures. Here, the sensing matrix used in the experiments incorporates the spectral response of the sensor and the imperfections between the aperture code and the camera.

3. To acquire a white reference measurement using as a target a diffuse reflectance standard. Then, to obtain a reconstruction of the white reference target from the acquired compressive measurement and the generated sensing matrix in step 2.

4. To acquire the electronic sensor noise by turning off the light source and preserving all camera parameters, e.g., the gain, integration time, and shutter. Then, the optical noise is suppresed.

5. To acquire the scenes using the white light source. Then, proceed to reconstruct the scene using the sampling matrix generated from the characterized codes (see step 2).

6. To calculate the reflectance via the following equation Colin D Christy. “Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy”. En: *Computers and electronics in agriculture* 61.1 (2008), págs. 10-19:

$$R(\lambda) = \frac{\hat{X}(\lambda) - \epsilon(\lambda)}{\tilde{W}(\lambda) - \epsilon(\lambda)} \tag{171}$$

where $\hat{X}$ represents the reconstructed scene, $\hat{W}$ represents the reconstructed white target, and ϵ is the dark count.

Notice that in a non-compressive system, the light correction starts in step three, and the reconstruction process is omitted.

Appendix B. Double Amici-prism design

One of the essential elements in the spectral sensing paradigm is the dispersive prism. This element separates the different wavelengths of an incident beam of light into different optical paths, where its dispersion curve depends on its geometry and substrate. Prisms are preferred over their grating counterparts since they possess better throughput, better light characteristics, and are more customizable. One of the essential elements in the spectral sensing paradigm is the dispersive prism. This element separates the different wavelengths of an incident beam of light into different optical paths, where its dispersion curve depends on its geometry and substrate. Prisms are preferred over their grating counterparts since they possess better throughput, better light characteristics, and are more customizable. For these reasons, in this dissertation, we concentrate on designing compressive spectral imagers based on dispersive prisms. Therefore, to deepen and better understand the spectral imagers design introduced in the first chapter and customize the dispersive element that aims to be used in the design of all the proposed optical systems in the present dissertation, we will study the design of a double Amici prism based on the first-order design principles.

Double Amici prism geometry

First prism geometry Let θ_0 the angle between an arbitrary incoming ray converging to the first medium and the optical axis, and β_1 the resulting left angle of split the apex angle α_1 with a vertical line, as is illustrated in Fig. 93. Then, the incident angle θ_1 can be calculated as

$$\theta_1 = \theta_0 - \beta_1, \tag{172}$$

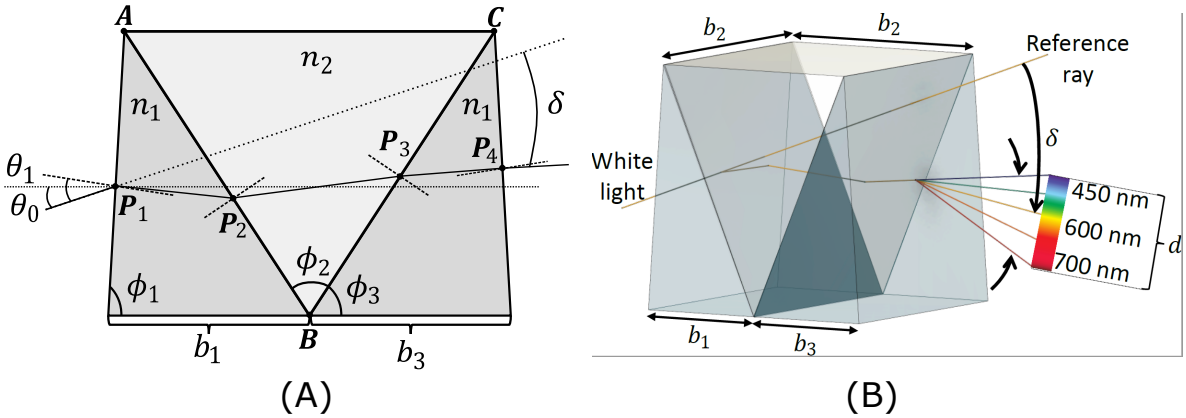


Figura 92. Double Amici prism geometry sketch. (A) Illustration for the central wavelength ray dispersion. (B) Wavelength-dependent ray dispersion.

which is derivated from the angle sum relationship

$$\theta_0 + \left(\frac{\pi}{2} - \theta_1\right) + (-\beta_1) = \frac{\pi}{2}, \tag{173}$$

where θ_0 and θ_1 are positive angles and β_1 is a negative angle, following the modern sign notation as is illustrated in Fig. 93. Here, it is worth noting that $\beta_1 = -\alpha_1 + \frac{1}{2}\phi_2$. After entering to the first prism, the reflected angle θ_1' is determined by Snellâ€™s Law as

$$\theta_1' = \arcsin\left(\frac{1}{n_1}\sin\theta_1\right), \tag{174}$$

where n_1 is the prism first-medium refractive index and θ_1' is a positive angle. Having estimated the reflected angle θ_1', the second incident angle between the second surface and the propagated ray can be calculated by

$$\theta_2 = \theta_1' - \alpha_1, \tag{175}$$

which is derived from the internal angle sum relationship of $\angle P_1AP_2$

$$\left[\frac{\pi}{2} - (-\theta_2)\right] + \left[\frac{\pi}{2} - (\theta_1')\right] + [\gamma_1 + (-\beta_1)] = \pi \tag{176}$$

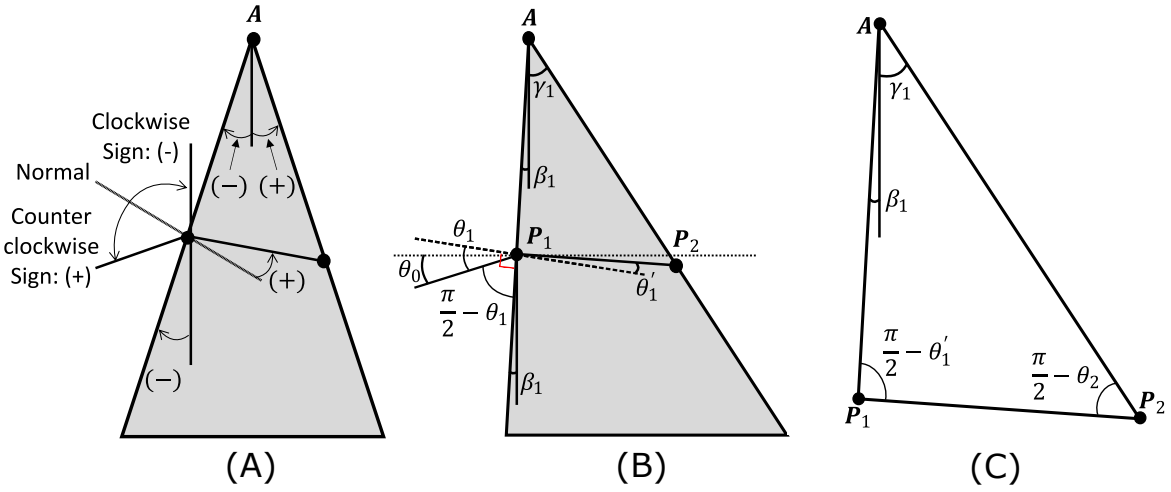


Figura 93. First-prism sketch. (A) Sign notation where θ_0, θ_1, and γ_1 are positive angles, and β_1 is negative. (B) First-prism and ray propagation geometry. (C) Auxiliary triangle ($\angle P_1 A P_2$) geometry used to calculate the second incident angle θ_2. where θ_2 is a negative angle and $\alpha_1 = \gamma_1 - \beta_1$ J.E. Greivenkamp. *Field Guide to Geometrical Optics*. Field Guides. Society of Photo Optical, 2004, as is illustrated in Fig. 93.

Second prism geometry After entering to the second prism, similarly to Eq. (174), the reflected angle θ_2' is determined by Snell's Law as

$$\theta_2' = \arcsin\left(\frac{n_1}{n_2}\sin\theta_2\right), \tag{177}$$

where n_2 is the prism second-medium refractive index and θ_2' is a negative angle, as is shown in Fig. 94 (left image). Then, the third incident angle between the third surface and the propagated ray can be obtained by

$$\theta_3 = \theta_2' - \phi_2, \tag{178}$$

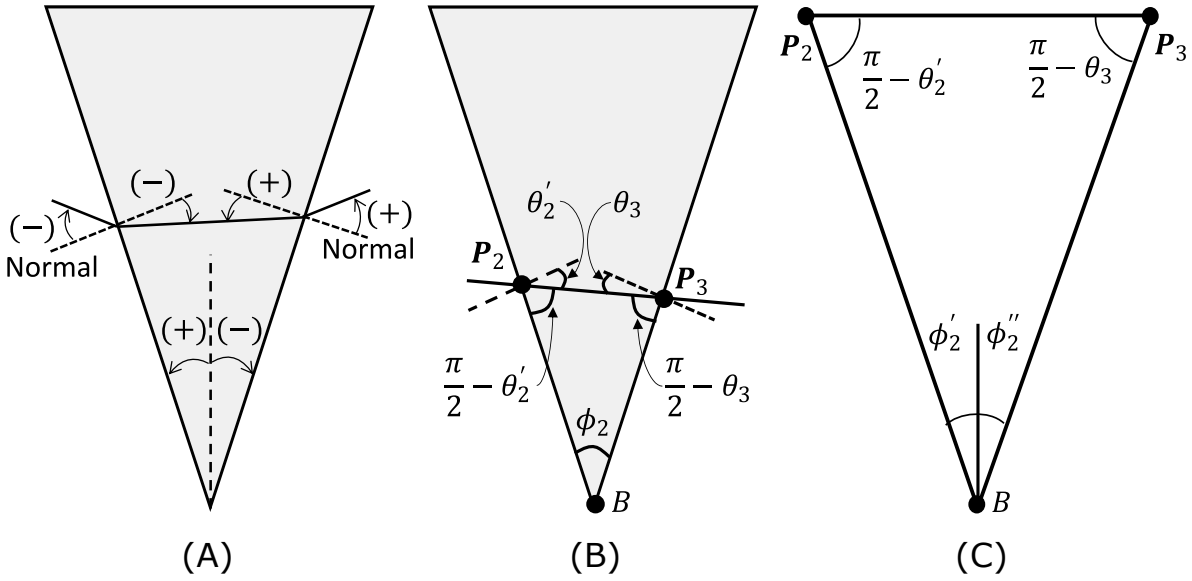


Figura 94. Second-prism sketch. (A) Sign notation where θ_3, and θ_2' is positive angles, and β_1 is negative. (B) Second-prism and ray propagation geometry. (C) Auxiliary triangle ($\angle P_2BP_3$) geometry used to calculate the second incident angle θ_3.

which is derived from the internal angle sum relationship of $\angle P_2BP_3$ (Fig. 94)

$$\left[\frac{\pi}{2}-\theta_3\right]+\left[\frac{\pi}{2}-(-\theta_2')\right]+[(-\phi_2'')+(\phi_2')]=\pi, \tag{179}$$

where θ_3 and ϕ_2' are positive angles, ϕ_2'' is a negative angle, and the second apex angle is equal to $\phi_2=\phi_2''-\phi_2'$, as is illustrated in Fig. 94.

Third prism geometry After entering to the third prism, similarly to Eq. (174), the reflected angle θ_2' is determined by Snell's Law as

$$\theta_3'=arcsin\left(\frac{n_2}{n_1}\sin\theta_3\right), \tag{180}$$

where θ_3' is a positive angle,as is shown in Fig. 95. Then, the fourth incident angle between the fourth surface and the propagated ray can be obtained by

$$\theta_4=\theta_3'-\alpha_1, \tag{181}$$

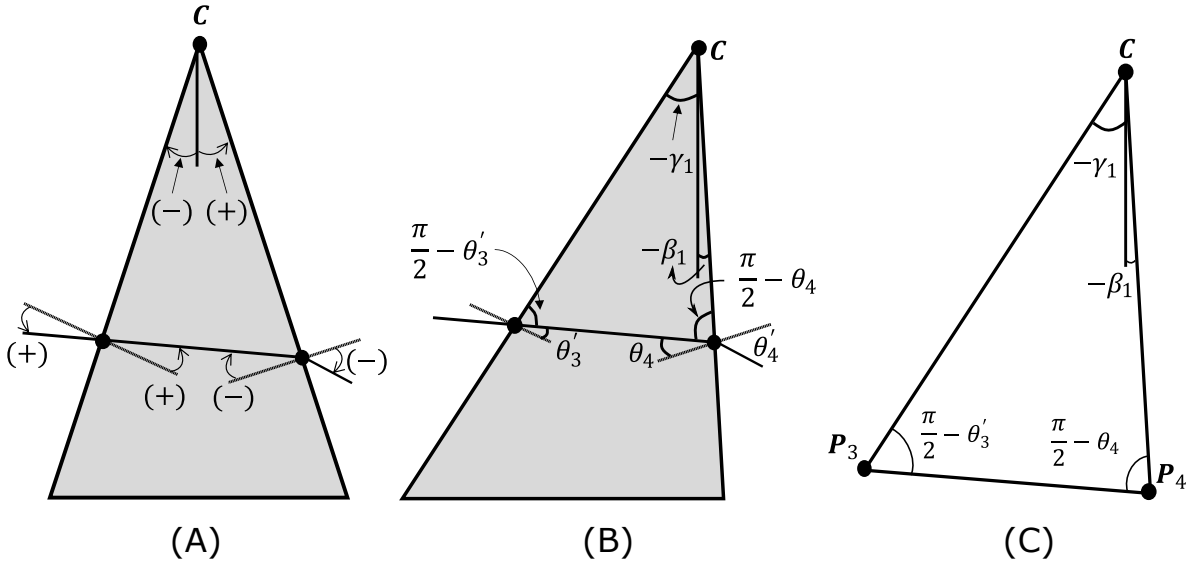


Figura 95. Second-prism sketch. (A) Sign notation where θ_3', and θ_4' is positive angles, and β_1 is negative. (B) Third-prism and ray propagation geometry. (C) Auxiliary triangle ($\angle P_3CP_4$) geometry used to calculate the second incident angle θ_4.

which is derived from the internal angle sum relationship of $\angle P_3CP_4$

$$\left[\frac{\pi}{2}-(-\theta_4)\right]+\left[\frac{\pi}{2}-(\theta_3')\right]+[-(-\beta_1)-(\gamma_1)]=\pi, \tag{182}$$

where θ_4 is a negative angle and $\beta_1-\gamma_1=-\alpha_1$. Similarly to Eqs. (177) and (180), the reflected angle of the outcoming ray can be determined by Snell's Law as

$$\theta_4'=\arcsin n_1\sin\theta_4, \tag{183}$$

where θ_4' is a negative angle. Then, the angle between the optical axis and the outcoming ray is defined as

$$\theta_5=\theta_4'-\beta_1, \tag{184}$$

which is derivated from the angle sum relationship

$$-(-\beta_1) + \left[\frac{\pi}{2} - (-\theta_4')\right] + (-\theta_5) = \frac{\pi}{2}, \tag{185}$$

where θ_5 is a negative angle. Finally, the prism deviation angle δ is defined as

$$\delta = \theta_0 + (-\theta_5)\,. \tag{186}$$

Since the prism refractive indices vary with wavelength, the deviation angle also does. The difference in deviation angle between the two extreme wavelengths transmitted by the prism can be estimated by

$$\Delta = \delta(\lambda_{max}) - \delta(\lambda_{min}) \tag{187}$$

where Δ represents the prism dispersion. Therefore, the following subsection will be studying the prism behavior for two particularly glass materials with wavelength-dependent refractive index.

Wavelength-dependent refractive index Inspired on Ashwin A Wagadarikar y col. "Video rate spectral imaging using a coded aperture snapshot spectral imager". En: *Optics Express* 17.8 (2009), págs. 6368-6388, the glass materials selected for the prism design are N-SK2 and SF4 N.P. Bansal y R.H. Doremus. *Handbook of Glass Properties*. Academic Press handbook series Handbook of glass properties. Elsevier Science, 2013. The refractive index n_1 and n_2 are mathematically modeled as

$$n_1(\lambda) = \left[1 + \frac{1.28189012\lambda.^2}{\lambda^2 - 0.0072719164} + \frac{0.257738258\lambda^2}{\lambda^2 - 0.0242823527} + \frac{0.96818604\lambda^2}{\lambda^2 - 110.377773}\right]^{\frac{1}{2}}, \tag{188}$$

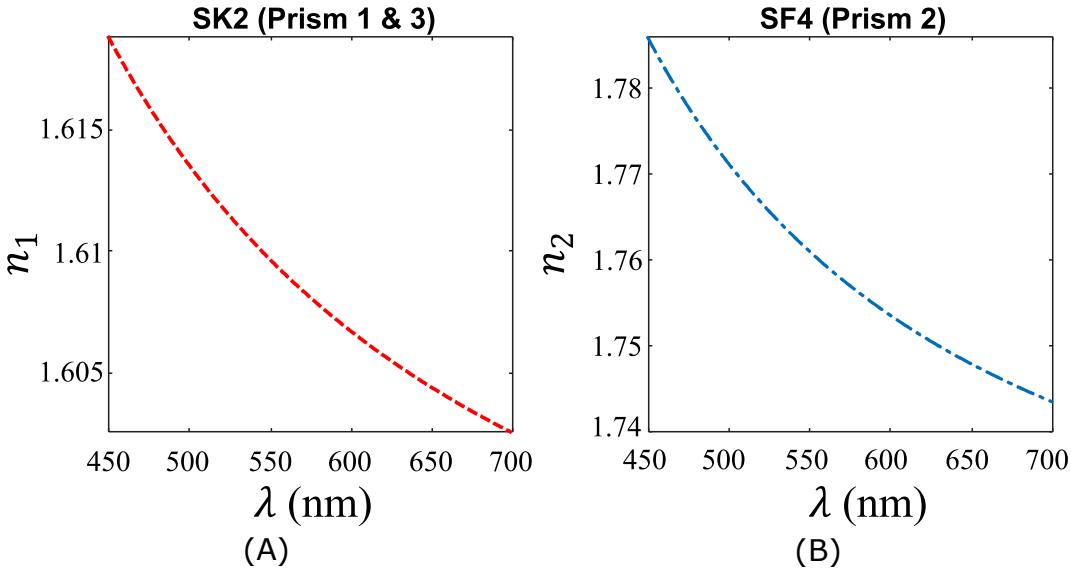


Figura 96. Wavelength-dependent refractive index for SK2 (n_1) and SF4 (n_2) glasses, resulting from the Eqs. (188) and (189), respectively.

and

$$n_2(\lambda) = \left[1 + \frac{1.61957826\lambda^2}{\lambda^2 - 0.0125502104} + \frac{0.339493189\lambda^2}{\lambda^2 - 0.0544559822} + \frac{1.02566931\lambda^2}{\lambda^2 - 117.652222}\right]^{\frac{1}{2}}, \quad (189)$$

respectively, where λ is defined in micrometers $[\mu m]$. Based on Eqs. (188) and (189), the refractive index functions are illustrated in Fig. 96 for the spectral range 400-700 nm.

ϕ_1 and ϕ_3 optimum range Having defined the refractive index glass materials (n_1 and n_2) and the Double Amici prism propagation geometry, we proceed to calculate the angles ϕ_1, and ϕ_3, which are obtained by establishing the following restriction: prism deviation angle must be zero for the central wavelength (i.e., $\delta(\lambda_c) = 0$). Here, the central wavelength is $\lambda_c = 570$ nm, where the first and last wavelengths are 450 nm and 700 nm, respectively. This spectral range is selected because, in this dissertation, we focused only on the visible spectral range (450-700 nm). The prism

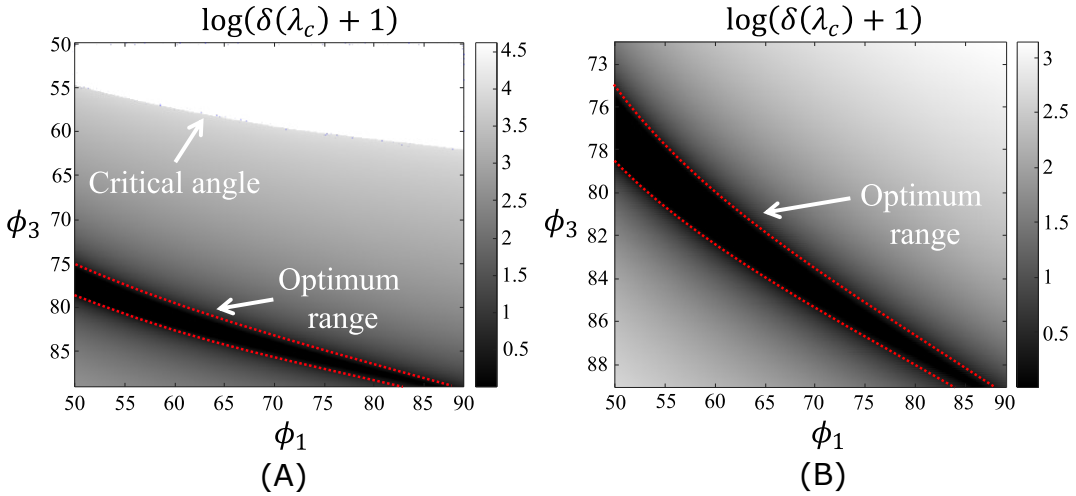


Figura 97. Prism deviation angle behavior in function of the ϕ_1 and ϕ_3. (A) and (B) deviation angle results for $\theta_0 = 0°$ in the range $50 \leq \phi_3 \leq 89°$

deviation is analyzed in two scenarios: a single incident angle to the propagation axis $\theta_0 = 0$ and the average for the range $-2.29° \leq \theta_0 \leq 2.29°$, as shown in Fig. 97 (A) and (B), respectively. Here, the angular bounds are obtained by calculating the resulting angular-field-of-view of a lens-detector configuration with a distance of 100 mm and a detector width of 4 mm. This bound correspond to an optical system with an angular field of view (FOV) equal to $2 \cdot \tan^{-1}\left(\frac{4mm}{(2)(100mm)}\right) = 2.29°$. From Fig. 97, we found that the linear function

$$\phi_3 = 0.3544\,(\phi_1 - 50) + 77.5, \tag{190}$$

describes the optimum range for ϕ_1 and ϕ_3, i.e., $\delta(\lambda_C) \approx 0$. Based on

To generated Fig. 97 the **Algorithms** 7 and **8** were developed, which are based on the Eqs. (172)-(189).

Shearing function characterization $S(\lambda)$ Having established the prism parameters, we formulate a shearing function that returns the total spatial displacement using the central wavelength λ_c as the reference point, i.e., the displacement for λ_c

Algorithm 7 Double amici prism function

Result: δ

1. $\theta_0, \phi_1, \phi_3, \lambda$
2. **Initialization:** $\alpha_1 = \pi - (\phi_1 + \phi_3), \alpha_2 = -(\pi - 2\phi_3), \beta_1 = -\alpha_1 - \frac{1}{2}\alpha_2$
3. $n_1(\lambda) \leftarrow$ via Eq. (188)
4. $n_2(\lambda) \leftarrow$ via Eq. (189)
5. $\theta_1 = \theta_0 - \beta_1,$
6. $\theta_1' = \arcsin\left(\frac{1}{n_1}\sin\theta_1\right),$
7. $\theta_2 = \theta_1' - \alpha_1,$
8. $\theta_2' = \arcsin\left(\frac{n_1}{n_2}\sin\theta_2\right),$
9. $\theta_3 = \theta_2' - \alpha_2,$
10. $\theta_3' = \arcsin\left(\frac{n_2}{n_1}\sin\theta_3\right),$
11. $\theta_4 = \theta_3' - \alpha_1$
12. $\theta_4' = \arcsin\left(n_1 \sin\left(\theta_4\right)\right),$
13. $\theta_5 = \theta_4' - \beta_1$
14. $\delta = \theta_0 - \theta_5$

Algorithm 8 Prism deviation angle

1. **Input:** $\lambda, t_0, t_1, t_2, t_3, t_4, t_5$
2. $a_1 \leftarrow 0$ 3. **for** $\phi_1 \leftarrow t_0$ until t_1
4. $a_1 \leftarrow a_1 + 1$
5. $a_2 \leftarrow 0$
6. **for** $\phi_3 \leftarrow t_2$ until t_3
7. $a_2 \leftarrow a_2 + 1$
8. $a_3 \leftarrow 0$
9. $r \leftarrow 0$
10. **for** $\theta_0 \leftarrow t_4$ until t_5
11. $a_3 \leftarrow a_3 + 1$
12. $\delta \leftarrow$ **Algorithm 1** with input $(\theta_0, \phi_1, \phi_3, \lambda)$
13. $r \leftarrow r + \delta$
14. $C_{a_1,a_2} \leftarrow \frac{r}{a_3}$
15. **Return:** C

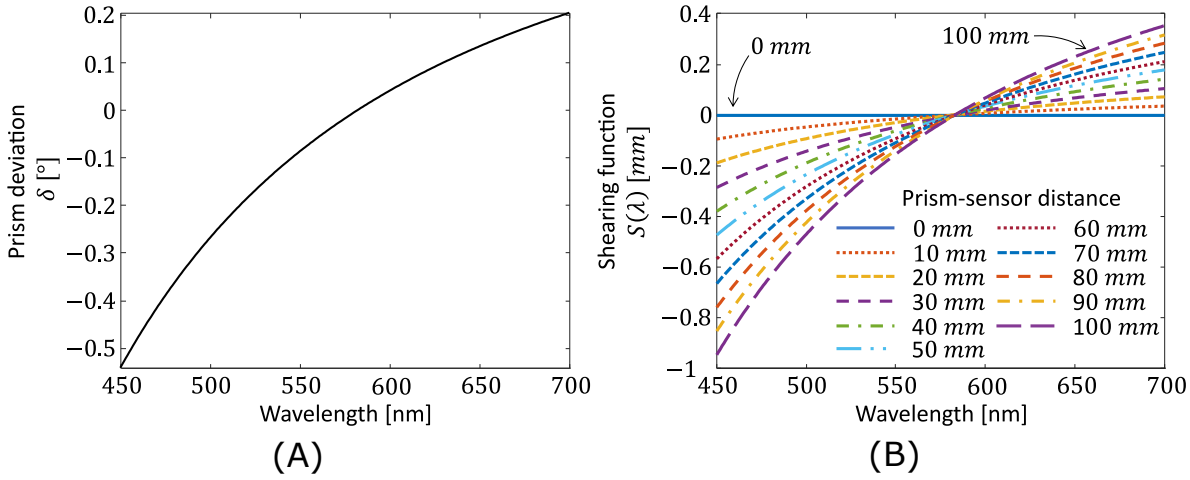


Figura 98. Customize double Amici prism spectral behavior with $\phi_1 = 85.9$, $\phi_3 = 73.7$, parameters. (A) Resulting prism deviation δ for $\theta_0 = 0°$. ((B) Shearing function $S(\lambda)$ for a prism-detector distance in the range $0 - 100$ [mm] and $\theta_0 = 0°$.

is 0. This function can be expressed as

$$S(\lambda) = x \cdot [\tan(\theta_5) + \tan(\delta(\lambda) + \theta_5)], \tag{191}$$

where x is the prism-sensor distance. Figure 98 illustrates the theoretical (A) prism deviation δ and the shearing $S(\lambda)$ functions for the customized double Amici prism with parameters $\phi_1 = 85.9$ and $\phi_3 = 73.7$. Specifically, the shearing function $S(\lambda)$ was tested with eleven different prism-sensor distances from 0 mm to 100 mm in steps of 10 mm. We customize an Amici double prism with the Shanghai Optics company using the parameters selected in this section $\phi_1 = 85.9$ and $\phi_3 = 73.7$ with central wavelength $\lambda_c = 570$ nm. Figure 99 (A) shows the customized double Amici prism which is stored in the HDSP optics laboratory. The prism is mounted in a customized prism mount attached to a precision 30 mm cage rotation mount. Following, we characterized the prism shearing function by mounting a 4f system where a DMD in the object plane is located, loaded with a binary calibration pattern, and a camera detector in the image plane. Specifically, Fig. 99 (B) illustrates the shearing function characterization for a prism-camera distance of 30 mm using a

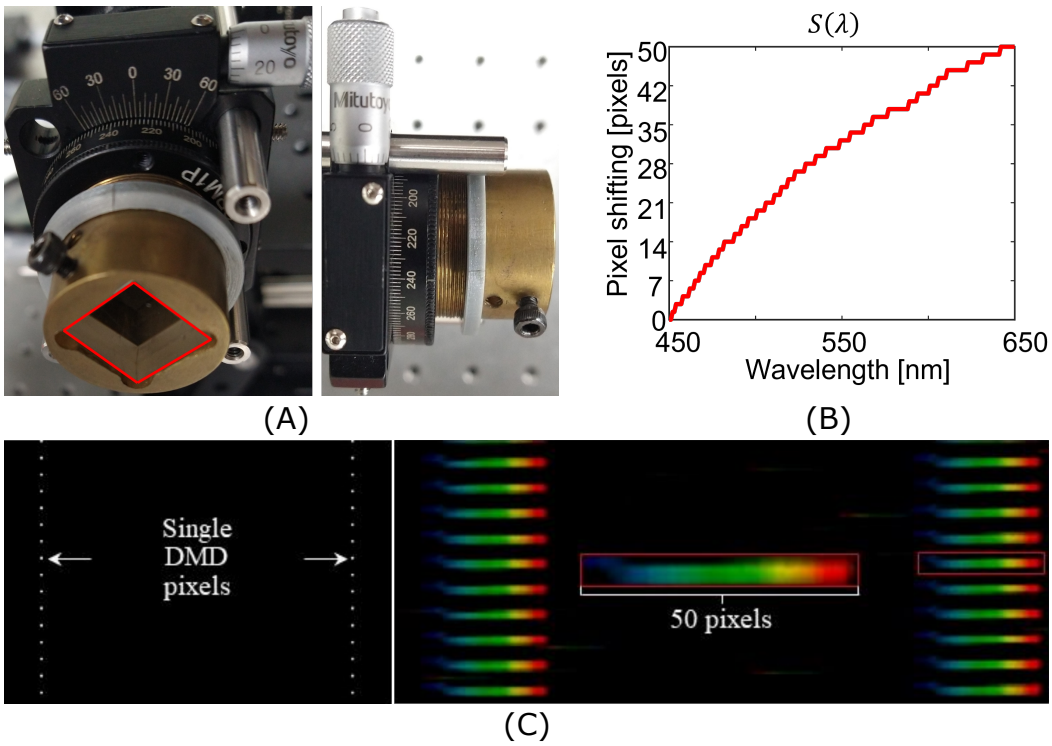


Figura 99. (A) Customized double Amici prism with the company shanghai optics. (B) Shearing function characterization using a Stingray camera with pixel size $6\mu m$. (C) Wavelengh depent point spread function (PSF) characterization using as point source a DMD with pixel size of $\approx 13\mu m$.

camera with a pixel pitch of 6 μm. This characterization results in a shearing function of 50 camera pixels in the 450-650 nm wavelength range, matching the theoretical results illustrates in Fig. 98(B). Figure 99(C) illustrates the DMD loaded pattern (left) and a color composite version of the calibrated wavelength depending PSF-prism.

Appendix C. Supplementary VSI

Beam-splitting-ratio Analysis via SNR

For illustration, in Fig. 100, the averaged intensity of the rows of the split measurements, $\mathbf{g}_T$ and $\mathbf{g}_R$, is compared to the noise, $\boldsymbol{\epsilon}$, setting an SNR=20 dB and using one of the 50 images described in the simulations. The SNR between each signal and the total noise is depicted at the different splitting-ratios $\alpha_R : \alpha_T$. Here, it can be noted that from an SNR perspective, a similar signal intensity between both measurements is achieved at around $\alpha_T = 0.6$.

Beam-splitting-ratio Analysis via dynamic range

Figure 101 illustrates the dynamic range analysis between the compressed measurements $\mathbf{g}_T$ and $\mathbf{g}_R$. This analysis is developed by defining an average absolute error metric as, $|\mathbf{D}_x\left((\mathbf{g}_T) \oslash (\mathbf{g}_R) - (\mathbf{g}_R) \oslash (\mathbf{g}_T)\right)|$, where $\frac{(\mathbf{g}_T)_i}{(\mathbf{g}_R)_i} = \frac{(\mathbf{g}_R)_i}{(\mathbf{g}_T)_i} = 1$ is the desired dynamic range performance, $\mathbf{D}_x = \frac{1}{N_x}(\mathbf{I}_{N_y+N_\lambda-1} \otimes \mathbf{1}_{N_x})$, $\mathbf{D}_x \in \mathbb{R}^{(N_y+N_\lambda-1)\times m}$ is a decimation matrix, and $\oslash$ represents the element-wise Hadamard division operator. Here it can be observed that the splitting ratio that promotes the lowest cumulative absolute error is located at $\alpha_T = 0.7$, and as expected, the worst results are located at $\alpha_T = 0$ and $\alpha_T = 1$.

Performance Evaluation for Different Additive Noise Scenarios

In this subsection, we measure the CASSI-VSI reconstruction performance under three noisy scenarios with 15, 20 and 30 dBs of SNR. Figure 102 summarizes the reconstruction results in terms of PSNR, SSIM, RMSE, and SAM, for the three different noise scenarios, varying the splitting ratio and working with the 50 spectral images

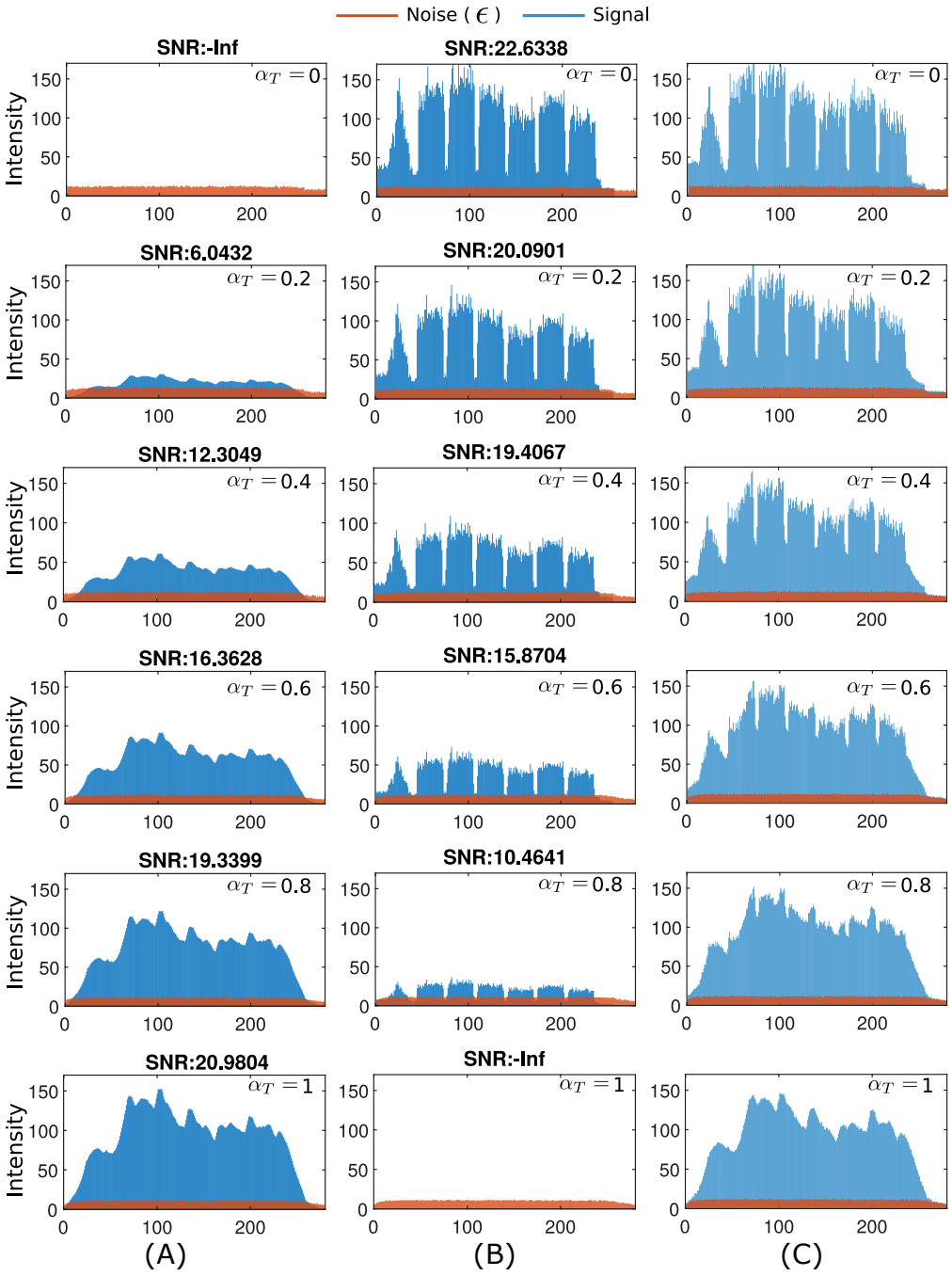


Figura 100. Signal-to-noise ratio analysis between the split signal intensities and the total sensor noise, assuming AWGN of 20 dB added to $\mathbf{g} = \mathbf{g}_T + \mathbf{g}_R$, and using one of the 50 images used in the simulations. (A) $\mathbf{g}_T$ vs. noise (B) $\mathbf{g}_R$ vs. noise. (C) $\mathbf{g}$ vs. noise. The signal intensities are plotted in blue and the noise in orange. The x-axis in each subplot represents the column pixels of the measurement, and the intensity is the averaged intensity along the rows of the sensor measurements.

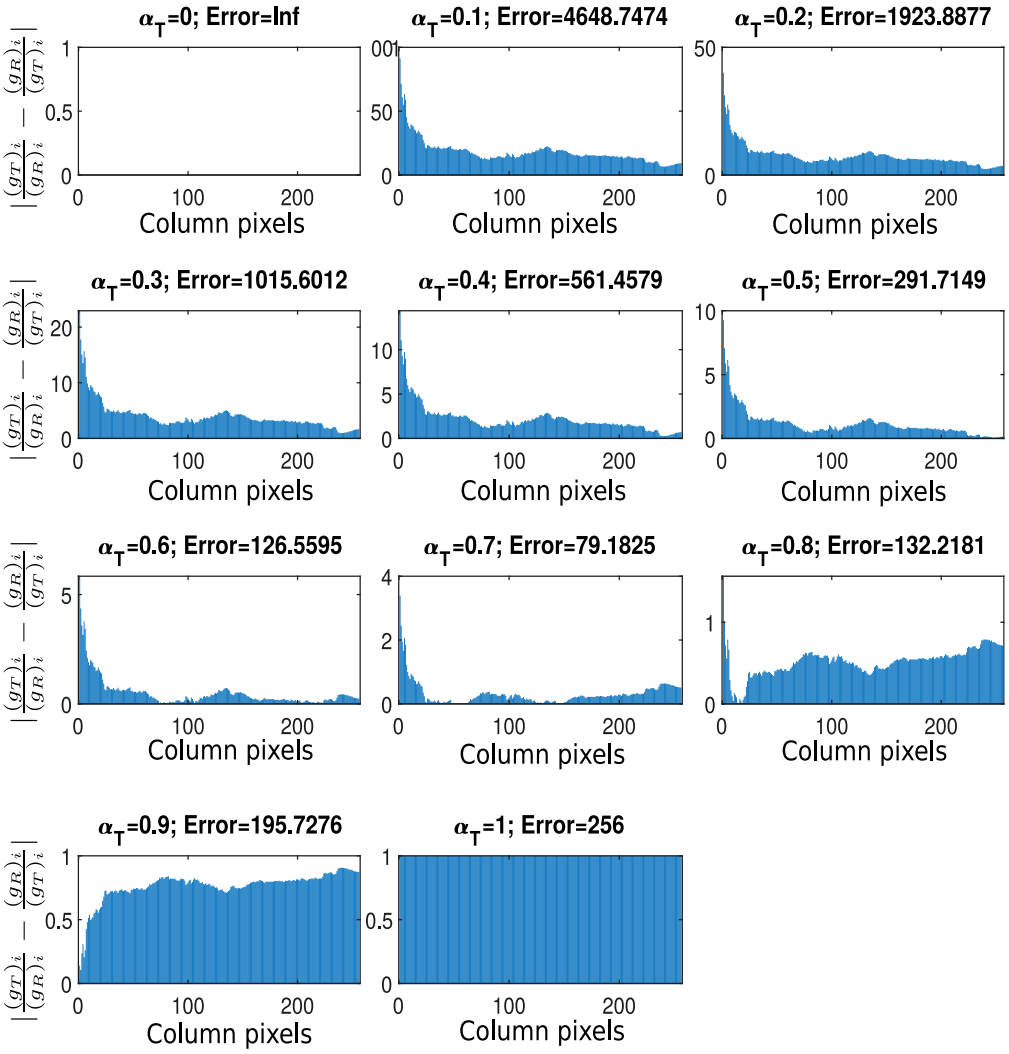


Figura 101. Absolute error analysis between the intensity ratios $\frac{(\mathbf{g}_T)_i}{(\mathbf{g}_R)_i}$ and $\frac{(\mathbf{g}_R)_i}{(\mathbf{g}_T)_i}$, where a ratio of 1 entails the same scale in both measurements, and thus the absolute difference $|\frac{(\mathbf{g}_T)_i}{(\mathbf{g}_R)_i} - \frac{(\mathbf{g}_R)_i}{(\mathbf{g}_T)_i}|$ must tend to 0. The minimum absolute error difference is attained at $\alpha_T = 0.7$.

used in the simulations. There it can be noticed that in terms of the PSNR, SSIM, and RMSE, the CASSI-VSI (blue squares) overcomes the CASSI (green squares), on average, for the three noisy scenarios. In terms of the SAM metric, the performance of the CASSI-VSI closely resembles that of the CASSI, but the latter yields a better result, as expected.

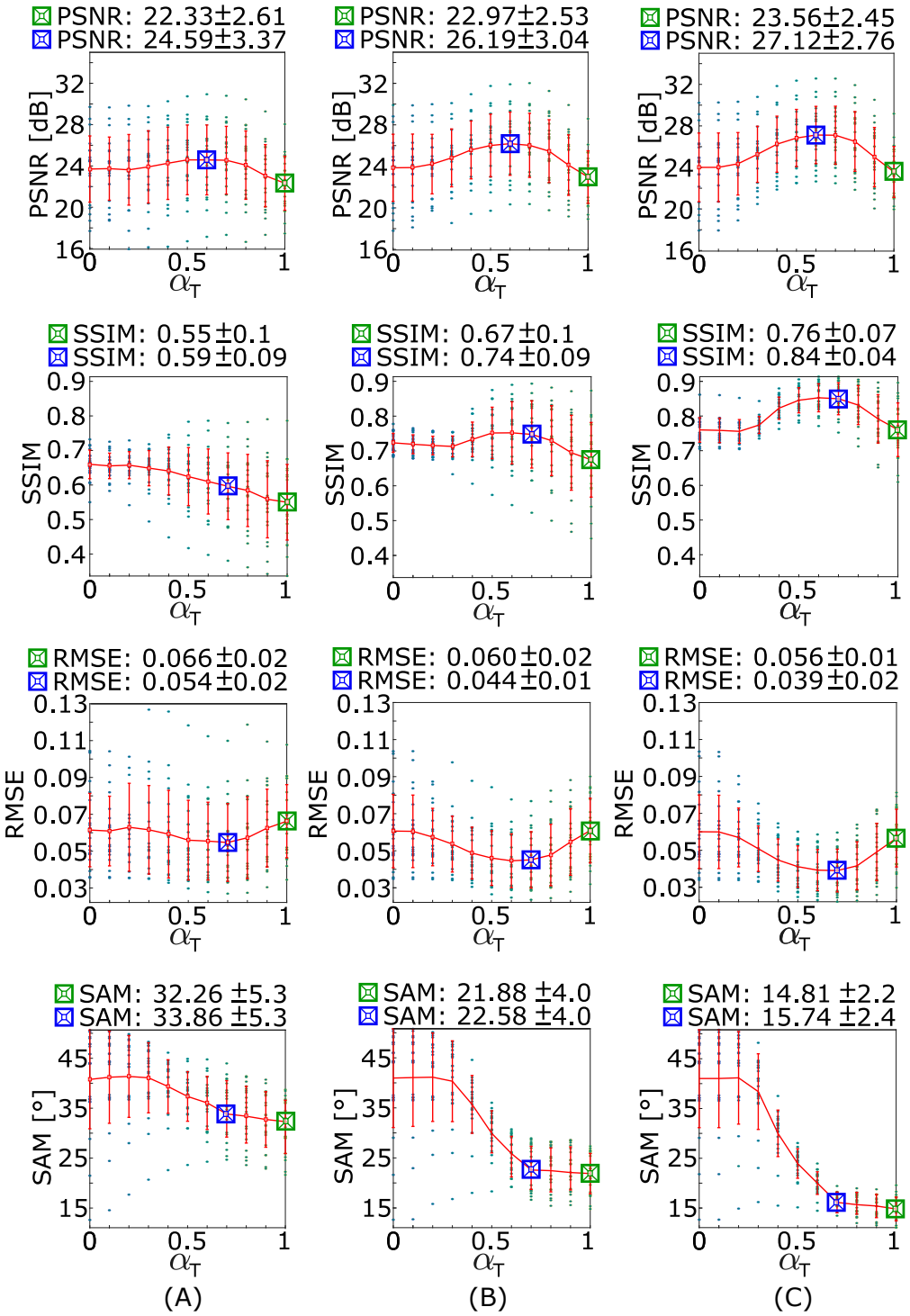


Figura 102. A-VSI reconstruction results in terms of PSNR, SSIM, RMSE, and SAM. (A) 15 dB, (B) 20 dB and (C) 30dB noise scenarios, where green and blue boxes highlights the CASSI and CASSI-VSI results.

Appendix D. Supplementary C^{3}SLFI

Tensor notation and preliminaries

To facilitate the understanding of the terminology used throughout the paper, and to make it self-contained, this section details the useful preliminaries and notation. In this work, scalars are denoted as lowercase letters $(a, b, \ldots)$, vectors as bold-faced lowercase letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices as bold-faced capital letters $(\mathbf{A}, \mathbf{B}, \ldots)$, and tensors as underlined bold-faced capital letters $(\underline{\mathbf{A}}, \underline{\mathbf{B}}, \ldots)$. Some useful product, decomposition, and tensorial representation properties are reviewed as follows Tamara G Kolda y Brett W Bader. "Tensor decompositions and applications". En: *SIAM review* 51.3 (2009), págs. 455-500.

- *Tensor notation:* A T-dimensional tensor is denoted as $\underline{\mathbf{X}} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_T}$, which can be written as a vector $\mathbf{x} = \text{vec}(\underline{\mathbf{X}})$, via the function $\text{vec}(.) : \{\mathbb{R}^{N_1 \times N_2 \times \ldots \times N_T} \rightarrow \mathbb{R}^{N_1 N_2 \ldots N_T}\}$. The letters $i_1, i_2, ..., i_T$ are used to index the $N_1, ..., N_T$ dimensions of the tensor $\underline{\mathbf{X}}$, respectively; that is, $i_k \in \{1, \ldots, N_T\}$, for $k = 1, \ldots, T$. Therefore, each element of the tensor $\underline{\mathbf{X}}$ is defined as $x_{i_1, i_2, \ldots, i_T}$.

- *Unfolding mode:* The unfolding mode-n version of a T-dimensional tensor $\underline{\mathbf{X}} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_T}$ is denoted by

$$\mathbf{X}_n = [\text{vec}(\underline{\mathbf{X}}_{:,:,\ldots,1,\ldots,:,:}), \ldots, \text{vec}(\underline{\mathbf{X}}_{:,:,\ldots,N_n,\ldots,:,:})], \tag{192}$$

  where $\mathbf{X}_n \in \mathbb{R}^{(\prod_{j=1, j \neq n}^{T} N_j) \times N_n}$.

- *n-mode product:* The n-mode (matrix) product of a tensor $\underline{\mathbf{X}}$ with a matrix $\mathbf{A} \in \mathbb{R}^{J \times N_n}$ is denoted by $\underline{\mathbf{X}} \times_n \mathbf{A}$, with size $N_1 \times \ldots \times N_{n-1} \times J \times N_{n+1} \times \ldots \times N_T$.

Element-wise, it is written as

$$(\underline{\mathbf{X}} \times_n \mathbf{A})_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_K} = \sum_{i_n=1}^{N_n} x_{i_1 i_2 \ldots i_K} a_{j i_n}. \tag{193}$$

- *Tucker model:* The tucker model seeks for a decomposition of a tensor $\underline{\mathbf{X}}$ as mode products of a core tensor $\underline{\mathbf{\Theta}} \in \mathbb{R}^{M_1 \times M_2 \times \ldots \times M_K}$ and K mode matrices $\mathbf{A}_{(K)} \in \mathbb{R}^{N_K \times M_K}$,

$$\underline{\mathbf{X}} \approx \underline{\mathbf{\Theta}} \times_1 \mathbf{A}_{(1)} \times_2 \mathbf{A}_{(2)} \ldots \times_K \mathbf{A}_{(K)}, \tag{194}$$

  which can be written in an element-wise form as

$$x_{i_1 i_2 \ldots i_K} \approx \sum_{j_1, j_2, \ldots, j_K} \theta_{j_1 j_2 \ldots j_K} a^{(1)}_{i_1 j_1} a^{(2)}_{i_2 j_2} \ldots a^{(K)}_{i_K j_K}. \tag{195}$$

Proof of Lemma 1

From Eqs. (73) and (92) it can be obtained that

$$\tilde{y}_{i_1, i_2, k} = \sum_{i_3=1}^{N_3} \underline{\tilde{F}}_{i_1, i_2, i_3} \underline{\tilde{T}}_{i_1, i_2, i_3, k}, \tag{196}$$

for $i_1 = 1, ..., \tilde{N}_1$, $i_2 = 1, ..., \tilde{N}_2$, $i_3 = 1, ..., N_3$ and $k = 1, ..., K$. Compacting the K-dimension, Eq. (196) can be rewritten as

$$\begin{aligned}\sum_{k=1}^{K} \tilde{y}_{i_1,i_2,k} &= \sum_{k=1}^{K}\sum_{i_3=1}^{N_3} \underline{\tilde{F}}_{i_1,i_2,i_3}\underline{\tilde{T}}_{i_1,i_2,i_3,k} \\ &= \sum_{i_3=1}^{N_3} \underline{\tilde{F}}_{i_1,i_2,i_3} \sum_{k=1}^{K} \underline{\tilde{T}}_{i_1,i_2,i_3,k} \\ &= \sum_{i_3=1}^{N_3} \underline{\tilde{F}}_{i_1,i_2,i_3}. \end{aligned} \tag{197}$$

The latter is obtained after using the fact that $\sum_{k=1}^{K} \underline{\tilde{T}}_{i_1,i_2,i_3,k} = 1$. Therefore, using Eq. (**??**), Eq. (197) can be rewritten as

$$\underline{\tilde{\mathbf{Y}}} \times_3 \mathbf{1}_K^T = \underline{\tilde{\mathbf{F}}} \times_3 \mathbf{1}_{N_3}^T, \tag{198}$$

and the equality in Eq. (89) holds.

Proof of Lemma 2

To establish a light notation, this problem is decomposed in three steps
(i) $\underline{\mathbf{\Gamma}} = \underline{\tilde{\mathbf{\Theta}}} \times_1 \mathbf{I}_{\tilde{N}_1 \times \tilde{N}_1} \times_2 \mathbf{I}_{\tilde{N}_2 \times \tilde{N}_2} \times_3 \mathbf{d}^T$,

$$\gamma_{ij} = \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{\ell=1}^{L} \tilde{\theta}_{mn\ell} d_{1\ell} \delta_{im} \delta_{jn} = \sum_{\ell=1}^{L} \tilde{\theta}_{ij\ell} d_{1\ell}. \tag{199}$$

(ii) $\underline{\mathbf{G}} = \underline{\mathbf{\Gamma}} \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \mathbf{I}_{N_2 \times N_2}$,

$$\begin{aligned} g_{i'j'} &= \sum_{m'}^{M} \sum_{n'}^{N} \gamma_{m'n'} (\tilde{\psi}_1)_{i'm'} \delta_{j'n'} \\ &= \sum_{m'}^{M} \sum_{n'}^{N} \left[\sum_{\ell}^{L} \tilde{\theta}_{ij\ell} d_{1\ell} \right] (\tilde{\psi}_1)_{i'm'} \delta_{j'n'} \\ &= \sum_{m'}^{M} \left[\sum_{\ell}^{L} (\tilde{\theta})_{m'j'\ell} d_{1\ell} \right] (\tilde{\psi}_1)_{i'm'}. \end{aligned} \tag{200}$$

(iii) $\underline{\mathbf{Z}} = \underline{\mathbf{G}} \times_1 \mathbf{I}_{N_1 \times N_1} \times_2 \tilde{\mathbf{\Psi}}_2$

$$\begin{aligned} z_{ij} &= \sum_{m}^{M} \sum_{n}^{N} g_{mn} (\tilde{\psi}_2)_{jn} \delta_{im} \\ &= \sum_{m}^{M} \sum_{n}^{N} \left[\sum_{m'}^{M} \left[\sum_{\ell}^{L} \tilde{\theta}_{m'n\ell} d_{1\ell} \right] a_{mm'} \right] (\tilde{\psi}_2)_{jn} \delta_{im} \\ &= \sum_{n}^{N} \left[\sum_{m'}^{M} \left[\sum_{\ell}^{L} \tilde{\theta}_{m'n\ell} d_{1\ell} \right] a_{im'} \right] (\tilde{\psi}_2)_{jn} \\ &= \sum_{n}^{N} \sum_{m'}^{M} \left[(\tilde{\psi}_1)_{im'} (\tilde{\psi}_2)_{jn} \left[\sum_{\ell}^{L} \tilde{\theta}_{m'n\ell} d_{1\ell} \right] \right]. \end{aligned} \tag{201}$$

Based on Eq. (73), the right side of Eq. (198) can be expressed as

$$\begin{aligned} z_{ij} &= \sum_{n}^{N} \sum_{m}^{M} \sum_{\ell}^{L} (\tilde{\psi}_1)_{im} (\tilde{\psi}_2)_{jn} \tilde{\theta}_{mn\ell} d_{1\ell} \\ &= \sum_{n}^{N} \sum_{m'}^{M} \left[(\tilde{\psi}_1)_{im'} (\tilde{\psi}_2)_{jn} \left[\sum_{\ell}^{L} \tilde{\theta}_{m'n\ell} d_{1\ell} \right] \right]. \end{aligned} \tag{202}$$

Thus, based on Eqs. (201) and (202), the equality $\underline{\tilde{\mathbf{\Theta}}} \times_3 \mathbf{d}^T \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \tilde{\mathbf{\Psi}}_2 = \underline{\tilde{\mathbf{\Theta}}} \times_1 \tilde{\mathbf{\Psi}}_1 \times_2 \tilde{\mathbf{\Psi}}_2 \times_3 \mathbf{d}^T$ in Eq. (91) holds.

Tensorial inversion via the Sherman-Morrison-Woodbury lemma

Optimization problem in Eq. (101) is equivalent to

$$\arg\min_{\tilde{\boldsymbol{\theta}}} \mu||\tilde{\boldsymbol{\theta}} - \boldsymbol{\omega}^{\iota} + \boldsymbol{\nu}^{\iota}||_F^2 + ||\tilde{\mathbf{y}} - \mathbf{H}\tilde{\boldsymbol{\theta}}||_F^2, \tag{203}$$

where $\mathbf{H} = \tilde{\boldsymbol{\Phi}}_3\tilde{\boldsymbol{\Psi}}_3 \otimes \tilde{\boldsymbol{\Phi}}_2\tilde{\boldsymbol{\Psi}}_2 \otimes \tilde{\boldsymbol{\Phi}}_1\tilde{\boldsymbol{\Psi}}_1$, $\mathbf{H} = \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Psi}}$, $\tilde{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}}_3 \otimes \tilde{\boldsymbol{\Phi}}_2 \otimes \tilde{\boldsymbol{\Phi}}_1$, $\tilde{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}_3 \otimes \tilde{\boldsymbol{\Psi}}_2 \otimes \tilde{\boldsymbol{\Psi}}_1$ with $\tilde{\boldsymbol{\Psi}}^T\tilde{\boldsymbol{\Psi}} = \mathbf{I}$, $\tilde{\boldsymbol{\theta}} = \text{vec}(\underline{\tilde{\boldsymbol{\Theta}}})$, $\boldsymbol{\omega}^{\iota} = \text{vec}(\underline{\boldsymbol{\Omega}}^{\iota})$, $\tilde{\mathbf{y}} = \text{vec}(\underline{\tilde{\mathbf{Y}}})$, and $\boldsymbol{\nu}^{\iota} = \underline{\boldsymbol{\nu}}^{\iota}$. Problem in Eq. (203) has the closed-form solution

$$\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\Psi}}\left(\tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}} + \mu\mathbf{I}\right)^{-1}\left(\tilde{\boldsymbol{\Phi}}^T\tilde{\mathbf{y}} + \mu\tilde{\boldsymbol{\Psi}}^T(\boldsymbol{\omega} - \boldsymbol{\nu})\right). \tag{204}$$

Based on the Sherman-Morrison-Woodbury Lemma Mehmet A Akgün, John H Garcelon y Raphael T Haftka. “Fast exact linear and non-linear structural reanalysis and the Sherman–Morrison–Woodbury formulas”. En: *International Journal for Numerical Methods in Engineering* 50.7 (2001), págs. 1587-1606, Eq. (204) can be rewritten as

$$\tilde{\boldsymbol{\theta}} = \mu^{-1}\tilde{\boldsymbol{\Psi}}\left[\mathbf{I} - \mu^{-1}\tilde{\boldsymbol{\Phi}}^T\left(\mu^{-1}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Phi}}^T + \mathbf{I}\right)^{-1}\tilde{\boldsymbol{\Phi}}\right]\left(\tilde{\boldsymbol{\Phi}}^T\tilde{\mathbf{y}} + \mu\tilde{\boldsymbol{\Psi}}^T(\boldsymbol{\omega} - \boldsymbol{\nu})\right). \tag{205}$$

The complexity of problem Eq. (205) is higher than that of problem Eq. (101), due to the larger dimension of the matrix product involved in $\tilde{\boldsymbol{\Phi}}$. However, based on the sensing geometry of the proposed architecture in Eq. (**??**), operations involving $\tilde{\boldsymbol{\Phi}}$ can be done faster than its size suggests by exploiting its tensorial structure. For

arbitrary vectors $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, we have

$$\begin{aligned} \mathbf{\Phi f} &= \text{vec}\left(\prod_{j=1}^{3} \underline{\tilde{\mathbf{F}}} \times_j \tilde{\mathbf{\Phi}}_j\right) \\ &= \text{vec}\left(\text{Arrange}\left(\left[\underline{\tilde{\mathbf{T}}} \circ \left(\underline{\tilde{\mathbf{F}}} \times_4 \mathbf{1}_K\right)\right] \times_3 \mathbf{1}_{N_3}^T\right)\right) \\ &= \text{vec}\left(\underline{\tilde{\mathbf{B}}}\right), \end{aligned} \tag{206}$$

and

$$\begin{aligned} \mathbf{\Phi}^T\mathbf{y} &= \text{vec}\left(\prod_{j=1}^{3} \underline{\tilde{\mathbf{Y}}} \times_j \tilde{\mathbf{\Phi}}_j^T\right) \\ &= \text{vec}\left(\sum_{k=1}^{K}\left[\underline{\tilde{\mathbf{T}}}_{:::k} \circ \left(\underline{\tilde{\mathbf{Y}}}_{::k} \times_3 \mathbf{1}_{N_3}\right)\right]\right), \end{aligned} \tag{207}$$

where $\underline{\tilde{\mathbf{B}}}_{:,:,k} = \left(\underline{\tilde{\mathbf{T}}}_{:,:,:,k} \circ \underline{\tilde{\mathbf{F}}}\right) \times_3 \mathbf{1}_{N_3}^T$, $\mathbf{1}_{N_3} \in \mathbb{R}^{N_3}$ is a vector of ones, $\underline{\tilde{\mathbf{T}}}^k \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3}$ is a 3D tensor representing the color-coded aperture for the kth snapshot and $\circ$ represents the Hadamard product. The right side of Eqs. (206) and (207) can be computed very efficiently with $\mathcal{O}(\tilde{N}_1\tilde{N}_2N_3)$ cost, in contrast to the left side which has a computational cost of $\mathcal{O}(m\tilde{N}_1\tilde{N}_2N_3)$. Similarly, operations involving $\tilde{\mathbf{\Psi}}$ can be performed with a lower complexity by using Eq. (90), which allows to reduce the complexity from $\mathcal{O}(\tilde{N}_1^2\tilde{N}_2^2N_3^2)$ to $\mathcal{O}(\max(\tilde{N}_1, \tilde{N}_2, \tilde{N}_3)\tilde{N}_1\tilde{N}_2\tilde{N}_3)$. Based on the fact that the tensor $\underline{\tilde{\mathbf{T}}} \in \mathbb{R}^{\tilde{N}_1 \times \tilde{N}_2 \times N_3 \times K}$ meets the complementary coded apertures condition

$$||\text{vec}(\mathbf{1}_{\tilde{N}_1}^T\mathbf{1}_{\tilde{N}_2} \times_3 \mathbf{1}_{N_3}) - \underline{\tilde{\mathbf{T}}} \times_4 \mathbf{1}_K^T||_2 = 0 \tag{208}$$

the inversion of the matrix $(\mathbf{I} + \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^T)$ can be calculated in a computationally efficient way by using the tensorial notation for

$$\underline{\mathbf{R}}_{:,:,k_1} = \sum_{k_2=1}^{K}\left[\underline{\mathbf{H}}_{:,:,k_1,k_2} \circ \underline{\tilde{\mathbf{Y}}}_{:,:,k_1}\right], \tag{209}$$

$$\underline{\mathbf{H}}_{i_1,i_2,:,:} = \text{Rearrange}\left((\text{Arrange}(Q_{i_1,i_2,:,:}))^{-1}\right) \tag{210}$$

$$Q_{i_1,i_2,k_1,k_2} = \left[\sum_{i_3=1}^{N_3} \tilde{T}_{i_1,i_2,k_1}\tilde{T}_{i_1,i_2,k_2}\right] + \left(1 - e^{-\delta(k_1-k_2)}\right) \tag{211}$$

with $k_1 = \{1, ..., K\}$, Arrange$(\cdot) : \mathbb{R}^{1\times1\times K\times K} \rightarrow \mathbb{R}^{K\times K}$ is a function that collapses the first and second dimension, Rearrange$(\cdot) : \mathbb{R}^{K\times K} \rightarrow \mathbb{R}^{1\times1\times K\times K}$ is a function that rearranges a second-order tensor in a fourth-order tensor, $\delta(\cdot)$ represent the delta Dirac function, where $\tilde{\mathbf{\Gamma}}_3 \otimes \tilde{\mathbf{\Gamma}}_2 \otimes \tilde{\mathbf{\Gamma}}_1 = (\tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^T + \mathbf{I}_{m\times m})$ with $m = \tilde{N}_1\tilde{N}_2K$. Here it is worth noting that the tensorial process defined in Eq. (205) is equivalent to $\left[\tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^T + \mathbf{I}\right]^{-1}\tilde{\mathbf{\Phi}}vec(\underline{\mathbf{G}})$. To demonstrate this tensorial equivalence, Fig. 103 shows the cumulative absolute error, for one thousand executions, between Eq. (104) and $\left[\tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^T + \mathbf{I}\right]^{-1}\tilde{\mathbf{\Phi}}vec(\underline{\mathbf{G}})$ using random $\underline{\mathbf{G}}$ and $\underline{\tilde{\mathbf{T}}}$ for each execution, with $\tilde{N}_1 = \tilde{N}_2 = 32$, $N_3 = 8$ and $K = 4$. In general, for two T-order tensors $\underline{\mathbf{F}}$ and $\underline{\mathbf{Y}}$ with vector counterparts $\mathbf{f} =$ vec$(\underline{\mathbf{F}})$ and $\mathbf{y} =$ vec$(\underline{\mathbf{Y}})$, $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, the traditional linear matrix model $\mathbf{y} = \mathbf{\Phi f}$, where $\mathbf{\Phi} = \mathbf{A}^{(T)} \otimes \mathbf{A}^{(T-1)} \otimes ... \otimes \mathbf{A}^{(1)}$, $\mathbf{A}^{(i)} \in \mathbb{R}^{\sqrt[T]{n}\times\sqrt[T]{n}}$, can be expressed in tensorial notation as $\underline{\mathbf{Y}} = \prod_{i=1}^{T} \underline{\mathbf{F}} \times_i \mathbf{A}^{(i)}$; the computational complexity of the former is $O(n^2)$ while the latter requires $O(nT\sqrt[T]{n^2})$.

Complexity and convergence analysis of the proposed Algorithm 5

The complexity of the proposed **Algorithm 5** is calculated as follows. The initialization of $\underline{\mathbf{Z}}$ in Line 1, is a tensor product with $O(K^2\tilde{N}_1\tilde{N}_2)$, and the calculation of $\mathbf{Z}_1$ in Line 2 is $O(\tilde{N}_1\tilde{N}_2)$. To estimate $\tilde{\mathbf{\Psi}}_1$ and $\tilde{\mathbf{\Psi}}_2$, the closed-form solutions to Eqs. (92)- (93) are obtained via Tucker decomposition in Lines 3-4, with a computational complexity of $O(\min\{\tilde{N}_1\tilde{N}_2^2, \tilde{N}_1^2\tilde{N}_2\})$. Matrices $\underline{\mathbf{Q}}$ and $\underline{\mathbf{H}}$ are initialized in Lines 6-7, exhibiting a computational complexity of $O(K\tilde{N}_1\tilde{N}_2N_3^2)$ and $O(K\tilde{N}_1\tilde{N}_2)$, respectively. To solve Eq. (101), Algorithm 5 summarizes the required steps in Lines 10, 11 and 12, whose

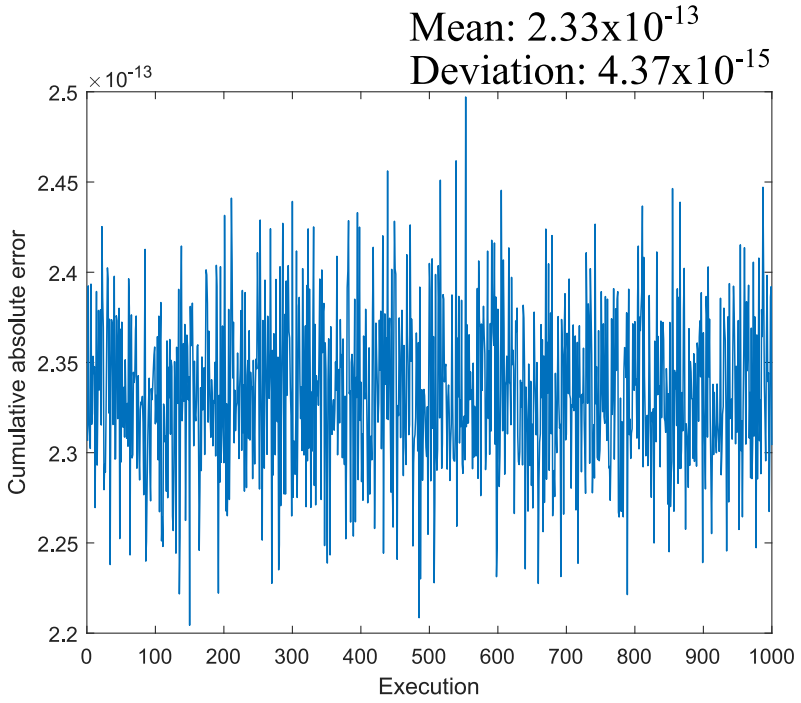


Figura 103. Error estimation between the proposed tensorial inverse and the traditional matrix inversion through one thousand runs.

computational complexities are given by $O(\max\{\tilde{N}_1, \tilde{N}_2, N_3\}\tilde{N}_1\tilde{N}_2N_3)$, $O(K\tilde{N}_1\tilde{N}_2N_3)$ and $O(\max\{\tilde{N}_1, \tilde{N}_2, N_3\}\tilde{N}_1\tilde{N}_2N_3)$, respectively. To estimate $\underline{\tilde{\mathbf{\Omega}}}$, the closed-form solution to Eq. (102) is obtained via soft-thresholding in Line 13, with a computational complexity of $O(\tilde{N}_1\tilde{N}_2N_3)$. To update $\tilde{\mathbf{\Psi}}_3$, Lines 16, 17, and 18 are executed, whose computational complexities are given by $O(\max\{\tilde{N}_1, \tilde{N}_2, N_3\}\tilde{N}_1\tilde{N}_2N_3)$, $O(\tilde{N}_1\tilde{N}_2N_3)$ and $O(\min\{\tilde{N}_1\tilde{N}_2N_3^2, \tilde{N}_1^2\tilde{N}_2^2N_3\})$, respectively. In general, the computational complexity of **Algorithm 5** is bounded by $O(\max\{\tilde{N}_1, \tilde{N}_2, N_3\}n_1)$ with $n_1 = \tilde{N}_1\tilde{N}_2N_3$ or similarly $O(\max\{\tilde{N}_1, \tilde{N}_2, N_3\}n)$. Assuming $\tilde{N}_1 = \tilde{N}_2 = N_3 = \sqrt[3]{n}$, the computational complexity of the proposed algorithm is given by $O(3n\sqrt[3]{n^2})$. In general, the computational complexity for a T-dimensional tensor with $N_1 = N_2 = ... = N_T = \sqrt[T]{n}$ is bounded by $O(Tn\sqrt[T]{n^2})$.

To solve the optimization problems established in Eqs. (92)- (93), and (97), the Tucker decomposition approach is used, which guarantees that the resulting factor ma-

trices $\mathbf{\Psi}_1$, $\mathbf{\Psi}_2$, $\mathbf{\Psi}_3$, are orthogonal and represent the principal components of each unfolding-tensor mode. Note that, the Tucker decomposition is regarded as a higher order generalization of the singular value decomposition (SVD) Kolda y Bader, “Tensor decompositions and applications”. The optimization problem established in Eq. (100) is composed of three proper convex and closed functions, $f_1(\cdot)$, $f_2(\cdot)$ and $\|\cdot\|_F^2$. Therefore, if $\mu > 0$, we may invoke the theorem [Eckstein y Bertsekas, “On the Douglasâ€”Rachford splitting method and the proximal point algorithm for maximal monotone operators.”, Theorem 8] by Eckstein and Bertsekas, which, guarantees the convergence of **Algorithm 5**.