End-to-end Optimization of a Coded Snapshot Computational Imaging System for Spectral-Depth Estimation

M.Sc. Edwin Vargas

Doctoral thesis to qualify for the title of Doctor of Engineering, electronics area

Advisor

Henry Arguello

Ph.D. in Electrical and Computer Engineering

Universidad Industrial de Santander

School of Physical-Mechanical Engineering

Department of Electrical, Electronics, and Telecommunications Engineering

Bucaramanga

2023

Content

List of Figures

page

List of Tables

page

List of Appendices

page

List of Abbreviations

TMCA Time Multiplexed Coded Aperture

DBD Dual Blind Deconvolution

SI Spectral Imaging

HSI Hyperspectral Imaging

CS Compressive Sensing

CSI Compressive Spectral Imaging

SSC Sparse Subspace Clustering

LRR Low Rank Representation

SC Subspace Clustering

PCA Principal Component Analysis

SVD Singular Value Decomposition

LSA Local Subspace Affinity

LLMC Locally Linear Manifold Clustering

SLBF Spectral Local Best-fit Flats

OMP Orthogonal Matching Pursuit

RIP Restricted Isometry Property

RP Random Projection

3D-CASSI Spatial-spectral Coded Compressive Spectral Imager

SP Single-pixel Camera

Resumen

Título: Optimización de extremo a extremo de un sistema computacional codificado de única captura para la estimación de información espectral y de profundidad [*]

Autor: Edwin Mauricio Vargas Díaz [**]

Palabras Clave: Apertura codificada, obturador codificado, imágenes de profundidad y espectrales, Optimización de extremo a extremo.

Descripción: Las imágenes de profundidad y espectrales son tecnologías indispensables para multiples aplicaciones, pero tradicionalmente se han explorado de manera independiente. Esfuerzos recientes han estado dirigidos hacia la codificación óptica de información espectral y de profundidad (EP) en una sola captura del sensor, que posteriormente se decodifica mediante un algoritmo computacional. La efectividad de los sistemas EP de única captura depende principalmente de la modulación óptica, que se conoce como codificación, y las técnicas computacionales empleadas para extraer información EP de las medidas codificadas. La modulación óptica se realiza convencionalmente utilizando aperturas codificadas (AC), máscaras de fase, prismas o rejillas, iluminación activa, entre otros. Esta tesis propone una estrategia de modulación óptica (codificación) mediante la sincronización de una apertura codificada en fase y un obturador de píxeles arbitrario en conjunto con una apertura codificada de color en el sensor. El multiplexado en el tiempo de la apertura codificada (MTAC) mejora las estrategias de codificación convencionales sin agregar nuevos elementos ópticos. Esta tesis demuestra que el MTAC propuesto implica una función de dispersión de puntos (FDP) espacialmente variable para un profundidad constante en una escena, lo que facilita la distinción, y por tanto, una mejor recuperación de

---

[*] Tesis de doctorado

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones. Director: Henry Arguello Fuentes, Doctorado en Ingeniería.

la información de profundidad. Además, el filtrado selectivo de bandas espectrales específicas por parte del CCA codifica las características relevantes de la información espectral que se decodifica usando un algoritmo de reconstrucción. Para optimizar el MTAC, esta tesis aprovecha los avances de las técnicas de aprendizaje profundo para aprender conjuntamente la modulación óptica y el algoritmo de decodificación computacional en un marco de trabajo de extremo a extremo (E2E). El enfoque propuesto demuestra a través de simulaciones numéricas y escenarios reales que el enfoque propuesto supera a los sistemas EP de única captura del estado del arte. Finalmente, esta tesis doctoral muestra que el concepto fundamental MTCA logra un mayor rendimiento cuando se aplica a sistemas computacionales de imágenes espectrales compresivas e imágenes de campo de luz, demostrando así su amplia aplicabilidad.

Abstract

Title: End-to-end Optimization of a Coded Snapshot Computational Imaging System for Spectral-Depth Estimation [*]

Author: Edwin Mauricio Vargas Díaz [**]

Keywords: Coded aperture, coded exposure, spectral-depth imaging, end-to-end optimization.

Description: Depth and spectral imaging are indispensable technologies for a myriad of applications but have been traditionally explored independently from each other. Recent endeavors are directed toward the optical encoding of spectral-depth (SD) information in a single image sensor measurement, which is subsequently decoded by a computational algorithm. The effectiveness of single snapshot SD imaging systems primarily depends on the optical modulation function, which is referred to as codification, and the computational techniques employed for extracting SD information from the encoded measurements. The optical modulation is conventionally realized using coded apertures (CAs), phase masks, prisms or gratings, active illumination, and many others. This dissertation proposes an optical modulation (codification) strategy by synchronizing a phase-coded aperture and a spatially varying pixel shutter in conjunction with a color-coded aperture in the sensor. The proposed time-multiplexed coded aperture (TMCA) improves conventional coding strategies without adding new optical elements. This thesis demonstrates that the proposed TMCA entails a spatially-variant point spread function (PSF) for a constant depth in a scene, which, in turn, facilitates the distinguishability, and therefore, better recovery of the depth information. Further, the selective filtering of specific spectral bands by the CCA encodes relevant spectral information that is disentangled using a reconstruction algorithm. To optimize

---

[*] Doctoral Thesis

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones. Director: Henry Arguello Fuentes, Doctorado en Ingeniería.

TMCA, this thesis leverages the advances of deep learning techniques to jointly learn the optical modulation and the computational decoding algorithm in an end-to-end (E2E) framework. The proposed approach is demonstrated via numerical simulations and real scenarios showing that the proposed approach outperforms state-of-the-art snapshot SD imaging alternatives. Ultimately, this doctoral thesis shows that the foundational concept of TMCA achieves higher performance when applied to compressive spectral imaging and light field imaging, thereby demonstrating its broader applicability.

Introduction

Snapshot imaging techniques have evolved rapidly in recent years, with a primary focus on spectral and depth dimensions. These two quantities are of particular interest since radiation at multiple electromagnetic bands provides information about the object material, while the object location is crucial for 3D-perception tasks and scene understanding (Shaw and Burke, 2003b; Song et al., 2015). The simultaneous acquisition of light reflectance at multiple wavelengths and knowledge of the objects location within a scene enables potential applications in areas such as plant phenotyping (Liu et al., 2020), material analysis under challenging conditions (Wu et al., 2016), autonomous driving (Hansard et al., 2012), and biomedical imaging (Radosevich et al., 2008).

Hyperspectral (HS) sensors collect data that can be represented by a three-dimensional data cube (Hsu and Burke, 2003). This data cube, referred to as HS image, is a collection of 2D images, where each 2D image is captured at a specific wavelength. HS images are characterized by a high spectral resolution, which allows accurate identification of the different materials contained in the scene of interest. Analyzing the spectral information of HS images has allowed the development of many applications in the fields of remote sensing (Schaepman et al., 2009), medical imaging (Panasyuk et al., 2007), or astronomy (Schreiber et al., 2009). Traditional spectral imaging systems use scanning methodologies in the spectral and spatial domain, making it challenging to capture dynamic scenes. Furthermore, HS images require to acquire a large amount of data that must be stored and transmitted. More recently, based on the compressed sensing theory developments, new compressive spectral imaging (CSI) systems for snapshot imaging have been proposed yielding

impressive results (Gehm et al., 2007; Arce et al., 2014; Correa et al., 2016). CSI captures a reduced number of measurements compared to conventional spectral imaging devices, becoming attractive for many practical applications. Reconstructing the underlying spectral information from the compressed measurements is performed via sparse coding or, more recently, deep learning approaches (Arguello and Arce, 2014; Choi et al., 2017; Wang et al., 2018; Jeon et al., 2019).

On the other hand, imaging systems with depth estimation capabilities are essential for many computer vision applications such as autonomous navigation, action recognition, or robot guidance (Biswas and Veloso, 2012; Jalal et al., 2019; Wen et al., 2017). Several standard solution approaches usually employ sophisticated camera systems such as a stereo camera or a camera array, pulsed LiDAR (Altmann et al., 2016), or time of flight (Zhu et al., 2008). Alternatively, employing structured-light techniques along with algorithms based on geometric triangulation has also been proposed to estimate depth information. A more straightforward approach that has been an active research area is depth estimation from a single image with conventional cameras (Godard et al., 2017; Fu et al., 2018; Chang and Wetzstein, 2019). Even though a single image lacks the depth cues that exist in a stereo image pair, traditional approaches leverage predefined image features that are statistically correlated with depth, e.g., shading, perspective distortions, occlusions, texture gradients, and haze (Hoiem et al., 2007; Henderson and Ferrari, 2019; Paschalidou et al., 2019). Besides, depth cues have been estimated using learned features from complex convolutional neural networks (CNNs), improving the performance of approaches employing hand-crafted features (Eigen et al., 2014; Fu et al., 2018). In contrast, it has also been shown that coded defocus blur or chromatic aberrations have the potential to generate strong depth cues and help estimate an

accurate depth map. Thus, imaging systems that employ an amplitude or phase coding mask have been proposed in the last two decades to optically encode the depth-dependence of the scene (Levin et al., 2007; Haim et al., 2018; Chang and Wetzstein, 2019). Common to all these approaches is to promote optical cues with a point spread function (PSF) that is depth-dependent.

As described above, depth and spectral imaging have been mainly studied as individual problems. Recently, few efforts have been made to propose high-dimensional imaging systems that can simultaneously acquire spectral as well as depth information. Pioneer works in the joint imaging of depth and spectrum include the combination of complex imaging systems that independently capture spectral and depth information with separate hardware and combine or render them in a computational stage. For instance, Kim et. al., introduced a combined depth and spectral approach for measuring 3D patterns on solid objects (Kim et al., 2012). Other examples include a cross-modal stereo system (Wang et al., 2016) and, a 5D hyperspectral imaging system (Heist et al., 2018) using structured light. In general, these approaches employing two different imaging modalities to estimate spectrum and depth significantly increase the device size. Furthermore, depending on its complexity, the alignment and registration process of the depth cues across different spectral bands can be a challenging problem in terms of calibration. Under a compressive multimodal imaging framework, (Horisaki et al., 2013), the Coded Aperture Snapshot Spectral Imaging (CASSI) system has also been combined with light field or time of flight (TOF) imaging to propose a more compact snapshot monocular imaging system (Feng et al., 2016; Rueda-Chacon et al., 2019). However, these systems employ specialized optical coding strategies sensitive to ambient light, limiting outdoor usage. To reduce the form factor, and to alleviate the requirement of

external illumination, authors in (Baek et al., 2021) proposed a compact snapshot imaging system constituted of just a single diffractive optical element (DOE) and an RGB sensor.

Moreover, conventional spectral-depth imaging systems design the optical system first. Then, the parameters in the image processing algorithm are tuned or learned to obtain the best image recovery. Considering the limits of this sequential design approach, a more recent framework dubbed deep optics introduces the concept of optimizing the optical elements jointly with the reconstruction algorithm in an end-to-end fashion employing stochastic optimization Sitzmann et al. (2018). Based on differentiable reconstruction algorithms, this framework allows the optimization of domain-specific computational cameras. This idea has been investigated in applications such as extended depth of field (Sitzmann et al., 2018), high dynamic range imaging (Metzler et al., 2020), spectral imaging (Wang et al., 2018), depth estimation (Chang and Wetzstein, 2019) and more recently for spectral-depth imaging (Baek et al., 2021). In general, recovering spectral-depth (SD) information jointly from a single snapshot is an extremely ill-posed problem. The ability to recover these quantities accurately depends on the properties of the optical modulation and the employed computational algorithm. This doctoral thesis addresses the problem of spectral-depth imaging from a single snapshot by proposing a new optical codification called Time-Multiplexed Coded Aperture (TMCA). The fundamental concept behind TMCA is to acquire novel codifications by synchronizing an amplitude or phase coded aperture (CA) that changes its pattern in time –dubbed as a time-varying CA– and a spatially varying shutter function realizing a coded exposure. In general, this research demonstrates that TMCA produces better codification capabilities for 3 different applications: hyperspectral imaging, light field imaging, and spectral-depth imag-

ing application. Furthermore, this thesis proposes to learn the TMCA codification, inspired by works in deep optics (Martel et al., 2020; Chang and Wetzstein, 2019; Metzler et al., 2019; Wu et al., 2019; Nehme et al., 2020; Wetzstein et al., 2020). More precisely, this research performs the joint optimization of the TMCA codes along with a neural network (NN) used to recover the light fields, hyperspectral images or depth maps. The end-to-end (E2E) differentiable approach can be interpreted as an encoder-decoder framework in which the encoder performs the optical-electronic codification and a digital decoder (the NN) is used for reconstruction. The contributions of this thesis are summarized as:

- This research introduces a new codification strategy for compressive imaging systems called time-multiplexed coded aperture (TMCA).
- This research develops new forward models based on the use of TMCAs for three applications: hyperspectral imaging, compressive light field imaging, and SD imaging. In the first case, TMCA emulates an expensive color filter array. In the second, the TMCA is an angular-sensitive CA resulting in less correlations in the coding of angular views, while in the third, the TMCA entails a novel system with a spatially varying response for constant depths in a scene.
- This research learns, in simulation, differentiable optical electronic encoders realizing the TMCAs as well as NN decoders solving the reconstruction problems. The research demonstrates that the learned codifications compare favorably against baselines that use traditional optical codification as well as against TMCA baselines using non-optimized codes.

- This thesis employs a benchtop SD imaging setup to capture a dataset of hyperspectral reflectance images and depth maps. The acquired dataset is used to train the proposed SD model.
- This research builds three prototypes: a compressive light field imaging system, a compressive hyperspectral imager, and a snapshot SD imager. The learned TMCA codes are translated to hardware and real experiments to show that the best results of TMCA obtained in simulation do hold in real-world systems.

In addition to the principal contributions listed above, the imaging ideas and algorithms developed for solving the ill-posed SD imaging problem were employed in joint radar-communications applications. The following section lists all the author's publications.

Publications and Author Contributions

International Journal Papers

1. **Vargas, E.**, Rueda-Chacón, H., & Arguello, H. (2023). Learning Time-multiplexed phase-coded apertures for snapshot spectral-depth imaging. Optics Express, 31(24), 39796-39810.
2. **Vargas, E.**, Mishra, K. V., Jacome, R., Sadler, B. M., & Arguello, H. (2023). Dual-blind deconvolution for overlaid radar-communications systems. IEEE Journal on Selected Areas in Information Theory, 4, 75-93
3. Lopez, J., **Vargas, E.**, & Arguello, H. Depth Estimation from a Single Spectral Optical Encoded Image using a Learned Colored-Coded Aperture. IEEE Transactions on Computational Imaging. (*Submitted*)

4. Jacome, R., **Vargas, E.**, Mishra, K. V., Sadler, B. M., & Arguello, H. (2023). Multi-Antenna Dual-Blind Deconvolution for Joint Radar-Communications via SoMAN Minimization. IEEE Transactions on Signal Processing. (*Submitted*)

5. Goyes-Peñafiel, P., **Vargas, E**., Mäkinen, Y., Foi, A., & Arguello, H. (2023). Volumetric Filtering for Shot Gather Interpolation in Swath Seismic Acquisition. IEEE Geoscience and Remote Sensing Letters, 20, 1-5.

6. Goyes-Peñafiel, P., **Vargas, E.**, Correa, C. V., Sun, Y., Kamilov, U. S., Wohlberg, B., & Arguello, H. (2023). Coordinate-Based Seismic Interpolation in Irregular Land Survey: A Deep Internal Learning Approach. IEEE Transactions on Geoscience and Remote Sensing, 61, *In Press*.

7. Arguello, H., Bacca, J., Kariyawasam, H., **Vargas, E**., Marquez, M., Hettiarachchi, R., ... & Edussooriya, C. U. (2023). Deep optical coding design in computational imaging: a data-driven framework. IEEE Signal Processing Magazine, 40(2), 75-88.

8. Camacho, A., **Vargas, E**., & Arguello, H. (2022). Hyperspectral and multispectral image fusion addressing spectral variability by an augmented linear mixing model. International Journal of Remote Sensing, 43(5), 1577-1608.

9. Goyes-Peñafiel, P., **Vargas, E**., Correa, C. V., Agudelo, W., Wohlberg, B., & Arguello, H. (2021). A consensus equilibrium approach for 3-D land seismic shots recovery. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.

10. Arias, K., **Vargas, E**., Rojas, F., & Arguello, H. (2020). Fusion of Hyperspectral and Multispectral Images Based on a Non-locally Centralized Sparse Model of Abundance Maps. Tecnura, 24(66), 62-75.

11. **Vargas, E**., Arguello, H., & Tourneret, J. Y. (2019). "Spectral image fusion from compressive measurements using spectral unmixing and a sparse representation of abundance maps". in IEEE Transactions on Geoscience and Remote Sensing, 57(7), 5043-5053.

International Conference Papers

1. Monsalve, A., **Vargas, E.**, Mishra, K. V., Sadler, B. M., & Arguello, H. (2023, December). Beurling-Selberg Extremization for Dual-Blind Deconvolution Recovery in Joint Radar-Communications. In 2023 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (Accepted).

2. Jacome, R., **Vargas, E.**, Mishra, K. V., Sadler, B. M., & Arguello, H. (2023, December). Factor Graph Expectation Maximization Dual Blind Deconvolution for Dynamic Integrated Sensing and Communications. In 2023 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (Accepted).

3. Carlsson, M., Martinez, E., **Vargas, E.**, & Arguello, H. Efficient Data Processing for Coded Aperture Snapshot Spectral Imager Systems. In 2023 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (Accepted).

4. Martínez, E., **Vargas, E.**, & Arguello, H. (2022, August). Fast Disparity Estimation from

a Single Compressed Light Field Measurement. In 2022 30th European Signal Processing Conference (EUSIPCO) (pp. 1991-1995). IEEE.

5. Jacome, R., Mishra, K. V., **Vargas, E**., Sadler, B. M., & Arguello, H. (2022, July). Multi-dimensional dual-blind deconvolution approach toward joint radar-communications. In 2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC) (pp. 1-5). IEEE.

6. **Vargas, E**., Mishra, K. V., Jacome, R., Sadler, B. M., & Arguello, H. (2022, May). Joint radar-communications processing from a dual-blind deconvolution perspective. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5622-5626). IEEE.

7. **Vargas, E.**, & Arguello, H. (2021, August). Dictionary Learning of Binary Atoms by using a Smooth Approximation. In 2021 29th European Signal Processing Conference (EUSIPCO) (pp. 2144-2148). IEEE.

8. **Vargas, E**., Martel, J. N., Wetzstein, G., & Arguello, H. (2021). Time-multiplexed coded aperture imaging: Learned coded aperture and pixel exposures for compressive imaging systems. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 2692-2702).

9. Galvis, L., Ramírez, J. M., **Vargas, E**., Villarreal, O., Agudelo, W., & Arguello, H. (2020). Reconstruction of 2D seismic wavefields from nonuniformly sampled sources. Electronic

Imaging, 2020(14), 307-1.

10. Pinilla, S., Correa, C. V., **Vargas, E**., Agudelo, W., & Arguello, H. (2019, December). Salient motion detection for spectral video on the compressive domain. In 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (pp. 106-110). IEEE.

11. López, K. L., **Vargas, E**., Morales, F. R., & Arguello, H. (2019, December). Higher-Order Tensor Sparse Representation for Video-Rate Coded Aperture Snapshot Spectral Image Reconstruction. In 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (pp. 704-708). IEEE.

12. Hinojosa, C., Bacca, J., **Vargas, E**., Castillo, S., & Arguello, H. (2019, October). Single-pixel camera sensing matrix design for hierarchical compressed spectral clustering. In 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.

13. Bacca, J., Correa, C. V., **Vargas, E**., Castillo, S., & Arguello, H. (2019, October). Compressive classification from single pixel measurements via deep learning. In 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.

14. Arias, K., **Vargas, E**., & Arguello, H. (2019, September). Hyperspectral and multispectral image fusion based on a non-locally centralized sparse model and adaptive spatial-spectral

dictionaries. In 2019 27th European Signal Processing Conference (EUSIPCO) (pp. 1-5). IEEE.

General Objective

To design and implement a coded snapshot computational imaging system for spectral-depth estimation by using an end-to-end optimization of the sensing and the recovery processes.

Specific objectives

- To acquire a spectral and depth database in the optical laboratory to be used in the end-to-end optimization of the computational system.
- To design an optical system that simultaneously encode the spectral-depth dimensions and to develop a differentiable mathematical model that describes the light wave propagation through its optical elements.
- To implement a differentiable decoder for the recovery process such that the spectral-depth information can be estimated from the encoded database and acquired in the laboratory.
- To design the optical encoding elements and the decoder algorithm by using an end-to-end optimization of the encoder-decoder system.
- To implement the optimized snapshot coded spectral-depth imaging system in the optical laboratory and estimate the corresponding calibration parameters to adjust the computational system model.
- To evaluate the performance of the designed snapshot computational coded spectral-depth imaging system by comparing it with higher precision sensors.

1. Theoretical Background

1.1. General Notation

Throughout this thesis, $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{R}^D$ denotes the D-dimensional linear space. Also, the lowercase boldface letters denote vectors, such as $\mathbf{x} \in \mathbb{R}^D$, and uppercase boldface letters denotes matrices, such as $\mathbf{X} \in \mathbb{R}^{D\times N}$. The transpose of the matrix $\mathbf{X} \in \mathbb{R}^{D\times N}$ is denoted as $\mathbf{X}^T \in \mathbb{R}^{N\times D}$. The uppercase calligraphic letters denote sets, such as $\mathscr{X} \subseteq \mathbb{R}^D$. For any vector $\mathbf{x} = [x_1, \cdots, x_D] \in \mathbb{R}^D$ and $p \geq 1$, the ℓ_p-norm is defined as $\|\mathbf{x}\|_p = \left(\sum_{i=1}^D |x_i|^p\right)^{1/p}$. As p approaches infinity, the infinity norm is defined as $\|\mathbf{x}\|_\infty = \max_{i=1}^D |x_i|$. Another particular interesting case for this thesis is when $p = 0$, in such case, $\|\mathbf{x}\|_0$ is defined as the number of nonzero entries in $\mathbf{x}$.

1.2. Spectral Imaging

Spectral imaging sensors collect data that can be represented by a three-dimensional data cube $I(x, y, \lambda)$ with x and y as spatial coordinates and λ the wavelengths (Hsu and Burke, 2003). This data cube referred to as spectral image can be seen as a collection of 2D images, where each 2D image is a spatial image at a specific wavelength. The spectrum of a spatial point in a scene is represented by the distribution of its electromagnetic radiation over a range of wavelengths. Since many applications can be developed from detailed spectra, several acquisition systems for precise spectral measurements have been studied for decades. These include applications in the fields of remote sensing (Schaepman et al., 2009), medical imaging (Panasyuk et al., 2007; Backman et al., 2000), geology (Delalieux et al., 2009), biological science (Kester et al., 2011), scientific observation and many other fields (Gowen et al., 2007). Traditional spectral imaging systems use scanning

methodologies in the spectral and spatial domain, making it difficult to capture dynamic scenes. The traditional sampling methods for HS imaging are based on a scanning sequence of 1D or 2D sources, which are then concatenated into a single data cube. For instance, push-broom spectral imaging sensors (Green et al., 1998) capture a spectral cube with one focal plane array (FPA) measurement per spatial line of the scene (Arce et al., 2014), and whereas whisk-broom acquires a single spectral pixel at a time (Herrala et al., 1994), and tunable filter imagers (Morris et al., 1994) measure 2D images at a specific wavelength. Figure 1 illustrates three scanning methodologies corresponding to a point, line, and wavelength scanning methods. In all the traditional methods for

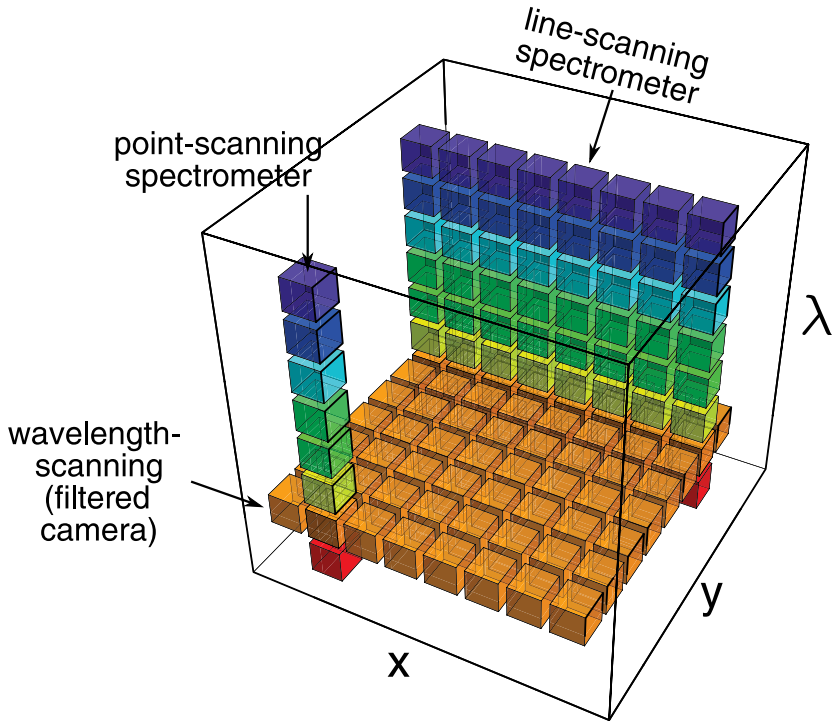

Figure 1. Scanning spectral imaging approaches for a spectral data cube with x and y as spatial coordinates and λ the wavelengths (Hagen and Kudenov, 2013).

HS data acquisition, due to technological reasons, there is a trade-off between spectral and spatial

resolution. Generally, HS images benefit from excellent spectroscopic properties with several or hundreds of thousands of contiguous bands but are limited by their relatively low spatial resolution (Shaw and Burke, 2003a). For instance, the Hyperion imaging spectrometer yields HS images with about 220 spectral bands, which extend from the visible region (0.4 to 0.7 μm) through the SWIR (about 2.5 μm), with a spatial resolution of 30 m (Middleton et al., 2013). Furthermore, due to the scanning acquisition methodology, these technologies suffer from two main problems: motion artifacts and low light collection efficiency. New non-scanning alternatives that could capture the full 3D data cube in a single measurement have been proposed to overcome these problems. These techniques are referred to as coded snapshot spectral imaging which will be presented in the following section.

1.3. Coded Snapshot Spectral Imaging

Scanning technologies for spectral imaging suffer from two main problems: motion artifact and poor light collection efficiency (Hagen and Kudenov, 2013). These have been long-standing problems that nowadays remain. The first techniques to overcome these problems were proposed in astronomy based on image slicing, fiber reformatting, and lenslet-based pupil array known as integral field spectroscopy (ISP) (Bacon et al., 1995). The latter are snapshot techniques that could capture the underlying spectral image in a single measurement time. Then, with the advances in computer capacity, new computational sensing systems were proposed. One of the first ideas dubbed as Computed Tomography Imaging Spectrometry (CTIS) (Okamoto and Yamaguchi, 1991) used a 2-D disperser to project a spectrally dispersed scene into a high-resolution CCD detector in such a way the spatial and spectral information are multiplexed. Since the detector array mea-

surements can be seen as tomographic projections of the 3D source, tomographic reconstruction techniques were employed to estimate the 3D spectral image. Motivated by the development of compressed sensing theory (CS), more recent approaches to acquire spectral data have been proposed. CS relies on two principles: sparsity and incoherence (Candès and Wakin, 2008). Sparsity is related to the signals of interest, and CSI exploits the fact that hyperspectral images can be sparsely represented on a proper basis Ψ. More formally, if we consider a hyperspectral image represented as a vector $\mathbf{f} \in \mathbb{R}^{N^2L}$, this can be expressed as $\mathbf{f} = \Psi\theta$, where θ is an S-sparse vector with $S \ll N^2L$. Incoherence relates to the sensing modality and expresses the idea that sparse vectors in Ψ must be spread out in the domain in which they are acquired (Candès and Wakin, 2008). There are mainly two approaches for compressing spectral data that have been implemented in practical applications: *i)* spatial coding based CS imagers, such as the coded aperture snapshot spectral imager (CASSI) (Wagadarikar et al., 2008), and *ii)* spatial and spectral coding-based spectral imagers, such as the spatio-spectral encoded compressive spectral imager (SSCSI) (Correa et al., 2015), or the Colored CASSI (C-CASSI) (Arguello and Arce, 2014). Note that these architectures have been implemented in practical applications (based on spatial and spectral coding). The coded aperture snapshot spectral imager (CASSI) is one of the most representative CSI architectures, which comprises a dispersive element and a coded aperture (Wagadarikar et al., 2008). The coded aperture is the spatial coding optical element defined as a block-unblock lithographic mask or a spatial light modulator (SLM) (Arce et al., 2014). The CASSI system is very similar to the CTI system. They differ mainly in the binary coding mask and the recovery algorithm. The CTI system employs orthogonal binary codes and a demultiplexing algorithm, while the CASSI system employs a random

binary pattern and a more complex reconstruction algorithm. The CASSI architecture encodes the 3D data only in the spatial domain, i.e., each pixel of the coded aperture blocks or let pass the entire spectral information. The colored CASSI (C-CASSI) is a different variation of the CASSI system, which replaces the binary masks by multiple-patterned arrays of selectable optical filters or colored-coded apertures to provide a richer modulation in both spatial and spectral domains (Arguello and Arce, 2014). In Figure 2, an instance of a colored-coded aperture pattern is shown.

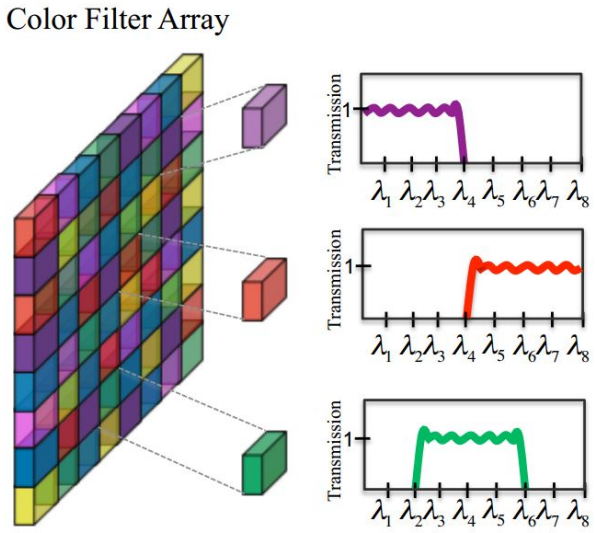

Figure 2. Colored coded aperture. Each pixel on the coded aperture is an optical filter whose spectral response can be selected. Source (Correa et al., 2015).

Each pixel on the colored-coded aperture is one of the possible optical filters whose spectral response can be selected. C-CASSI reduces the number of 2D measurements required to recover the underlying image because of the higher randomness of these 3D coded aperture structures. Further, based on properties of compressed sensing such as mutual coherence or the restricted isometry property the codification of the C-CASSI system has been designed in such that the

quality of the recovered spectral image is improved. The sensing representation of the C-CASSI system is depicted in Fig 3. Finally, another alternative for spectral estimation consists of a com-

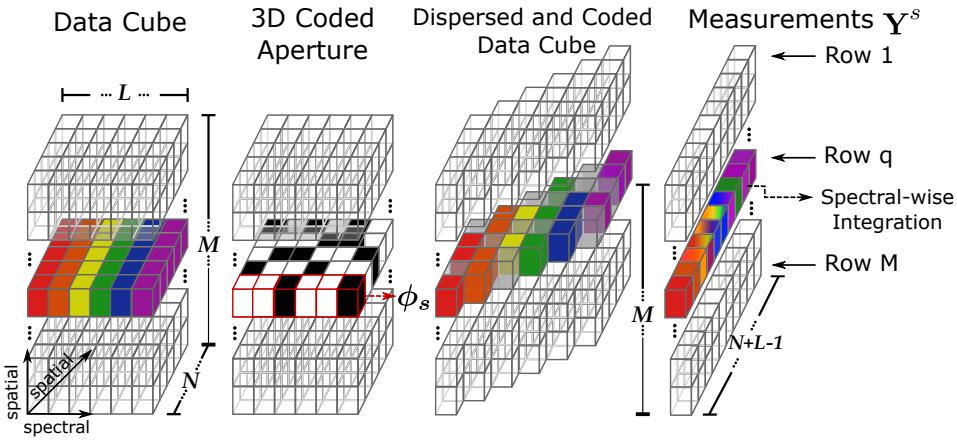

Figure 3. Schematic representation of the sensing phenomena behind CASSI system.

pact snapshot computational system based on diffractive imaging (Jeon et al., 2019). In (Jeon et al., 2019), the authors propose a diffractive optical element (DOE) that replaces complex optical elements such as a prism o a coded aperture mask to obtain a minimal form factor system. They take advantage of the wavelength-dependent property of Fresnel diffraction to design a DOE with a spectrally-varying point spread function (PSF) with an anisotropic shape. The phase codification of the diffractive optical element is fully determined by the height profile of the DOE. Since the spectral cues come only from a diffractive optical element, the ill-posed problem of recovering the spectral image becomes more challenging. Besides, they propose an end-to-end network that learns a spatial-spectral prior, designed explicitly for the deconvolution of the proposed point spread function. It is important to note that the optical system design is entirely analytical, and the end-to-end optimization networks only consider the recovering process.

1.4. Depth Imaging

Imaging systems with depth estimation capabilities are essential for many computer vision applications such as autonomous navigation, action recognition, or robot guidance (Biswas and Veloso, 2012; Jalal et al., 2019; Wen et al., 2017). Several standard solution approaches usually employ sophisticated camera systems such as a stereo camera or a camera array (Jeon et al., 2015), pulsed LiDAR (Altmann et al., 2016), time of flight (Zhu et al., 2008), or structured light (Zhang et al., 2002). The following figure shows a schematic representation of a single-photon LiDAR system to acquire depth information in several multispectral bands.

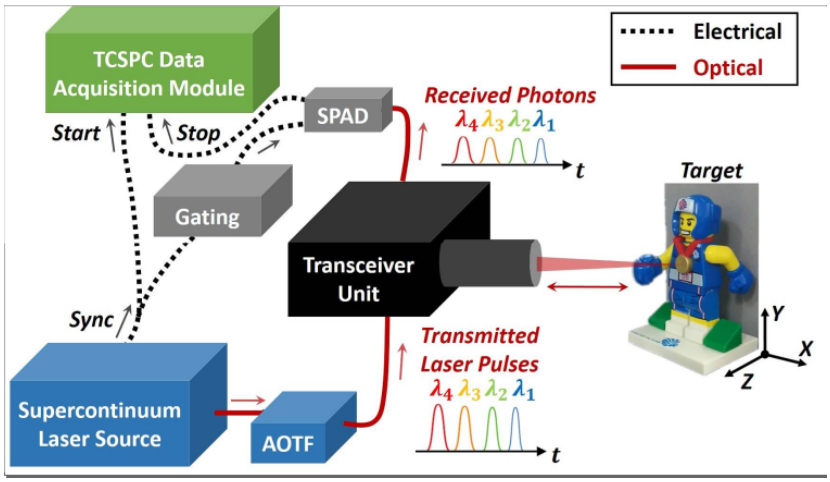

Figure 4. Schematic diagram of the multispectral depth imaging system based on single-photon LIDAR (Altmann et al., 2016).

The systems mentioned above are expensive, bulky, require sophisticated electronics, and are not compatible with custom sensors. For that reason, the challenge of depth estimation from a single image with conventional cameras has been an active area of research (Godard et al., 2017; Fu et al., 2018; Chang and Wetzstein, 2019). Even though a single image lacks the depth cues that

exist in a stereo image pair, traditional approaches leverage predefined image features that are statistically correlated with depth, e.g., shading, perspective distortions, occlusions, texture gradients, and haze (Hoiem et al., 2007; Henderson and Ferrari, 2019; Paschalidou et al., 2019). Besides, depth cues have been estimated using learned features from complex convolutional neural networks (CNNs), improving the performance of approaches employing hand-crafted features (Eigen et al., 2014; Fu et al., 2018). Figure 5 depicts the model architecture of a neural network for the estimation of high-detailed depth maps (Eigen et al., 2014). This architecture contains two main components: a coarse-scale network first predicts the depth map globally, and a fine-scale network refines the predicted global depth map within local regions. Note that the two stacks are applied to the original input; however, the fine-scale network also contains the coarse network's output in the first layer image features. Although the results of these approaches are consistent within similar images in the training dataset, they do not generalize well for different datasets since the problem of depth estimation with pictorial depth cues or RGB images alone is extremely ill-posed.

1.5. Coded Depth Imaging

In contrast to depth estimation from a single image, alternative approaches employ optical coding to improve the depth estimates. For instance, imaging systems that employ an amplitude or phase coding mask have been proposed in the last decade to optically encode the depth-dependence of the scene (Levin et al., 2007; Haim et al., 2018; Chang and Wetzstein, 2019). It has been shown that coded defocus blur or chromatic aberrations have the potential to generate strong depth cues and help estimate an accurate depth map. In (Levin et al., 2007), the authors proposed a variation of a conventional camera where a pattern code within the aperture of the camera lens modulates

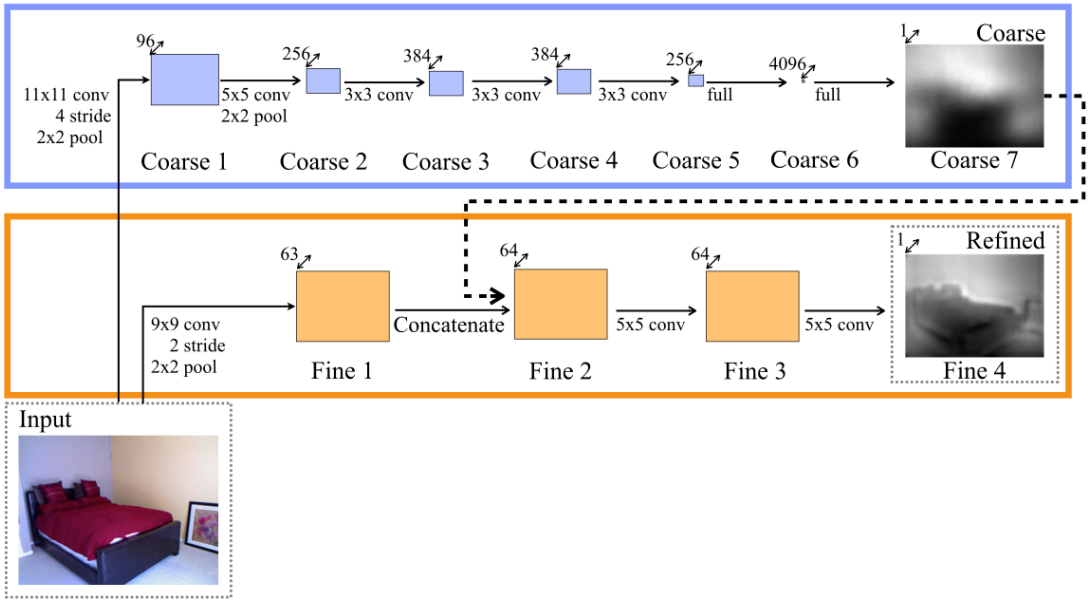

Figure 5. Schematic representation of the neural network architecture for monocular depth estimation (Eigen et al., 2014).

the incident light to generate a coded image rather than the final desired image. It allows the simultaneous extraction of reliable information such as an artifact-free image and a depth map estimate. In (Lin et al., 2013), coded focal stack photography is proposed for high-resolution 3D imaging and extended depth of field. This technology consists of a 3D volume composed of a sequence of images captured under different focal settings. This is achieved by sweeping the focal plane over a range of depths within the exposure time of a single image while selectively modulating the readout patterns of each pixel (Lin et al., 2013). Since this approach is based on compressive sensing, the employed reconstruction strategy is a sparse coding algorithm. Chang et al. (Chang and Wetzstein, 2019) propose an end-to-end design of the system's optical elements and the recovery algorithm for the monocular depth estimation problem, using coded defocus blur.

Specifically, the codification is made by a free-form lens that is jointly estimated with a CNN that recovers the depth map estimate. Similar to (Chang and Wetzstein, 2019), in (Haim et al., 2018) and (Wu et al., 2019) is exploited the emerging computational paradigm, where the optics and the computational algorithm are co-designed to maximize performance and accuracy. In both cases, the optical systems consist of a phase coded aperture with high light throughput. Common to all these approaches is the use of a coded image in order to promote optical cues. Meanwhile, these optical cues are obtained by having an optical system with a depth-dependent PSF. Figure 6 depicts a set of different PSFs for different depths for a system with chromatic aberrations.

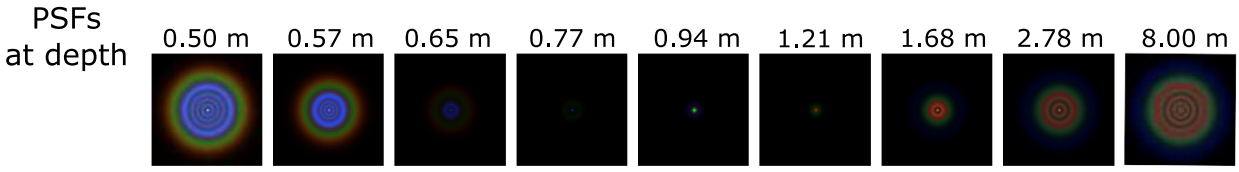

Figure 6. Point spread functions for different depths in a system with chromatic aberrations. The difference between PSFs allows the optical encoding of depth (Chang and Wetzstein, 2019).

On the other hand, there are active systems for depth estimation based on time of flight (ToF) technologies that have been widely applied since they obtain an excellent performance in terms of accuracy for different ranges. The basic operation principle of depth estimation employing ToF is based on measuring the time of the round-trip of a pulse of light. ToF sensors are able to produce a full depth frame simultaneously, therefore allowing their application to dynamic scenes. There exist a plethora of computational imaging systems based on ToF technology. For simplicity, to name just one, a single-pixel 3D imaging system using pulsed structured illumination and a high-speed photodiode were proposed for sampling the time of flight of the pulsed laser source

(Sun et al., 2016). The time-varying structured light illumination is implemented through a digital micromirror device (DMD). To recover the depth estimates, they employed a strategy based on compressive sensing achieving accurate depth maps and real-time high-resolution videos with a frame rate of 12 Hz. It is important to note that the optical codification coming from the DMD is set up to be Hadamard matrices, and the system is not further optimized. Figure 7 shows a summary of the passive and active technologies for depth codification.

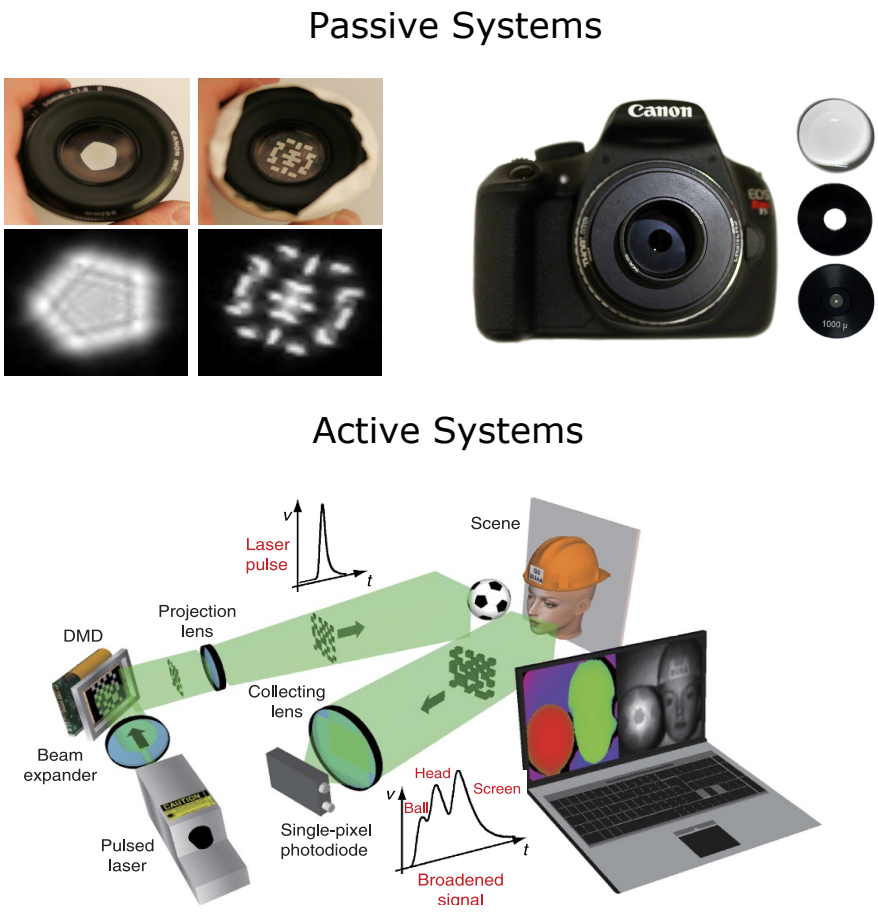

Figure 7. Top: Passive depth systems. (left) Conventional camera with a coded aperture (Levin et al., 2007) (right) Optimized free-form lens for depth estimation using an end-to-end approach (Chang and Wetzstein, 2019). Bottom: 3D single-pixel imaging using a time-varying structured light illumination (Sun et al., 2016). These images have been adapted from their original sources.

1.6. Coded Spectral-Depth Imaging

Multidimensional optical imaging has grown rapidly in the last years as a result of the development of three major technical advances: High-resolution focal plane arrays, developments of new computational techniques, and the development of high-precision microfabrication techniques (Gao and Wang, 2016). In conventional imaging techniques, only the two-dimensional spatial distribution of light is measured. However, optical imaging aims to capture the light even in up to nine dimensions, corresponding to the information of spatial coordinates, emittance angles, wavelength, time, and polarization of a single photon (Gao and Wang, 2016). In this field, efforts have been made to propose high dimensional imaging systems that can simultaneously acquire spectral and depth information. Pioneer works in the joint imaging of depth and spectrum include the combination of complex imaging systems that independently capture spectral and depth information with separate hardware and combine or render them in a computational stage. For instance, Kim et al., introduced a combined depth and spectral approach for measuring 3D patterns on solid objects (Kim et al., 2012). Other examples include a cross-modal stereo system (Wang et al., 2016) and, a 5D hyperspectral imaging system (Heist et al., 2018) using structured light. In general, these approaches employing two different imaging modalities to estimate spectral and depth significantly increase the device size. Furthermore, depending on its complexity, the depth cues' alignment and registration process across different spectral bands can be a very challenging problem in terms of calibration. Based on the advances in deep learning for the solution of inverse problems and the reconstruction of high-fidelity images, more compact computational systems

have been proposed for the acquisition of high temporal resolution spectral-depth images. In (Yao et al., 2019), they proposed a stereo-based computational imaging system that consists of RGB and spectral commercial cameras and a deep learning approach for the computational reconstruction of the temporal spectral-depth image field. The proposed system can achieve a spatial resolution of 1920×1024, 16 spectral bands, and a depth map with an accuracy of 2mm at distances around 0.6 m. Specifically, the system employs a convolutional neural network (CNN) to superresolve the acquired low-resolution spectral image and a deep learning network named as EnhaceNet for a texture enhancement guided by the complementary spatial information of the RGB image. The processing through EnhaceNet is the most important step in the reconstruction algorithm. Figure 8 briefly summarizes the proposed system and the reconstruction algorithm.

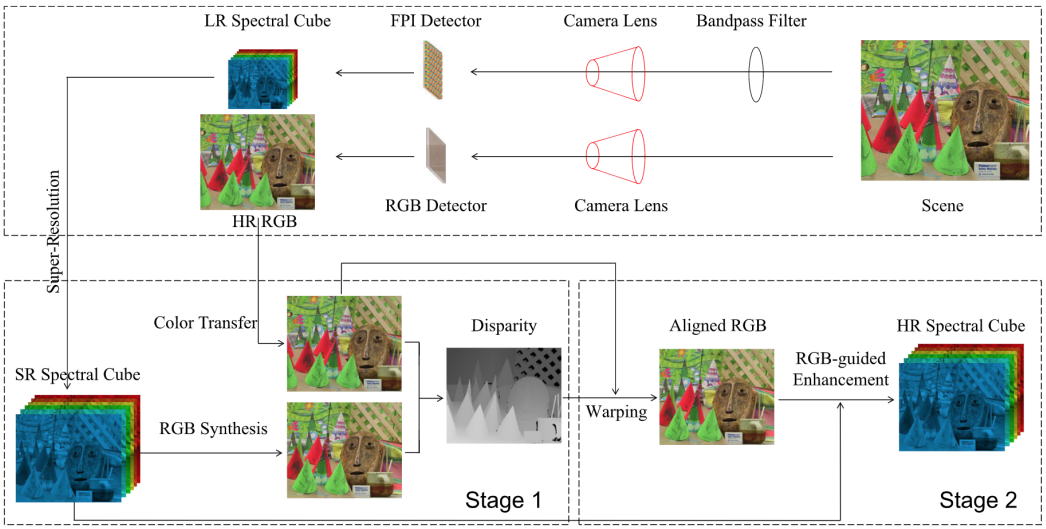

Figure 8. Sketch of the temporal spectral-depth computational imaging system with the reconstruction algorithm. The stages of superresolution and RGB-guided texture enhancement are performed using deep learning approaches. Source image: (Yao et al., 2019)

On the other hand, under a compressive multidimensional framework (Horisaki et al.,

2013), CASSI systems have also been combined with light field or time of flight (TOF) imaging to propose a more compact snapshot monocular imaging system for spectral-depth estimation(Feng et al., 2016; Rueda-Chacon et al., 2019). Specifically, in (Feng et al., 2016), they employ an integral imaging technique to modify the CASSI system to obtain spectral-depth information. Integral imaging is a technique that consists of acquiring multiple images from different perspectives (Horisaki et al., 2013). The integration of the integral imaging technique with the spectral imager CASSI is achieved by employing a microlens array. The reconstruction process consists of recovering a stack of images corresponding to depth slices that contain clear features if they are in the correct depth plane. This procedure is similar to obtain a focal stack for spectral images. In (Rueda-Chacon et al., 2019), the authors propose a snapshot compressive multispectral depth imaging system based on colored-coded aperture codification integrated with an off-the-shelf ToF sensor. Further, they optimize the colored coded aperture employing a spectral blue noise shaping technique improving the quality of the recovered spectral-depth scene. It is worth noting that in this work, they optimized the optical encoder employing a hand-crafted prior. However, they do not consider the image processing algorithm in the optimization. Fig. 9 shows a sketch of the proposed snapshot computational system in (Rueda-Chacon et al., 2019).

1.7. End-to-end Optimization of Optics and Image Processing Algorithms

Deep optics is a very recent framework that introduces the concept of optimizing the optical elements jointly with the reconstruction algorithm in an end-to-end fashion employing stochastic optimization. Based on differentiable reconstruction algorithms, this framework allows for the optimization of domain-specific computational cameras. Two key components of this framework are

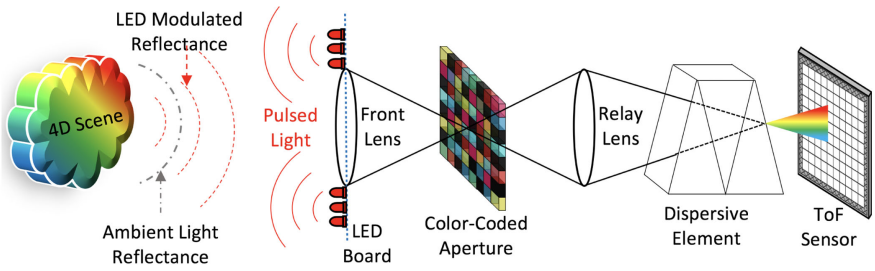

Figure 9. Schematic representation of the snapshot compressive multispectral depth imaging system based on the CASSI system and ToF sensor. Source image: (Rueda-Chacon et al., 2019)

wave optics or parameterized image formation model that contains object depth and wavelength-dependent effect as well as a differentiable reconstruction algorithm. In general, this idea can be integrated with several low-level and high-level image processing algorithms. This framework has been recently investigated on applications such as extended depth of field (Sitzmann et al., 2018), high dynamic range imaging (Metzler et al., 2020), spectral imaging (Wang et al., 2018), depth estimation (Chang and Wetzstein, 2019) and many others. More in detail in (Sitzmann et al., 2018), they demonstrate the efficacy of this approach by applying it to an achromatic extended depth of field. In this case, the optical system is a free-form lens whose impulse response is modeled using wave optics. In fact, this response is characterized by the height profile of the optimized free-form lens. This allows designing a very compact system that can be implemented with common optics. The image processing algorithm is a simple differentiable Wiener filtering producing excellent results; however, more complex and sophisticated algorithms can be employed at the expense of greater computational resources. Similarly, in (Metzler et al., 2020; Chang and Wetzstein, 2019), this framework is employed for high dynamic range imaging and depth estimation, where the optical system is also characterized by wave optics. In contrast to (Sitzmann et al., 2018)

that employ a simpler image processing algorithm, in (Metzler et al., 2020; Chang and Wetzstein, 2019), they employed a complex and task-specific deep neural network as the image processing algorithm for high dynamic range and depth estimation, respectively. In (Wang et al., 2018), an end-to-end optimization framework is also employed to reconstruct spectral images employing CASSI. In particular, the optimization parameters of the optical system are the distribution of the coded apertures in the CASSI system. A CNN architecture named HyperReconNet is proposed for the joint optimization of the coded apertures and the recovery algorithm. Considering the limits of computation for CNNs, the authors proposed a patch modeling approach and a repeated pattern of the coded aperture for making the training of the deep neural network feasible. Finally, to exploit the spatial and spectral intrinsic correlation of spectral images, the authors designed a spatial and spectral network separately which are finally concatenated to obtain the estimated image.

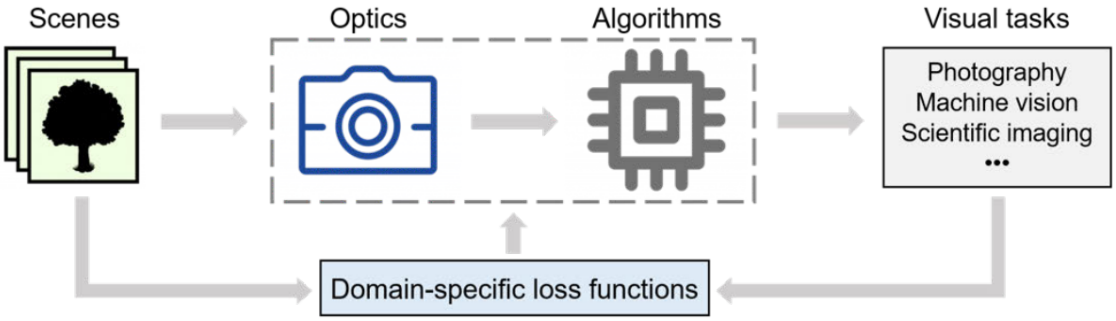

Figure 10. Schematic representation of the end-to-end optimization of optics and recovery algorithms for domain-specific tasks (Eigen et al., 2014).

To describe more formally the end-to-end framework, denote the variable of interest to be estimated, such as the irradiance of the all-in-focus image, a depth map, or the spectral image as $\mathbf{f}$. Further, denote the coded optical system by the encoding operator $\mathscr{M}_{\varphi}$, where we assume that it can be fully described by a set of parameters φ. Thus, the computation of the measurements of a

coded computational imaging system can be simply represented as $\mathbf{y} = \mathscr{M}_{\varphi}(\mathbf{f})$. To estimate $\mathbf{f}$ from the coded measurements $\mathbf{y}$ of a given computational coded imaging system, a differentiable reconstruction algorithm or decoder $\mathscr{D}_{\psi}$ is used. Similarly to the encoder, we assume that the decoder can be fully defined by the set of parameters ψ. Given a pre-trained encoder−decoder, one would capture measurements with the corresponding optical system $\mathbf{y} = \mathscr{M}_{\varphi}(\mathbf{f})$ and then estimate the variable of interest via $\tilde{\mathbf{f}} = \mathscr{D}_{\psi}(\mathbf{f})$. The recovery process is commonly referred to as the inference stage. To determine an optimal set of parameters φ, ψ, the optical-encoder, and electronic-decoder are trained jointly in an end-to-end fashion with some dataset containing P ground truth irradiance, spectral images or depth maps by minimizing an appropriate loss function $\mathscr{L}$ with some variant of stochastic gradient descent

$$\underset{\phi,\psi}{\operatorname{argmin}} \sum_{t=0}^{P-1} \mathscr{L}\left(\mathscr{D}_{\psi} \circ \mathscr{M}_{\varphi}(\mathbf{f}^t), \tilde{\mathbf{f}}^t\right). \tag{1}$$

where $\circ$ denotes the composition of functions. This is a self-supervised learning scenario in which the encoder implements a physically feasible coding function and the decoder recovers the target. Note also that both encoder and decoder are jointly optimized in this case, so the optical codes, parameterized by φ, influence the reconstruction algorithm ψ and vice versa. Moreover, the optimal choice of both encoder and decoder depends on the loss function, which could be a low-level image metric, such as mean-squared error (MSE), or a higher-level loss function such as image classification accuracy.

2. Time Multiplexed Coded Aperture (TMCA)

The principle of the proposed TMCA consists of synchronizing a time-varying CA with a shutter function performing a coded exposure. This idea can be applied to different systems depending on the specificities of the optical setup it is coupled with, i.e, the optical elements in between the CA and the sensor as shown in Figure 11. The CA and shutter functions are jointly optimized with a digital decoder for each application in simulation, considering the specific constraints of the spatial light modulators realizing the CA and of the sensor realizing the coded exposures. After training, the optimized TMCA can be deployed to physical devices (prototypes)

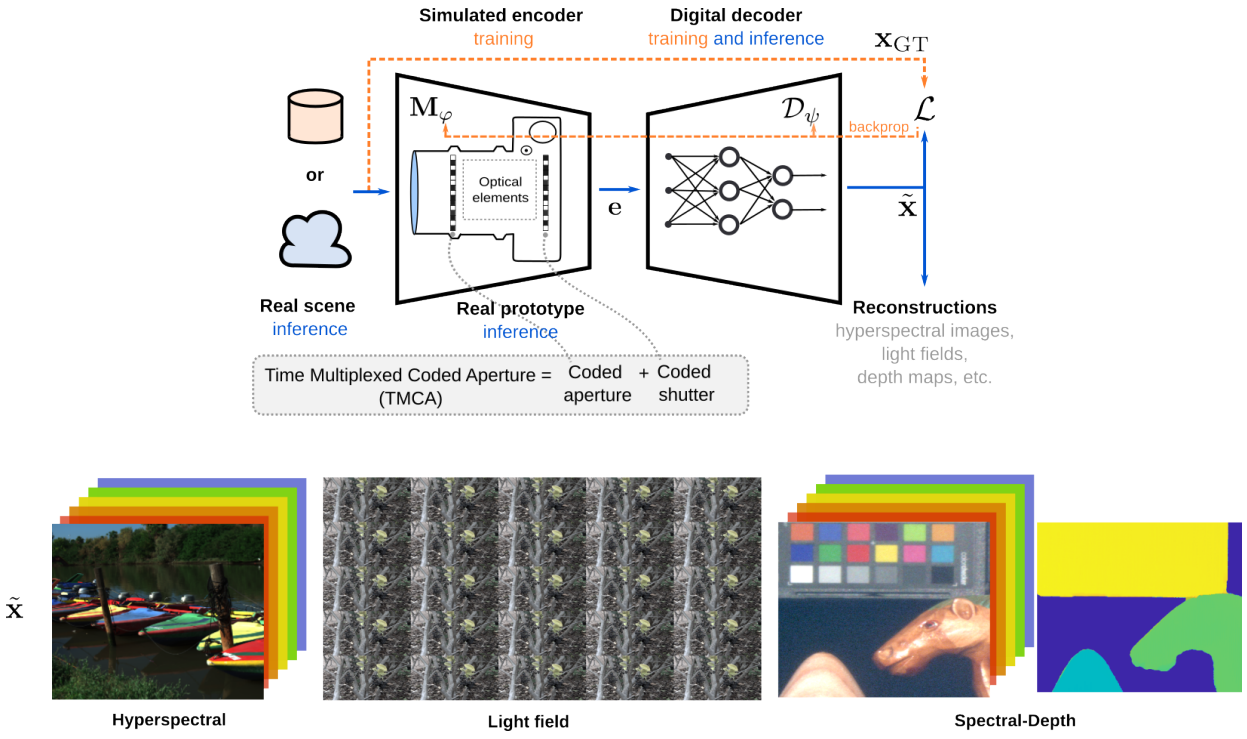

Figure 11. An illustration of the proposed Time Multiplexed Coded Aperture (TMCA) codification in the proposed end-to-end differentiable pipeline. We also show reconstruction examples for our hyperspectral, compressive light field, and spectral-depth imaging applications.

that capture real-world coded scenes.

The rest of this chapter presents the general forward model and codification induced by a TMCA. Then, the following chapters describe how this model can be used for three specific applications using different optical setups: in hyperspectral, light-field, and spectral-depth imaging. Particularly, this dissertation derives the codification for those three systems and demonstrates how the proposed TMCA improves the traditional CAs they use.

Generalities

We consider the irradiance f, invariant in time as the quantity we reconstruct from compressive imaging systems. Our proposed TMCA consists of two coding stages. The first stage optically encodes f into a field g incident to the sensor, using a time-varying CA $T(t)$. We model $g(t)$ as the response of a linear optical system $\mathscr{O}$ given input f:

$$g(t) = \mathscr{O}(T(t), f). \tag{2}$$

As we shall see in chapters 3, 4 and 5, the exact form of $\mathscr{O}$ is application dependent but is always a function of the time-varying coded aperture $T(t)$.

The second stage consists of a coded exposure. The irradiance $g(t)$ is captured by the sensor in a single snapshot of exposure time Δt. During this exposure, a spatially-varying shutter function $S_{i,j}(t)$ turns the pixel (i, j) “on" and “off" multiple times, resulting in the coded exposure $e_{i,j}(t)$ in

which the integration of $g(t)$ has been modulated:

$$e_{i,j}(t) = \int_{t}^{t+\Delta t} S_{i,j}(t')g_{i,j}(t')\,\mathrm{d}t'. \tag{3}$$

In particular, we consider binary shutter functions $S_{i,j}$ defined on K discrete time slots (sub-exposures of time δt): we have $\Delta t = K \cdot \delta t$. The coded exposure can then be rewritten as

$$e_{i,j} = \sum_{k=0}^{K-1} S_{i,j}^{k} g_{i,j}^{k}, \text{with } S_{i,j}^{k} \in \{0,1\}^{k}, \tag{4}$$

where $g_{i,j}^{k}$ and $S_{i,j}^{k}$ denote the irradiance incident on the sensor and shutter function at pixel (i,j) in the $k-$th time slot. Those slots allow us to synchronize the coded apertures with the coded exposures (both are indexed by k).

Another way to write this discrete model is using the matrix-vector notation $\mathbf{e} = \sum_k \boldsymbol{S}^k \mathbf{g}^k$, where $\mathbf{e}$ and $\mathbf{g}$ are the "vectorized" form of the exposure and coded irradiance and $\boldsymbol{S}$ is the measurement matrix representing the shutter function. The irradiance incident to the sensor is $\mathbf{g}^k = \boldsymbol{O}^k \mathbf{f}$, where the matrix $\boldsymbol{O}^k$ is the point spread function of the application-dependent optical system, also including the coded aperture, and $\mathbf{f}$ represents the irradiance f in its vector form. Using this notation, the forward model of the proposed TMCA can be simply written as

$$\mathbf{e} = \sum_{k=0}^{K-1} \boldsymbol{S}^k \boldsymbol{O}^k \mathbf{f} = \mathbf{M}\mathbf{f}, \tag{5}$$

defining the overall measurement matrix of our compressive imaging system $\mathbf{M} = \sum_{k=0}^{K-1} \boldsymbol{S}^k \boldsymbol{O}^k$.

Recovering $\mathbf{f}$ given the coded snapshot $\mathbf{e}$ amounts to solving an inverse problem.

In the following chapters, we employ this general model for three different applications. In our first

application of compressive light field imaging, we aim at recovering light fields and the irradiance f (and thus $\mathbf{f}$) considers multiple view angles. Our second application targets hyper-spectral imaging. In that case, f (and $\mathbf{f}$) considers multiple frequency bands. Our last application is the main goal of this thesis: spectral-depth imaging. In this case, f jointly considers multiple frequency bands and multiple depth maps.

3. TMCA for Compressive Light Field Imaging

This chapter presents how the proposed TMCA improves the performance in a compressive light field application. Light fields are a rich source of information about the geometry of the scene making them useful for depth estimation (Jeon et al., 2018; Wang et al., 2015; Martínez et al., 2022). Particularly, this chapter considers the same optical setup proposed for compressive light field imaging by Marwah et al. (Marwah et al., 2013) as shown in Fig. 12. A coded aperture mask T is placed between the objective lens and the sensor, at a distance d_l from the latter. The incident field $g(x,t)$ at the sensor is the spatially modulated light field projected along its angular dimension taken over the aperture area $\mathscr{V}$:

$$g(x) = \int_{\mathscr{V}} l(x,u)\, T\big(x + s(u-x)\big)\, \mathrm{d}u, \tag{6}$$

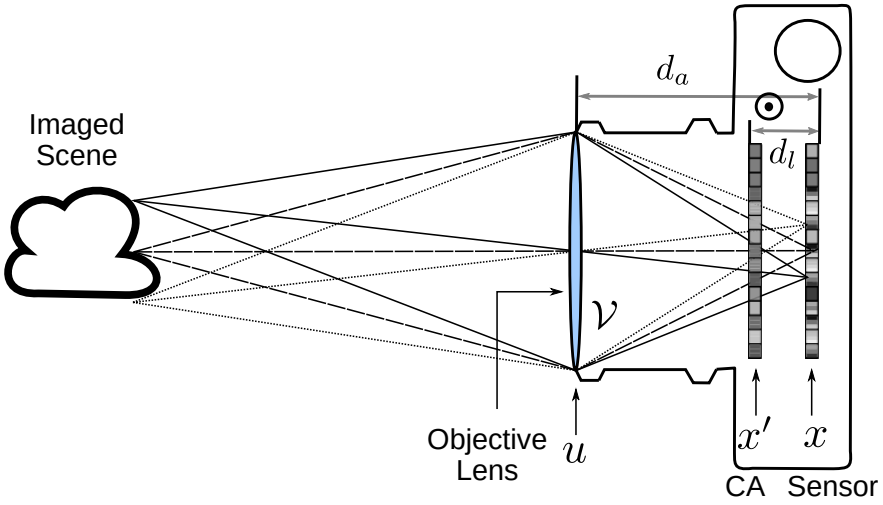

Figure 12. A diagram of the ray optics for compressive light field imaging using TMCAs.

where $s = d_l/d_a$, with d_a the distance from the sensor to the aperture plane, is the shear of the mask pattern with respect to the incident light field $l(x,u)$. Similar to (Levoy and Hanrahan, 1996), this thesis adopts a two-plane parameterization for the light field where x is the 2D spatial dimension on the sensor plane and u denotes the 2D position on the aperture plane (see Fig. 12). Now considering the time-varying coded aperture T and the shutter function S, introduced in the previous chapter, the measurement model yielding the exposure e is

$$e(x) = \int_{\mathscr{V}} \int_{\Delta t} S(x,t')\, l(x,u)\, T(x+s(u-x),t')\, \mathrm{d}u\, \mathrm{d}t'. \tag{7}$$

By defining the TMCA $\hat{T}$ as

$$\hat{T}(x,u) = \int_{\Delta t} S(x,t')\, T(x+s(u-x),t')\, \mathrm{d}t', \tag{8}$$

the model in Equation (7) can be simply rewritten as

$$e(x) = \int_{\mathscr{V}} l(x,u)\, \hat{T}(x,u)\, \mathrm{d}u. \tag{9}$$

The model from (Marwah et al., 2013) described in Equation (6) does not include the coded exposure. Comparing the latter with the proposed TMCA model in Equation (8), it is worth noting that the proposed model can be seen as an equivalent coded aperture denoted as $\hat{T}$. Furthermore, using a change of variable, this thesis shows that the proposed TMCA for compressive light field

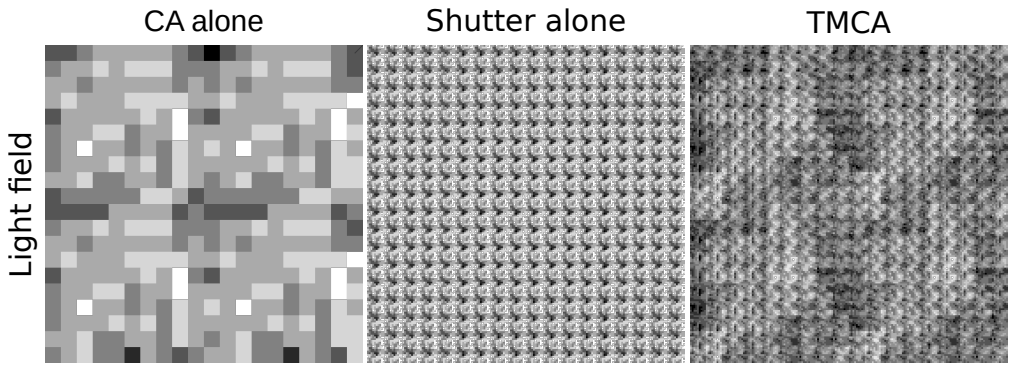

Figure 13. Illustrations of TMCA codification for light field imaging with $K = 8$. Right: full codification as in Eq. (8). Left and middle: codifications when $\boldsymbol{S} = \mathrm{Id}$. (left) and $\boldsymbol{T} = \mathrm{Id}$. (middle). Even though the codes are binary, $K = 8$ yields 2^8 possible shutter / CA values.

imaging can be expressed in the coded aperture plane as

$$\hat{T}(x',u) = \int_{\Delta t} S(x' + \hat{s}(u - x'), t') T(x', t') \, \mathrm{d}t', \tag{10}$$

using the spatial coordinates x' in the coded aperture plane (see Fig.12) and with $\hat{s} = d_\ell / (d_a - d_\ell)$.

Interestingly, (10) shows that using the proposed TMCA, the equivalent CA's pixels can be seen as responding differently to rays coming from different angles. If the shutter function is removed and the CA remains constant in time, then the TMCA reduces to (Marwah et al., 2013), where all the pixels respond equally for all angles. When exclusively analyzing the shutter function, without the CA codification, the time-modulation in the sensor plane averages all the views, which does not yield a clear coding advantage if the scene is static in time (since all the views would be the same). Figure 13 shows the coding elements distribution in the coded aperture plane for the aforementioned configurations.

3.1. Discrete Model

Considering K discrete time slots, the discretized form of the coded exposure in Eq. (9) for a given pixel m in the sensor array, can be written as

$$e_m = \sum_{k=1}^{K} \sum_{\ell=(U-1)/2}^{(U+1)/2} S_m^k \, T_{m+\ell}^k \, l_{m,\ell}, \tag{11}$$

where m, ℓ and k are the indexes for the discretized spatial, angular, and time dimensions, and M, U and K are the corresponding number of samples along these dimensions. Note the coded aperture is defined as shifted by a given number of pixels depending on the sub-aperture image ℓ.

Defining the discrete TMCA as

$$\hat{T}_m = \sum_{k=1}^{K} S_m^k \, T_{m+\ell}^k, \tag{12}$$

the discrete model of the coded exposure in (11) can be finally expressed as

$$e_m = \sum_{k=1}^{K} \sum_{\ell=(U-1)/2}^{(U+1)/2} \hat{T}_{m,\ell} \, l_{m,\ell}. \tag{13}$$

3.2. Simulation Results

The proposed end-to-end TMCA for compressive light field imaging is learned on a dataset aggregating real-world and synthetic light fields (LF). We use 100 real captured LF of size $7 \times 7 \times 376 \times 541$ from Lytro dataset (Kalantari et al., 2016), 22 synthetic LF images of size $5 \times 5 \times 512 \times$

512 from (Shi et al., 2019), and 33 synthetic LF images of size $5 \times 5 \times 512 \times 512$ from (Honauer et al., 2016). The aggregated dataset is randomly divided into 110 LF images for training, 20 for validation, and 25 for testing.

In this experiment, 5×5 angular views from a single snapshot with a resolution of 480×270 pixels are reconstructed. Randomly cropped patches of those images of spatial size 11×11 are employed and consider 5×5 angular views for training: randomly cropping the 4D LF into patches increases the number of samples at training while reducing the memory requirements to process an entire light field. The decoder is the deep spatial-angular convolutional sub-network proposed in (Guo et al., 2020), it is trained for 500 epochs using the L_2 norm. To optimize the TMCA under binary constraints, this thesis employs the sign function in the forward pass but employs a continuous approximation of the sign function an approach similar to the one presented in Martel et al. (2020). Specifically, a hard threshold is used for binary quantization in the forward pass while in the backward pass a smooth approximated sigmoid function is considered. After training, the estimated light field is obtained by merging the overlapping 4D patches with a median filter.

The proposed TMCA codification was compared against four different baselines: a) a reconstruction using the traditional dictionary learning and reconstruction approach of (Marwah et al., 2013), b) the same codification but using the deep network decoder from (Guo et al., 2020) c) our TMCA with random (non-optimized codes) using the deepnet decoder of (Guo et al., 2020) d) E2E optimization of (Marwah et al., 2013) with decoder (Guo et al., 2020) e) our full TMCA codification pipeline with learned codes. The results are compiled in Table 1, showing that on the two metrics we evaluated on (PSNR and SSIM) the TMCA is superior to the other baselines.

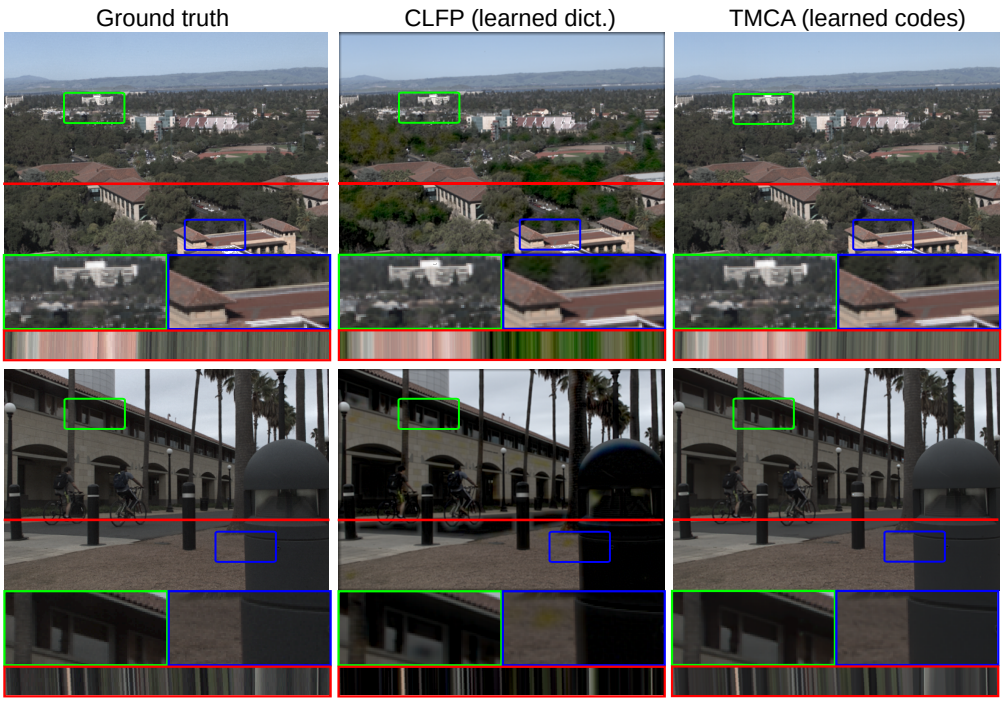

Figure 14. Examples of light-field reconstructions (central view) in simulation (shown with zoom-ins) comparing CLFP Marwah et al. (2013) and our proposed TMCA with ground truth.

Qualitative results are shown in Figure 14 where two recovered LFs of the testing set (Kalantari Lytro LFs) are reconstructed and compared with the ground truth. The learned coded apertures and shutter function in our pipeline are displayed in Figure 15. We use a 2D array of sub-images for visualization of the TMCA where each sub-image represents the response of the equivalent coded aperture to all rays arriving at one point on the coded aperture from all points on the aperture plane.

Sensitivity to noise and shutter length: Finally, quantitative results for additional experiments varying the Gaussian readout noise for a constant shutter length ($K = 8$) and varying the shutter length for a fixed level of noise are presented. The results can be found in Table 2 and 3 showing the less noise the better as expected, and an optimal shutter length value for $K = 8$.

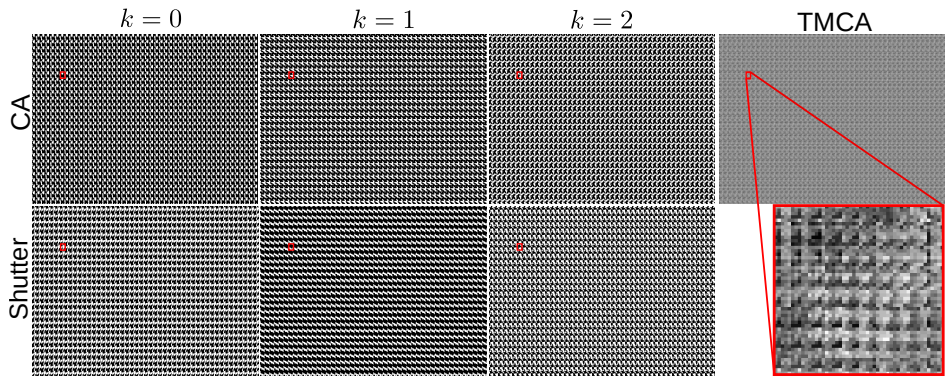

Figure 15. Learned TMCA for compressive light field imaging. Learned CA and shutter function for the first three time slots $k = 0, 1, 2$. We use a 2D array of sub-images for visualization of the TMCA where each sub-image represents the response of the equivalent coded aperture to all rays arriving at one point on the coded aperture from all points on the aperture plane. Thus, the resultant learned TMCA is equivalent to a coded aperture with sensitive angular pixels.

3.3. Experimental Evaluation

We also assess the proposed TMCA approach in a real experiment. We use a liquid crystal on silicon (LCoS) display (HOLOEYE PLUTO-2.1 LCoS SLM) where each pixel can independently change the polarization state of the incoming light field, in conjunction with a polarizing beam splitter and relay optics. As a single pixel on the LCoS cannot be well resolved with our setup, we treat blocks of 4×4 LCoS pixels as macro pixels, resulting in a CA of 480×270. We

Methods	PSNR(↑)	SSIM(↑)
CLFP(Marwah et al., 2013) (learned dict.)	30.06	0.82
CLFP(Marwah et al., 2013) (deep net.(Guo et al., 2020))	32.43	0.91
CLFP(Marwah et al., 2013) (E2E learned.)	33.21	0.91
TMCA (random codes)	34.03	0.93
TMCA (opt. codes)	**34.89**	**0.94**

Table 1
Compressive light field imaging: comparison of the proposed TMCA against baselines using the aggregated Lytro dataset.

$\sigma^2 =$	10^{-4}	10^{-3}	10^{-2}	10^{-1}
PSNR	34.67	33.57	31.65	28.98
SSIM	0.923	0.911	0.851	0.772

Table 2
Including additive Gaussian noise (Light-field Lytro dataset)

$k =$	2	4	8	16
PSNR	34.11	34.66	34.67	34.47
SSIM	0.894	0.908	0.911	0.916

Table 3
Varying the shutter length (Light-field Lytro dataset)

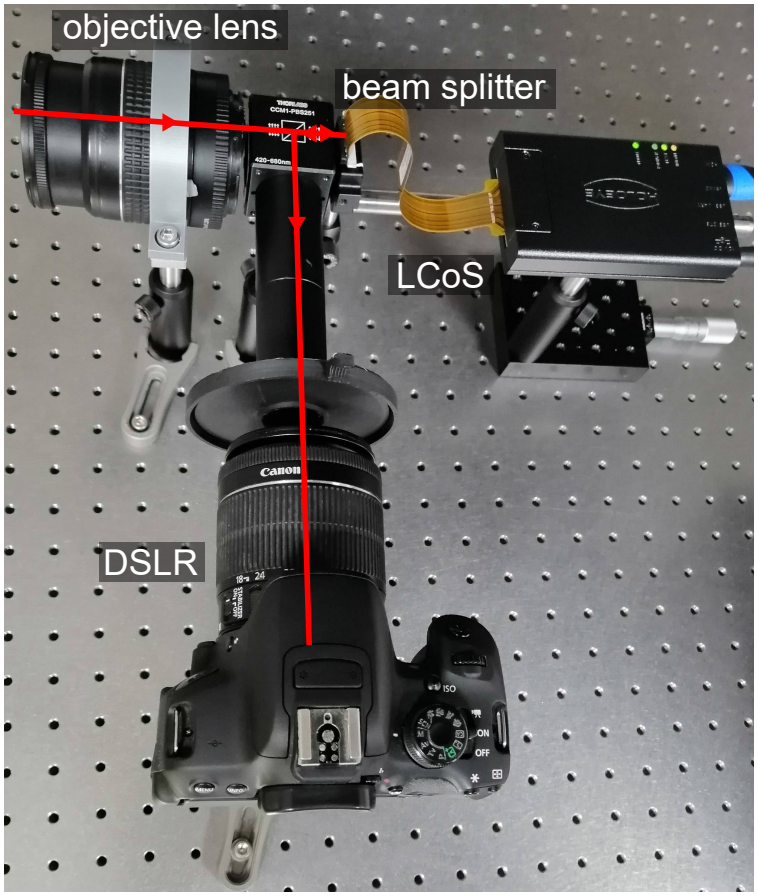

Figure 16. A photograph of the optical setup used in our compressive light field imaging experiments. The light path is shown in red and the various components are labeled.

Figure 17. Real captures of coded snapshots for light fields comparing CLFP (Marwah et al., 2013) and the proposed TMCA with their reconstructions.

reimage the LCoS with an SLR camera lens (Canon EF-S $18-55$ f/$4-5.6$ IS STM) which is not focused on the LCoS but in front of it, thereby optically placing the (virtual) image sensor behind the LCoS plane. A Canon EF 80 mm, f/5.6 II lens is used as the imaging lens and focused at a distance of 60cm. We follow the same procedure as in (Marwah et al., 2013) to adjust the distance between the mask (LCoS plane) and the virtual image sensor for capturing light fields with 5×5 angular resolution. A photograph of the optical setup is shown in Fig. 16.

The real captured measurements are reconstructed using the same deep networks as in simulation. The measurements and reconstructions are shown in Figure 17 which shows that the TMCA enables reconstructions featuring better spatial resolution in the angular views.

4. TMCA for Compressive Spectral Imaging

In this chapter, we delve into the second application of the proposed TMCA approach. Unlike Chapter 2, which focuses on analyzing the scene's geometry, this chapter explores the response to various wavelengths across the electromagnetic spectrum. Particularly, this thesis studied a compressive hyperspectral imaging application using the coded aperture spectral snapshot imager (CASSI) proposed in Wagadarikar et al. (2008) (see Figure 18). In this architecture, spectral dispersion is achieved using a prism between the lens and the sensor. The quantity we aim at recovering is the irradiance $I(x, y, \lambda)$. Note that we are now explicitly considering a second spatial dimension y (and will assume the prism disperses in the x dimension) and the spectral dimension λ. Similarly, we denote the spatial dependency of the coded aperture $T(x, y)$. The field impinging the sensor is now also dependent on the optical response of the prism h as well the spectral response

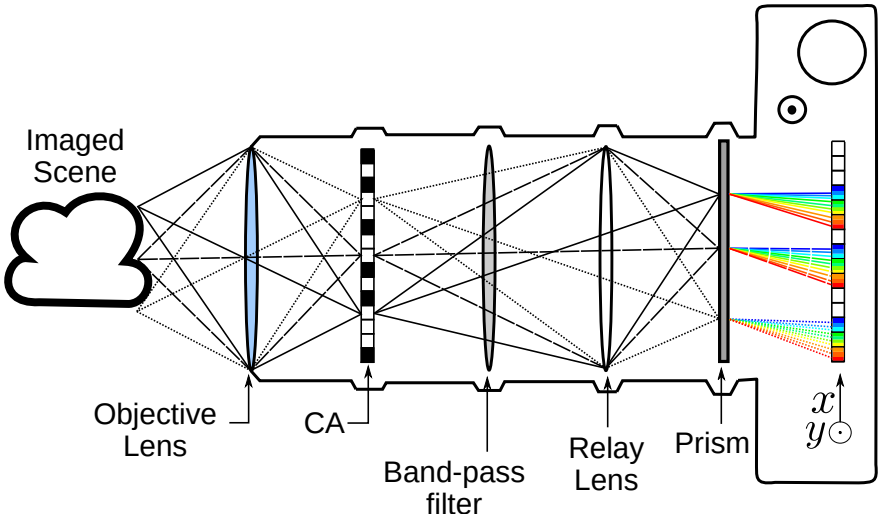

Figure 18. A diagram of the ray optics for coded aperture spectral snapshot imager (CASSI) using TMCA.

of the sensor κ:

$$g(x,y) = \iiint T(x',y')\, I(x',y',\lambda) h(x - \mathfrak{s}(\lambda) - x', y - y')\, \kappa(\lambda)\, \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}\lambda, \tag{14}$$

where $\mathfrak{s}(\lambda)$ is the wavelength-dependent spatial shift induced by the prism.

Using the shutter function S to create the TMCA, the optically encoded field g yields the coded exposure

$$e(x,y) = \int_{\Delta t} S(x,y,t') \iiint T(x',y',t')\, I(x',y',\lambda) h(x - \mathfrak{s}(\lambda) - x', y - y')\, \kappa(\lambda)\, \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}\lambda\, \mathrm{d}t'. \tag{15}$$

Since h is the propagation through unit magnification imaging optics and a dispersive element with linear dispersion, the impulse response can be expressed as $h(x - \mathfrak{s}(\lambda) - x', y - y') = \delta(x - \lambda - x', y - y')$. After substituting this expression in (15) simplifying and rearranging, we can express the coded measurements as

$$e(x,y) = \iiint \hat{T}(x',y',\lambda)\, I(x',y',\lambda) \delta(x - \lambda - x', y - y')\, \kappa(\lambda)\, \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}\lambda, \tag{16}$$

that use the TMCA $\hat{T}$ defined as

$$\hat{T}(x',y',\lambda) = \int_{\Delta t} S(x' + \lambda, y', t') T(x',y',t')\, \mathrm{d}t'. \tag{17}$$

In summary, adding the coded exposure to the model of Wagadarikar et al. (2008) generates

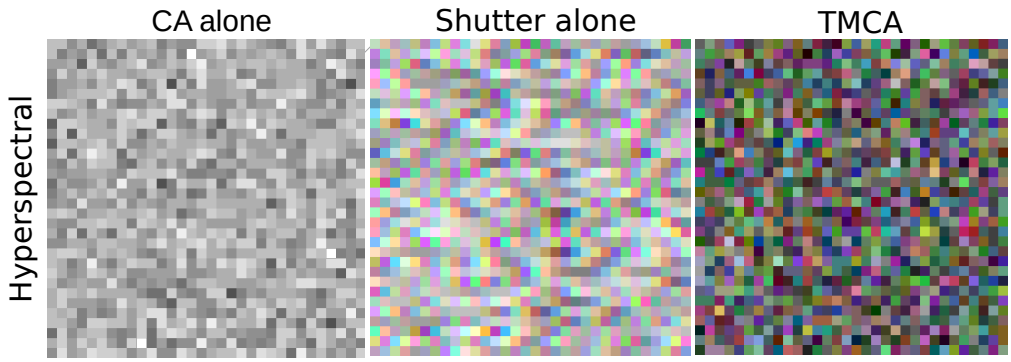

Figure 19. Illustrations of TMCA codification for compressive hyperspectral imaging with $K = 8$. Right: full codification as in Eq. (17). Left and middle: codifications when $\boldsymbol{S} = \text{Id.}$ (left) and $\boldsymbol{T} = \text{Id.}$ (middle). Even though the codes are binary, $K = 8$ yields 2^8 possible shutter / CA values.

a TMCA with a new coded aperture $\hat{T}$. The dependency of Equation (17) in λ shows that the proposed TMCA emulates a colored-coded aperture which would otherwise be expensive to create. Importantly, note that the spectral response is only true when both the shutter function and CA are jointly employed. If the shutter function is a constant, this is exactly the model in Wagadarikar et al. (2008): there is no spectral response of the CA. On the other hand, if the CA is removed, $\hat{T}$ exhibits a spectral response that shares the same code for every wavelength that is simply shifted in the x direction depending on the wavelength.

A final advantage of the TMCA codification is that even in the case we would restrict the coded exposure or coded aperture to binary values, for instance, because those would be simpler to realize physically, the proposed codification can still produce spatio-spectral patterns with non-binary values of attenuation. Figure 19 depicts the codification of the proposed TMCA for the CASSI system.

4.1. Discrete model

For a given pixel (m,n) (remember we index the spatial location in 2D and one of the spatial dimension is “special" and corresponds to the dimension in which the prism disperses light), and considering K discrete time slots, the measurement model can be written as:

$$e_{m,n} = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} S_{m,n}^{k} f_{i,j,\ell} T_{i,j}^{k} h_{m-i,n-j,\ell}, \tag{18}$$

where i, j are indices along the discretized spatial dimensions of M,N samples each, ℓ,k denote the indices for the discretized wavelength and time dimensions, and L,K are the corresponding number of samples along those.

The point spread function h corresponds to a propagation model through a unit magnification imaging optics. Further assuming the prism features linear dispersion, the PSF can be expressed as the shifted Dirac $\delta_{m-i,n-j,\ell}$. Substituting this expression in Equation (18), and simplifying, the exposure model becomes

$$e_{m,n} = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} S_{i,j+\ell}^{k} T_{i,j}^{k} f_{i,j,\ell} \delta_{m-i,n-j,\ell}. \tag{19}$$

Grouping the time variables in (19), we then define the discrete TMCA:

$$\hat{T}_{i,j,\ell} = \sum_{k=1}^{K} S_{i,j+\ell}^{k} T_{i,j}^{k}, \tag{20}$$

Methods	PSNR(↑)	UIQI(↑)	SAM(↓)	ERGAS(↓)	DD(↓)
CASSI (ADMM)	27.40	0.938	22.42	9.56	0.031
CASSI (U-Net)	29.66	0.968	15.99	7.04	0.022
CASSI (E2E learned)	30.23	0.971	15.11	6.72	0.020
TMCA (random)	31.39	0.978	13.01	5.74	0.019
TMCA (learned)	**32.72**	**0.981**	**11.92**	**5.27**	**0.016**

Table 4
Compressive spectral imaging: Comparison of the proposed TMCA with baselines on the ICVL 1 dataset.

yielding the coded measurements

$$e_{m,n} = \sum_{\ell=1}^{L} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \hat{T}_{i,j,\ell} \, f_{i,j,\ell} \, \delta_{m-i,n-j,\ell}. \tag{21}$$

4.2. Simulation Results

The model for hyperspectral imaging was learned using the ICVL dataset. It consists of 200 spectral images. We randomly select 160 hyperspectral images for training, 20 for validation, and 20 for testing, cropped at a size of 256×256 with $L = 12$ spectral bands. We set the number of time slots in the TMCA encoder to $K = 8$. The U-Net is trained for 500 epochs. An L_2 norm is selected as the objective function in the E2E optimization process. As described in the previous chapter, the binary constraints of the shutter and CA are considered using the approximation in Martel et al. (2020).

The proposed TMCA codification was compared against four different baselines: a) the traditional CASSI codification using random binary patterns and reconstructed using the alterna-

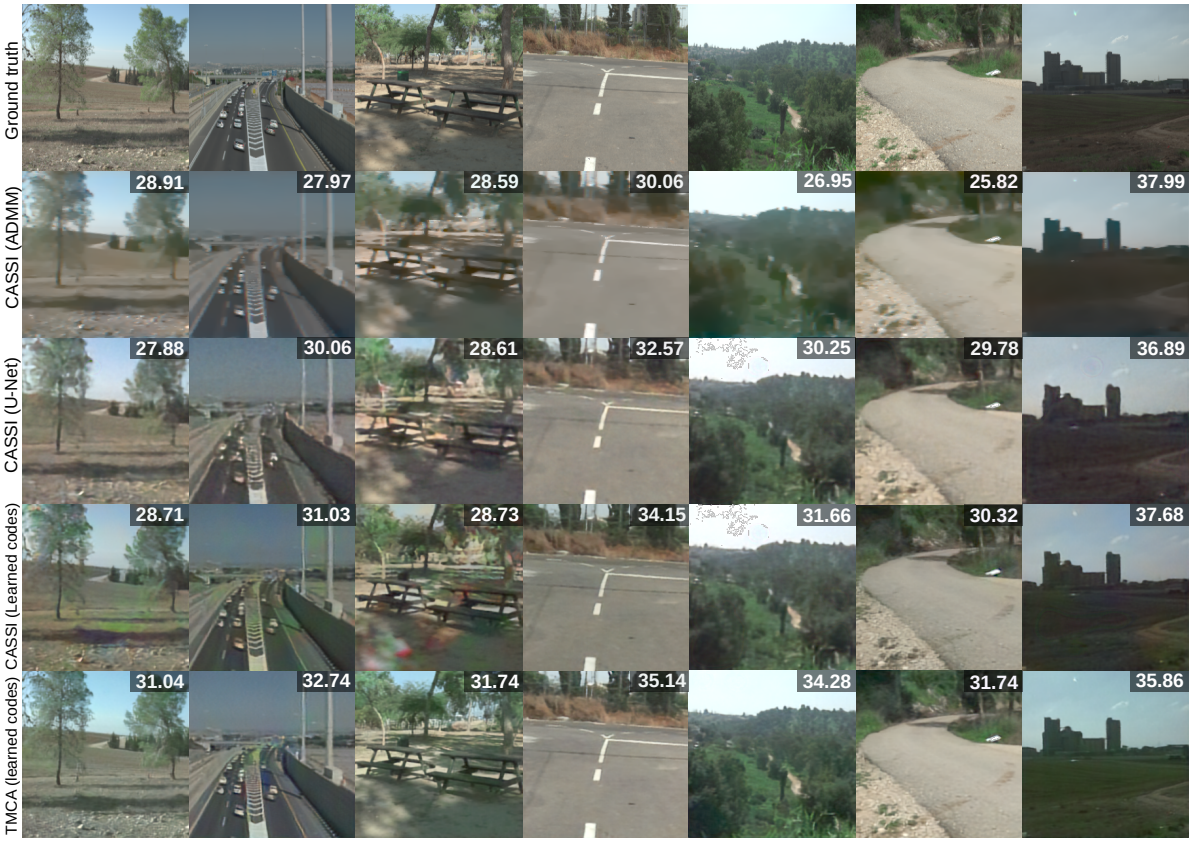

Figure 20. Results for compressive spectral imaging on the ICVL 1 test fold comparing our TMCA (fifth row) against the CASSI using a conventional optimization technique (ADMM) as a decoder (second row), a U-Net decoder (third row), and jointly learned codes and U-Net decoder (fourth row). The PSNR (dB) between the reconstructions and the ground truth images is shown in the lower-right corner.

tive direction method of multipliers (ADMM) Boyd et al. (2011) b) the CASSI codification using a trained U-Net as a decoder c) the proposed TMCA codification and reconstruction pipeline using random (non-optimized) codes d) the CASSI codification with joint learned codification and trained U-Net as a decoder and e) our full TMCA codification with learned codes. The quantitative results performed on our random test fold of ICVL are presented in Table 4, they show that on all the metrics we evaluated on (PSNR, UIQI, SAM, ERGAS, and DD) our full TMCA pipeline performs better than all the other baselines. Qualitative results of a few reconstructed hyperspec-

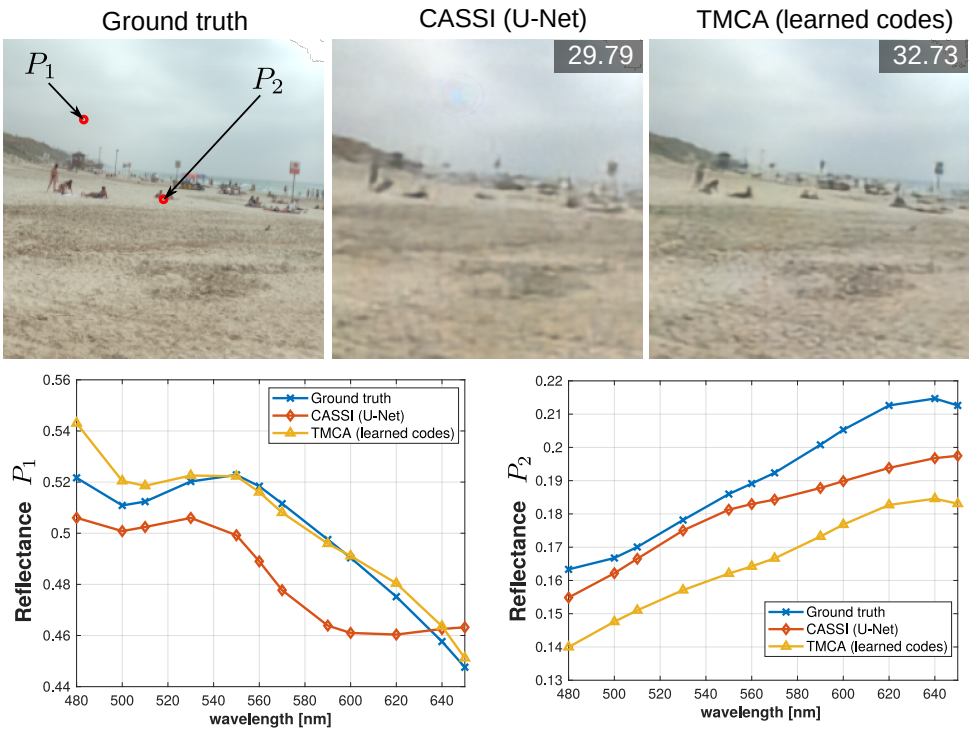

Figure 21. Spectral signatures for two points in an image randomly sampled from the ICVL 1 dataset. The $L = 12$ spectral bands are shown for a point in the sky and a point in the sand on the reconstructions performed by our method using TMCA and learned codes and the conventional CASSI with a U-Net.

tral images are shown in Figure 20. The hyperspectral images are mapped to an RGB composite image by selecting three spectral channels corresponding to wavelengths of 650, 550 and 480 nm. These images illustrate how the proposed TMCA appears closer to the ground truth both in terms of spatial accuracy (they are less blurry) as well as spectral accuracy (colors match better).

Spectral signatures Qualitative and quantitative comparisons of full spectral signatures ($L = 12$) for two different points taken in a randomly sampled image of the ICVL 1 dataset are shown in Figure 21 while $L = 6$ spectral bands for two other randomly sampled images are shown in Figure 22.

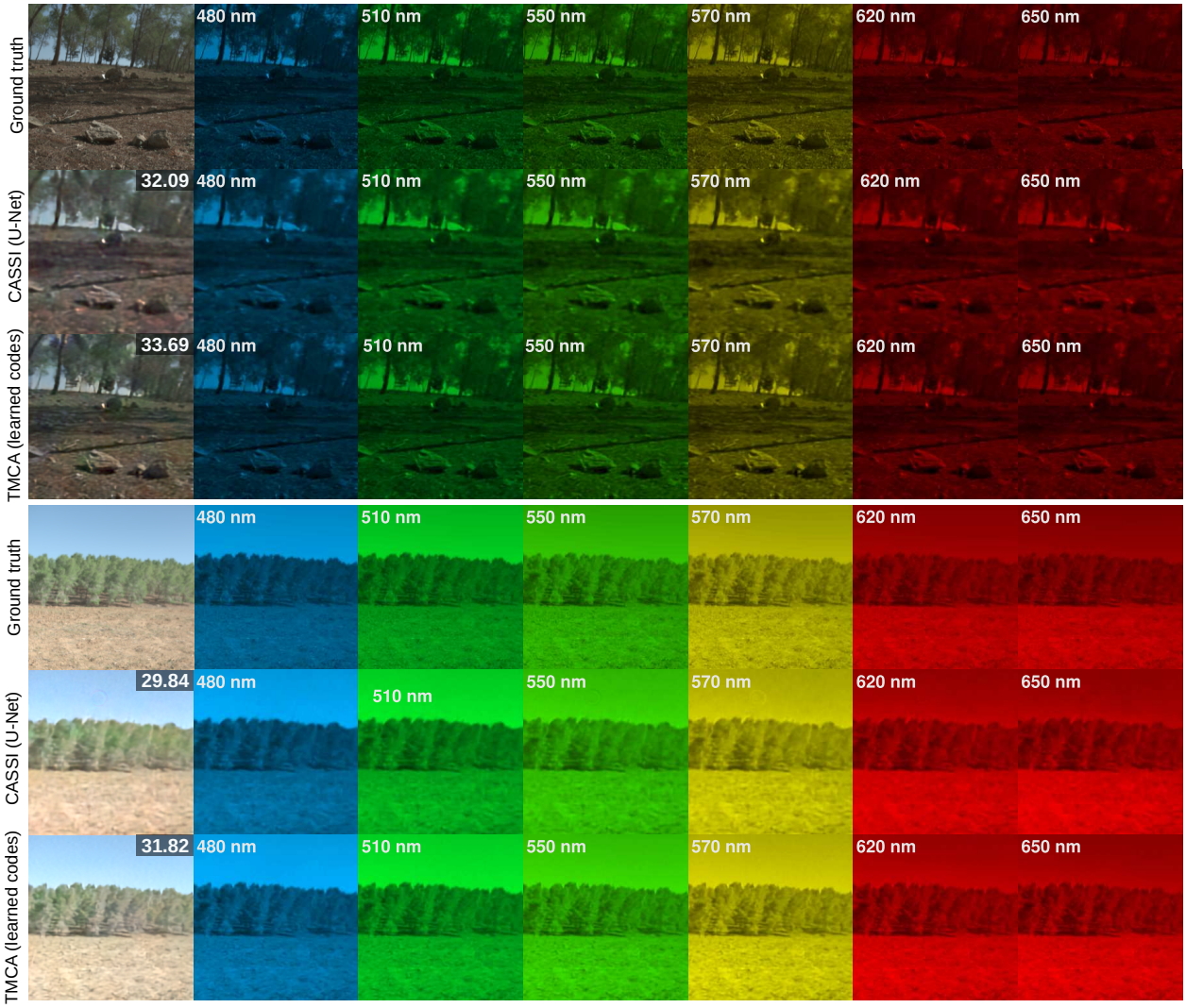

Figure 22. Results of our compressive spectral imaging reconstruction on two synthetic images of the ICVL 1 Dataset showing 6 different spectral bands (out of $L = 12$).

Sensitivity to noise and shutter length: We performed additional experiments varying the Gaussian readout noise for a constant shutter length ($K = 8$) and varying the shutter length for a fixed level of noise. The results can be found in Table 5 and 6 showing the less noise the better. We find an optimal shutter length value to be $K = 16$.

$\sigma^2 =$	10^{-4}	10^{-3}	10^{-2}	10^{-1}
PSNR	32.051	31.723	31.311	30.207
UIQI	0.979	0.976	0.977	0.965
SAM	5.43	5.73	5.98	6.86
ERGAS	12.57	12.92	13.49	15.41
DD	0.017	0.018	0.019	0.022

Table 5
Including additive Gaussian noise (ICVL 1 dataset)

$K =$	2	4	8	16
PSNR	31.032	30.937	31.723	31.858
UIQI	0.971	0.972	0.976	0.978
SAM	6.29	6.21	5.73	5.54
ERGAS	14.17	14.04	12.92	12.63
DD	0.021	0.020	0.018	0.018

Table 6
Varying the shutter length (ICVL 1 dataset)

4.3. Experimental Evaluation

This research built a prototype of our system to evaluate the proposed TMCA approach for real-world scenes. The system consists of an achromatic objective lens with 50mm focal length (Thorlabs AC254-050-A-ML), a digital micromirror device (DMD DLi4120), an F/8 relay lens, a custom double Amici prism with center wavelength 550nm captured through a monochrome CCD sensor Stingray F-080B with 4.65μm pixel size. The built prototype is shown in Figure 23. The same prototype was used to capture measurements for the traditional CASSI codification and the proposed TMCA. The spectral images are reconstructed using the U-Nets trained in simulation, but using the real measurements. Results showing the capture of a book cover are shown in Figure 24 demonstrating that TMCA imaging can be implemented in a real-world imaging prototype that

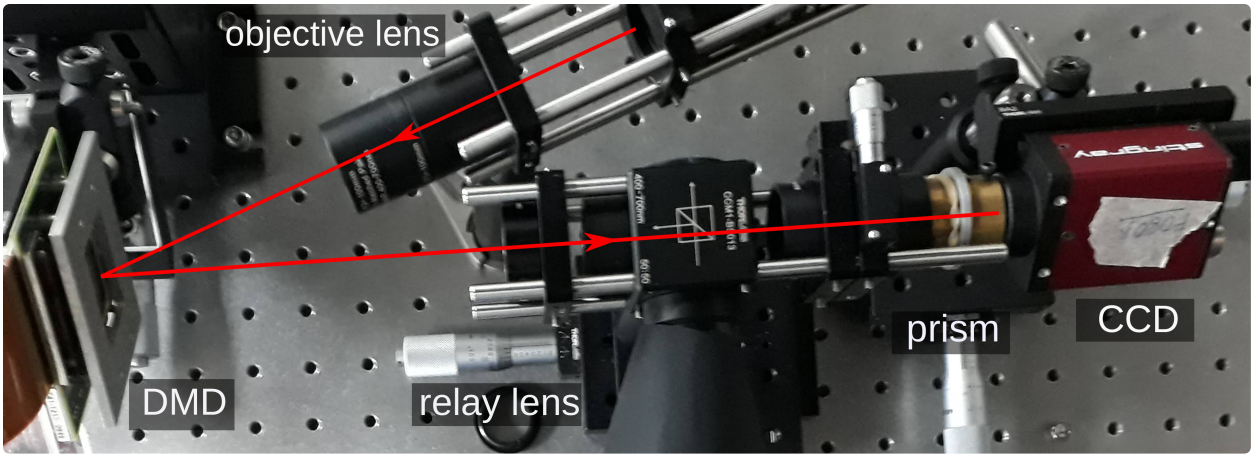

Figure 23. A photograph of the optical setup implementing our hyperspectral compressive imaging system. The light path is shown in red and the various components are labeled.

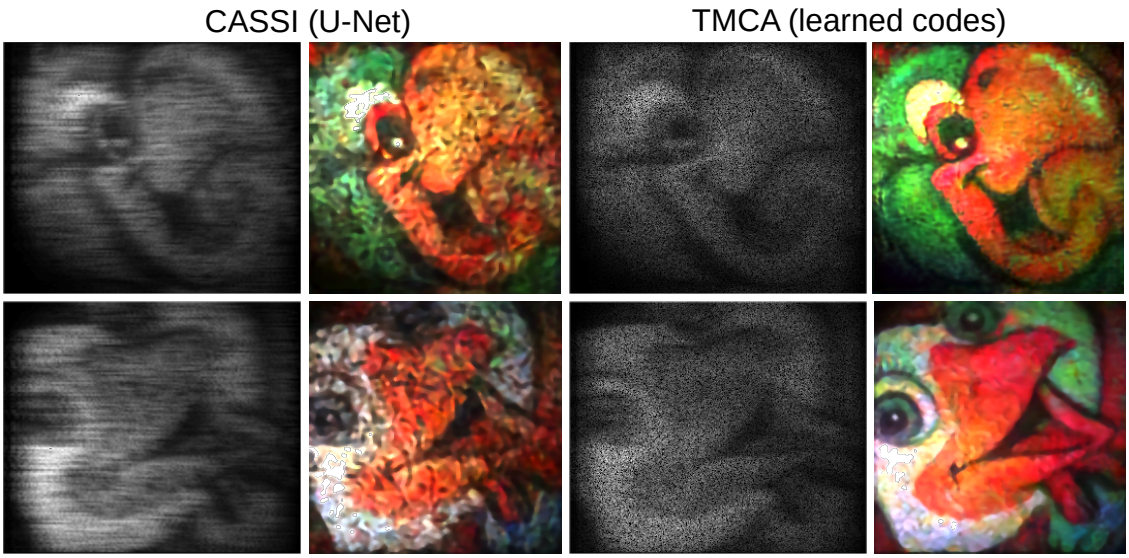

Figure 24. Examples of real captures of coded snapshots for hyperspectral imaging and their reconstructions.

still presents a higher visual quality than traditional CASSI.

5. TMCA for Spectral-Depth Imaging

Building upon the previous two chapters, this chapter presents our approach leveraging TMCA for joint imaging of spectral-depth (SD) information from a single snapshot. The proposed snapshot SD system consists of a free-form phase mask, with arbitrary thickness profile h, in the aperture plane and a CCA on the sensor plane (see Figure 25). The distance separating the aperture plane and the sensor plane is denoted as s. The incident irradiance to the system is a SD source density $f(x,y,z,\lambda)$ that responds differently to every source point depending on the phase mask height map h, depth z, and wavelength λ. Here, wave optics is employed to obtain the PSF (Goodman, 2005). In contrast with previous applications described in chapters 3 and 4 that employ amplitude-coded aperture $T(t)$, it is important to note that here the TMCA codification employs a phase-coded aperture ($h(t)$) synchronized with the shutter function $S(t)$. To start describing the

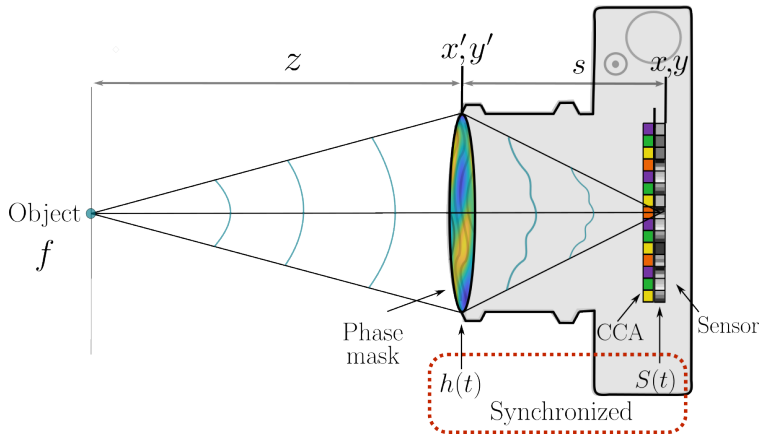

Figure 25. Schematic optical setup of the proposed TMCA codification strategy for snapshot SD imaging. The phase mask varies in time and the sensor includes a shutter function. The CCA encodes the spectral information before being integrated into the sensor.

mathematical model of the proposed model, first, consider the spherical light wave emitted by a point source at a distance z from the aperture plane. The unit-amplitude complex-valued electric field immediately before the phase mask is given by

$$U_{\text{in}}(x',y') = \mathrm{e}^{ik\sqrt{x'^2+y'^2+z^2}}, \tag{22}$$

where $k = 2\pi/\lambda$ is the wavenumber. The wave then travels through the phase-mask h; the equation governing the field, immediately after, is obtained by multiplying the impinging field $U_{\text{in}}(x',y')$, by $b(x',y')$. This models the delay at each location (x',y') on the phase-mask, induced by its thickness and depending on its refractive index. Then, the phase transformation is given by:

$$b(x',y') = \mathrm{e}^{i\phi_d(x',y')}, \tag{23}$$

where

$$\phi_d(x',y') = \frac{2\pi\Delta n}{\lambda} h(x',y'), \tag{24}$$

and Δn is the refractive index difference between air and the material of the refractive or diffractive optical element constituting the free-form phase mask. Additionally, we model an aperture by inserting an amplitude function $A(x',y')$ modulating the unit-amplitude field. The aperture blocks all the light in regions outside its opening. Then, the electric field immediately after the phase mask is calculated by multiplying both the amplitude modulation (introduced by the aperture) and

the phase modulation (introduced by the free-form lens) with the impinging electric field:

$$U_{\text{out}}(x',y') = A(x',y') \cdot b(x',y') \cdot U_{\text{in}}(x',y'). \tag{25}$$

Finally, the field propagates a distance s from the phase-mask to the focal plane assuming the transfer function Goodman (2005):

$$H_s(k_x,k_y) = \mathrm{e}^{iks\sqrt{1-(\lambda k_x)^2-(\lambda k_y)^2}}, \tag{26}$$

where (k_x,k_y) are spatial frequencies. This transfer function, in the Fourier domain, produces an electric field in front of the sensor described as

$$\begin{aligned} p(x'',y'',z,\lambda) &= U_{\text{FPA}}(x'',y'') \\ &= \left|\mathscr{F}^{-1}\left[\mathscr{F}\left[U_{out}(x',y')\right] H_s(k_x,k_y)\right]\right|^2, \end{aligned} \tag{27}$$

where $\mathscr{F}$ denotes the 2D Fourier transform. The mathematical model in (27) allows obtaining a 2D PSF for each wavelength (λ) and depth (z) of interest. Based on (27), we can approximate the incident irradiance g as the sum of the different contributions of every point source of the scene. If we consider the scene is composed of a planar object located at a given depth z, g can be modeled as the simple convolution of the PSF in (27) and f along the sensor dimensions. However, since natural scenes contain depth variations, the incident irradiance is modeled as a spatially-variant convolution. Here, we consider a two-dimensional manifold in 3D space with local intensity; that

is, we consider an object parameterized with spatial coordinates (x,y) and depth coordinate $z(x,y)$, resulting in the irradiance $f(x,y,z(x,y),\lambda)$. For simplicity in the notation, hereafter we denote the irradiance function as $f(x,y,\lambda)$. Thus, the encoded irradiance g in the sensor plane can be approximated as

$$g(x,y,\lambda) = \iint f(x',y',\lambda)p(x-x',y-y',z(x',y'),\lambda)\,\mathrm{d}x'\,\mathrm{d}y', \tag{28}$$

where p is the spatially-variant PSF of the system that varies with depth $z(x',y')$.

We further consider an approximated *layered* depth model, with D depth planes, that lead to a finite number of PSFs corresponding to each depth plane. Under this approximation, the spatially variant PSF can be linearly separated as

$$p(x-x',y-y',z(x',y'),\lambda) = \sum_{d=0}^{D-1} \omega_d(x',y')p_d(x-x',y-y',\lambda), \tag{29}$$

where $p_d(x,y,\lambda) = p(x,y,z_d,\lambda)$ corresponds to the PSF at depth plane z_d, and ω_d are binary weights, such that, $\omega_d = 1$ if spatial location (x',y') corresponds to the d-th depth layer. Note that the linear approximation in (28) can accurately model the response from constant depth regions, but it fails at depth discontinuities. To improve the accuracy of the forward model, Ikoma et al. (2021) proposed a nonlinear differentiable model based on alpha compositing. For simplicity, we stick to the linear model, but a similar conclusion can be derived for the nonlinear model in

Ikoma et al. (2021).

With the approximation in (29), the spatially-variant convolution in (28) simplifies to a sum of spatially-invariant convolutions of the form

$$g(x,y,\lambda) = \sum_{d=0}^{D-1} \iint f(x',y',\lambda)\,\omega_d(x',y')\,p_d(x-x',y-y',\lambda)\,\mathrm{d}x'\,\mathrm{d}y'. \tag{30}$$

Now considering a time-varying height profile h, denoted by $h(t)$, the CCA denoted as $C(x,y,\lambda)$, and the shutter function S, the measured coded exposure of the proposed system is given by

$$e(x,y,\lambda) = \int C(x,y,\lambda)\,S(x,y,t) \sum_{d=0}^{D-1} \iint f(x',y',\lambda)\,\omega_d(x',y')\,p(x-x',y-y',z_d,\lambda,t)\,\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}t, \tag{31}$$

where p varies in time due to the variations in h. After regrouping the time-varying terms, the coded exposure can be succinctly expressed as

$$e(x,y,\lambda) = \iint f(x',y',\lambda)\,\tilde{p}(x-x',y-y',z(x',y'),\lambda)\,\mathrm{d}x'\,\mathrm{d}y', \tag{32}$$

where

$$\begin{aligned}\tilde{p}(x-x',y-y',z(x',y'),\lambda) = \\ \sum_{d=0}^{D-1} \omega_d(x',y') \int p(x-x',y-y',z_d,\lambda,t)\,C(x+x',y+y',\lambda)\,S(x+x',y+y',t)\,\mathrm{d}t\end{aligned} \tag{33}$$

is the equivalent spatially-variant PSF realized by the TMCA codification. Note that, compared

to the traditional PSF in (29), the proposed TMCA entails a different PSF for every position (x', y'), for a constant depth plane in the scene. We leverage this characteristic of the TMCA to enhance depth estimation by sacrificing spatial resolution, an advantage that cannot be achieved in a spatially-invariant system. Furthermore, the synchronized time variations of the phase mask and shutter function increase the expressibility of the equivalent PSFs, thus providing a wider range of possibilities to design the PSF.

5.1. Discrete model

For a given pixel (m,n) in the sensor plane, and considering K discrete time slots, the discrete measured exposure in (32) is given by

$$e_{m,n,\ell} = \sum_i \sum_j f_{i,j,\ell} \tilde{p}_{m-i,n-j,z(i,j),\ell}, \tag{34}$$

where

$$\tilde{p}_{m-i,n-j,z(i,j),\ell} = \sum_{d=0}^{D-1} \omega_{d,i,j} \sum_{k=0}^{K-1} p_{m-i,n-j,d,\ell,k} C_{m+i,n+j,\ell} S_{m+i,n+j,k}, \tag{35}$$

i, j are indices along the discretized spatial dimensions of M,N samples, ℓ, k denote the indices for discretized wavelength and time dimensions, and L and K are the corresponding number of samples along them. According to (35), the number of distinctive PSFs depends on the spatial resolution of the shutter and the CCA, thus, leading to a considerable number of PSFs and model complexity. To reduce this number, we constrain the structure of the shutter and CCA to be periodic with period

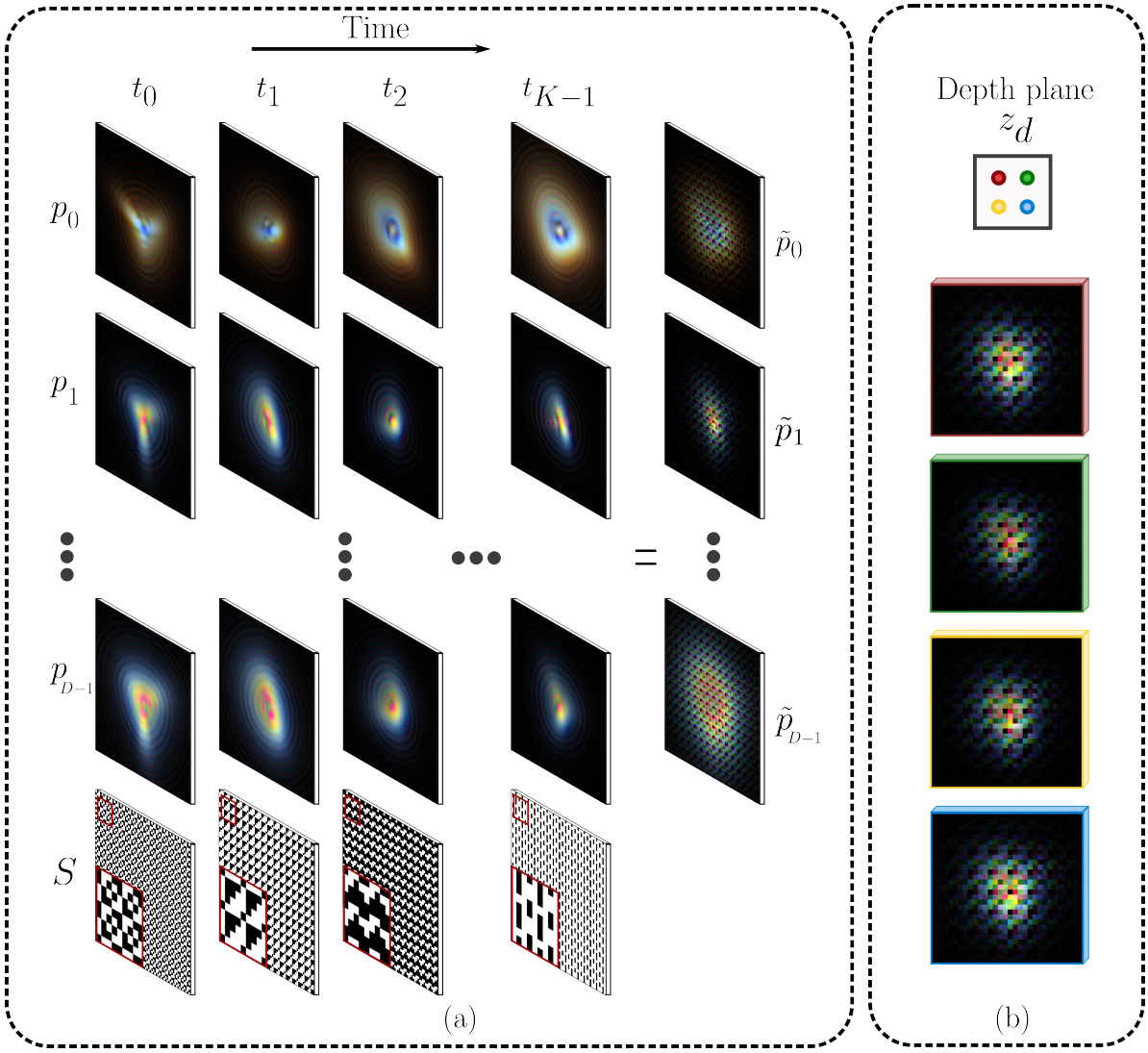

Figure 26. (a) Depth-and-time varying PSFs p, shutter function S, and equivalent depth-variant PSF $\tilde{p}$ of a point source from a given position (i, j) in the image plane. (b) PSFs of different points at the same depth z_d.

$Q \ll \{N, M\}$. In this case, the number of all PSFs reduces from $MNDL$ to Q^2DL. Moreover, the CCA filters cannot be arbitrary spectral responses in real-world scenarios. To alleviate this constraint, we adopt the methodology in Arguello et al. (2021), where the spectral response of the CCA filters is constrained to be a linear combination of R feasible filters. Figure 26(a) shows the depth-and-time varying PSFs p, the shutter function S, and the equivalent depth-variant PSF $\tilde{p}$ in (35), for a point source in a given position (i, j) at the image plane. We recall that the proposed TMCA yields a different PSF for every position, for a constant depth plane in the scene, leading to accurate depth estimation by sacrificing spatial resolution. This characteristic is illustrated in Fig. 26(b) by showing the PSFs of 4 different points in the image plane, at the same depth z_d. In comparison to depth from defocus approaches that encounter difficulties in textureless regions, the proposed TMCA pattern mitigates this ambiguity, associated with depth information, with the spatially-variant response.

Further, in an imaging system, the recorded image in a channel c is an integration of the spectral information weighted by the spectral response of the c-th sensor channel. Considering a classical image formation model, the recorded image per channel, y^c, can be finally expressed as

$$y^c_{m,n} = \sum_{\ell} e_{m,n,\ell} \kappa^c_{\ell} + w_{m,n}, \tag{36}$$

where κ^c_{ℓ} represents the c-th spectral response of the sensor, at the ℓ-th spectral band, and $w_{m,n}$ accounts for additive Gaussian noise.

Figure 27. Comparison of the recovered spectral images, mapped to RGB, using our proposed TMCA method and the competitive baseline Baek et al. (2021). Visually, the colors of the proposed approach better resemble the ground truth, demonstrating a better spectral estimation. The PSNR between the recovered and the ground truth images is reported in the lower-right corner of each inset.

5.2. Simulation Results

One important aspect of a learning model is the dataset used to supervise the optimization. Even though there is a plethora of spectral and depth map datasets, there exist very limited SD datasets. This research uses a recent benchmark SD dataset proposed in Baek et al. (2021). It consists of 18 spectral-and-depth map image pairs with a spatial resolution of 2824×4240 pixels. The spectral images span along 25 wavelengths from 440 nm to 680 nm, with a 10 nm interval. The depth images varies within $[0.4 - 2.0]$ m. 14 spectral images are randomly selected for training, and 4 for testing. At *training* time, patches of size 256×256 are randomly selected. At *testing*

time, the learned model on non-overlapping patches of size 256×256 are employed to construct the whole image. The SD image is cropped to a spatial resolution of 2816×4096 to obtain an integer number of patches. The number of time slots in the TMCA encoder is set to $K = 4$, the number of depths to $D = 8$, and the number of feasible color filters to $R = 4$. The neural network is trained for 1000 epochs with a learning rate of 0.001. Our models are implemented in PyTorch Paszke et al. (2017) and trained on a Titan RTX GPU using the ADAM optimizer Kingma and Ba (2014). The proposed TMCA codification method is compared against a baseline that employs

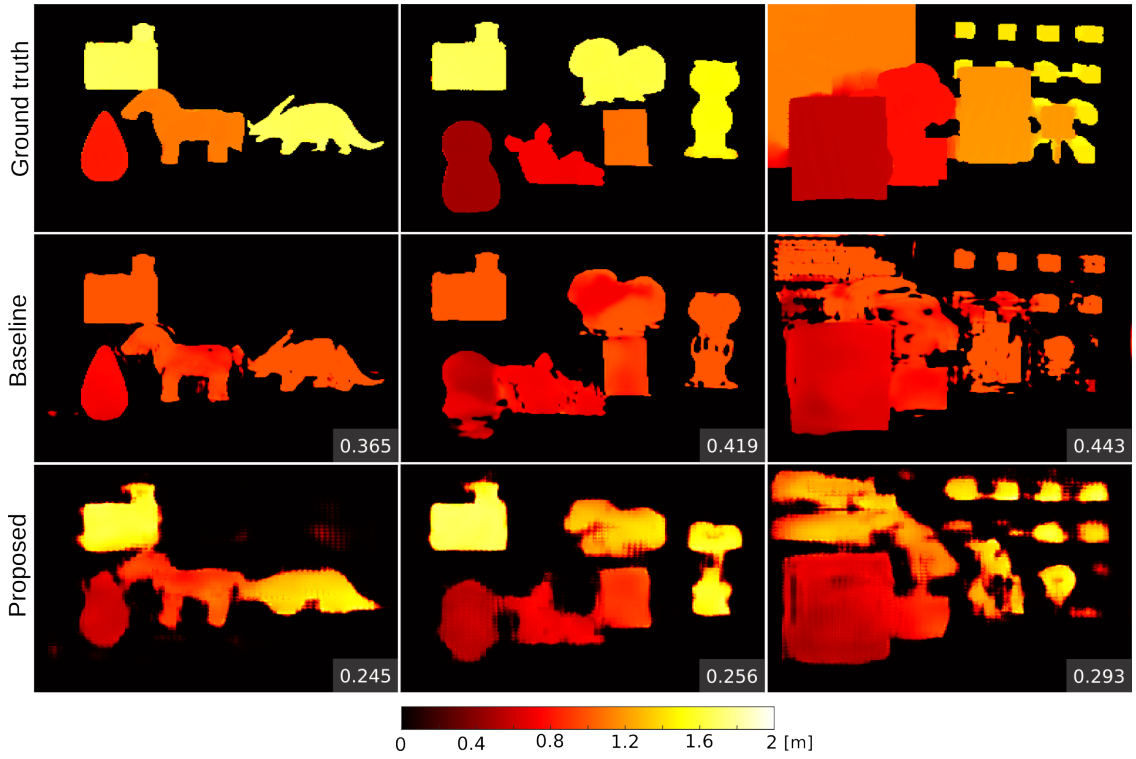

Figure 28. Comparison of the estimated depth maps using the proposed TMCA method and the competitive baseline Baek et al. (2021). The estimated depth maps by our proposed approach attain lower spatial resolution but exhibit higher fidelity to the ground truth depth values. The MAE between the estimated and ground truth depth maps is reported in the lower-right corner of each inset.

Table 7
Spectral and depth imaging performance in simulation. Comparison of the proposed TMCA method against the baseline Baek et al. (2021).

	Spectral metrics			Depth metrics	
	PSNR [dB] (↑)	SAM (↓)	UIQI (↑)	MAE (↓)	RMSE (↓)
Baseline Baek et al. (2021)	37.1367	8.7454	0.9756	0.3770	0.5167
Proposed	**37.9176**	**8.1457**	**0.9883**	**0.2581**	**0.4034**

DOE-only codification Baek et al. (2021). To obtain the results in Baek et al. (2021), we modified our own implementation removing the time-varying elements, the coded shutter and the CCA. The quality of the spectral reconstruction is evaluated using the peak signal-to-noise ratio (PSNR), the spectral angle mapper (SAM) metrics Kruse et al. (1993), and the universal image quality index (UIQI) Wang and Bovik (2002); the depth map reconstruction is evaluated using the root-mean-square error (RMSE) and the mean absolute error (MAE). Quantitative results evaluated on the test set are presented in Table 7 showing that the proposed TMCA codification performs slightly better than the baseline in terms of spectral performance, while showing superior performance in the depth estimation. Qualitative results of 3 recovered spectral images, depicted in Fig. 27, show that the proposed method can recover better spectral information (better color matching to the ground truth), while attaining a similar spatial resolution. The corresponding estimated depth maps from scenes used in Fig. 27 are shown in Fig. 28, which demonstrates that the TMCA yields depth maps that attain lower spatial resolution but featuring higher fidelity to the ground truth depth maps. Finally, recovered spectral signatures of 3 different colors in the color checker, shown in Fig. 29, confirm the better recovery of the spectral information.

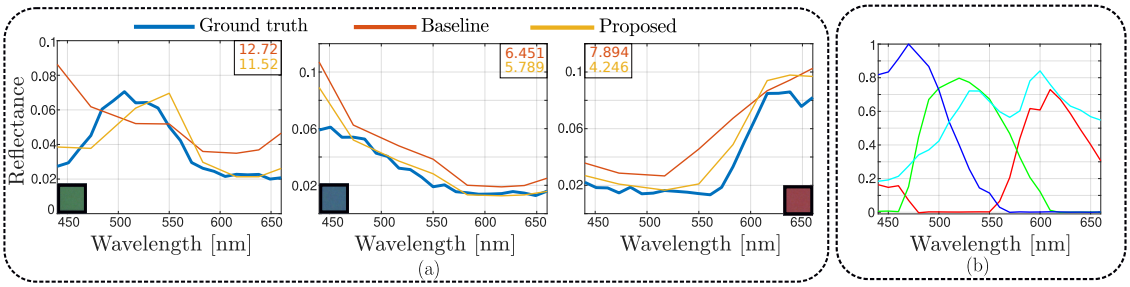

Figure 29. (a) Comparison of the reconstructed spectral signatures of 3 color patches from the color checker in Fig. 27, between the proposed TMCA and the competitive method Baek et al. (2017). The SAM metric between the estimated and ground truth signatures is reported in the upper-right corner of each inset (b) Spectral response of the feasible color filters used in the CCA Arguello et al. (2021).

5.3. Experimental Evaluation

This research built a proof-of-concept prototype to demonstrate the proposed TMCA for real-world scenes. The prototype, shown in Fig 30, consists of a reference objective lens (Canon EF 28-80 mm), which focuses the light of the scene at a distance of 40 mm. Employing a pair of 100 mm focal length lenses (achromatic doublet Lens f=100 mm, Thorlabs, AC254-100-A-ML) and a beam splitter (BS) (Thorlabs CCM1-BS013, 30 mm non-polarizing), we generate a 4F system, and modulate the wavefront at 2F using a deformable mirror (DM, actuator piezo, Thorlabs, DMP40-P01-40) with a spectral range of 450-2000 nm. The phase modulation of the DM is dictated through the learned Zernike coefficients that are configured using the official software from Thorlabs. Finally, through the BS and the relay lens, an image is formed on the sensor of a Canon EOS M50 camera. We emulate the CCA, shutter function, and time integration. To emulate the

CCA, we illuminate the target scene using a tunable light source (Newport TLS130B[1]) and obtain the spectral band intensity with the gray-scale version of the camera. Once the coded measurements are obtained, the spectral-and-depth image pair is reconstructed using the same networks as in the simulation. For the experimental evaluation, we fine-tuned the network learned in simulation by using our own spectral-depth dataset. This dataset contains 63 spectral images with their corresponding depth map, see Fig. 31. The acquisition of the spectral images was carried out by illuminating the scene with a tunable light source (Newport TLS130B). We used wavelengths between $400-700$ nm with intervals of 10 nm. The depth maps vary in a range of $0.4-1.6$ m, estimated using structured light (SL) setup and the phase unwrapping via graph cuts algorithm Bioucas-Dias and Valadao (2007). Fig. 31 shows a few samples of the acquired spectral depth dataset and the optical setup employed to acquire them. A summary of the coded snapshot acquired, the corresponding recovered spectral image, and the estimated depth map, are shown in Fig. 30. This result demonstrates that the proposed approach can recover SD information with high quality in a real-world scenario.

[1] https://www.newport.com/p/TLS130B-300X

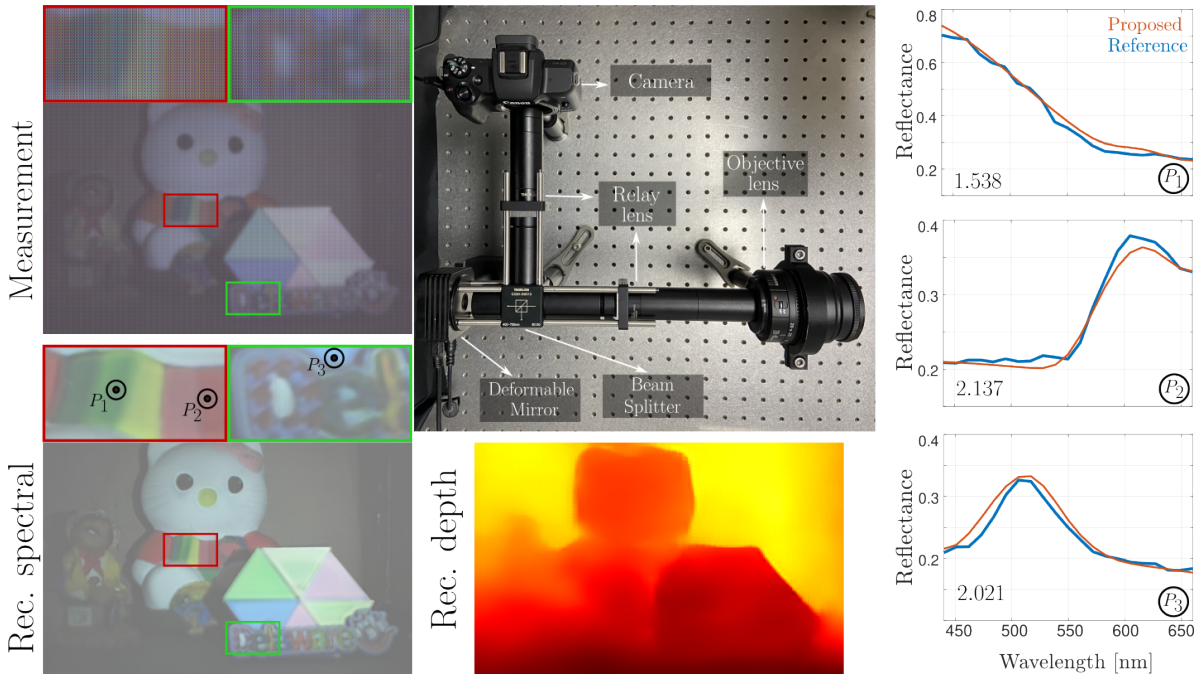

Figure 30. Proof-of-concept testbed prototype of the proposed TMCA snapshot SD system, along with a real coded acquisition and the corresponding recovered spectral-and-depth information using the optimized decoder. The estimated reflectances of three spectral signatures at points P_1, P_2, and P_3 are compared with reference signatures measured with a spectrometer and report the SAM metric in the lower-left corner of the inset.

Figure 31. (left) Optical system used to acquire our own spectral-depth dataset using spectral scanning and structured light. The light projector (1) emits a collimated beam of structured light patterns that is captured by the camera (2). The tunable light source (4) emits selective wavelengths to travel through an optical fiber (3) to illuminate the scene which is also captured with camera (2). (right) A few samples of the acquired dataset.

6. Extensions: Dual Blind Deconvolution for Joint-Radar Communications

The last chapter of this thesis presents an extended idea that came out in the fulfillment of the objectives. Specifically, this chapter delves into the joint radar-communications processing application for spectrum efficiency. With the advent of new wireless communication systems and novel radar technologies, the electromagnetic spectrum has become a contested resource Mishra et al. (2019). This has led to the development of various system engineering and signal processing approaches toward optimal, secure, and dynamic sharing of the spectrum Paul et al. (2016). In general, spectrum-sharing technologies follow three major approaches: co-design Liu et al. (2022), cooperation Bică and Koivunen (2018), and coexistence Wu et al. (2022). While co-design requires developing new systems and waveforms to efficiently utilize the spectrum Duggal et al. (2020), cooperation requires an exchange of additional information between radar and communications to enhance their respective performances Paul et al. (2016). In the coexistence approach, radar and communications independently access the spectrum and spectrum-sharing efforts are largely focused on receiver processing. Among all three approaches, separating the radar and communications signals at a coexistence receiver is highly challenging because the uncoordinated sharing of resources does not help in clearly distinguishing the two received signals. In this paper, we focus on the coexistence problem.

In general, coexistence systems employ different radar and communications waveforms and separate receivers, wherein the management of interference from different radio systems is key to retrieving useful information Bică and Koivunen (2018). When the received radar signal reflected

off from the target is overlaid with communications messages occupying the same bandwidth, the knowledge of respective waveforms is useful in designing matched filters to extract the two signals Paul et al. (2016). Usually, the radar signal is known and the goal of the radar receiver is to extract unknown target parameters from the received signal. In a typical communications receiver, the channel is known (or estimated) and the unknown transmit message is of interest in the receive processing. However, these assumptions do not extend to a general scenario. For example, in a passive Sedighi et al. (2021) or multi-static Dokhanchi et al. (2019) radar system, the receiver may not have complete information about the radar transmit waveform. Similarly, in communications over dynamic channels such as millimeter-wave Mishra et al. (2019) or Terahertz-band Elbir et al. (2021), the channel coherence times are very short. As a result, the channel state information changes rapidly and cannot be determined *a priori*. Moreover, different transmitters are employed in the spectral coexistence scenario leading to a lack of synchronized arrivals of both signals at a common receiver.

In this thesis, we consider the aforementioned general spectral coexistence scenario, where both radar and communications channels and their respective transmit signals are unknown to the common receiver. We model the extraction of all four of these quantities as a *dual-blind deconvolution* (DBD) problem. This formulation is a variant of *blind deconvolution* (BD) — a longstanding problem that occurs in a variety of engineering and scientific applications, such as astronomy, image deblurring, system identification, and optics — where two unknown signals are estimated from the observation of their convolution Jefferies and Christou (1993); Ayers and Dainty (1988); Abed-Meraim et al. (1997); Ahmed and Demanet (2018). This problem is ill-

Table 8
Comparison with Prior Art

Problem	Measurements	Unknown parameters	Algorithm
Multi-dimensional SR[a]Xu et al. (2014)	$[\mathbf{y}]_{\widetilde{n}} = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}]_\ell e^{-\mathrm{j}2\pi \mathbf{f}_\ell^T \mathbf{n}}$	$\boldsymbol{\alpha}, \mathbf{f}_\ell$	Multi-dimensional ANM
1-D BD[b] Chi (2016)	$[\mathbf{y}]_n = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}]_\ell e^{-\mathrm{j}2\pi n [\boldsymbol{\tau}]_\ell} [\mathbf{g}]_n$	$\mathbf{g}_n, \boldsymbol{\alpha}, \boldsymbol{\tau}$	ANM
Blind 1-D SR[c]Yang et al. (2016a)	$[\mathbf{y}]_n = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}]_\ell e^{-\mathrm{j}2\pi n [\boldsymbol{\tau}]_\ell} [\mathbf{g}_\ell]_n$	$\mathbf{g}_\ell, \boldsymbol{\alpha}, \boldsymbol{\tau}$	ANM
SR radar[d]Heckel et al. (2016)	$\mathbf{y} = \mathbf{G} \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}]_\ell \mathbf{a}(\mathbf{r}_\ell)$	$\mathbf{r}_\ell, \boldsymbol{\alpha}$	ANM
Blind 2-D SR[e]Suliman and Dai (2022)	$[\mathbf{y}]_{\widetilde{n}} = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}]_\ell \mathbf{a}(\mathbf{r}_\ell)^H \mathbf{G}_{\widetilde{n}} \mathbf{g}_\ell$	$\boldsymbol{\alpha}, \mathbf{r}_\ell, \mathbf{g}_\ell$	ANM
BdM[f] Ling and Strohmer (2017)	$\mathbf{y} = \sum_{i=1}^{I} \mathrm{diag}(\mathbf{B}_i \mathbf{h}_i) \mathbf{A}_i \mathbf{x}_i$	$\mathbf{h}_\ell, \mathbf{x}_\ell$	Nuclear norm minimization
1-D SR-dM[g]Li and Chi (2019)	$[\mathbf{y}]_n = \sum_{i=1}^{I} [\mathbf{g}]_{i,n} \cdot \left(\sum_{\ell=0}^{L_i-1} [\boldsymbol{\alpha}_i]_\ell e^{-j2\pi n [\boldsymbol{\tau}_i]_\ell} \right)$	$\boldsymbol{\alpha}_i, \boldsymbol{\tau}_i$	SoAN minimization
This paper (DBD)[h]	$[\mathbf{y}]_{\widetilde{n}} = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell \mathbf{b}_n^H \mathbf{u} e^{-\mathrm{j}2\pi n \tau_c} e^{-\mathrm{j}2\pi (n [\boldsymbol{\tau}_r]_\ell + p [\mathbf{v}_r]_\ell)} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q \mathbf{d}_{\widetilde{n}}^H \mathbf{v} e^{-\mathrm{j}2\pi (n [\boldsymbol{\tau}_c]_q + p [\mathbf{v}_c]_q)}$	$\boldsymbol{\alpha}_r, \mathbf{u}, \boldsymbol{\tau}_r, \boldsymbol{\nu}_r, \boldsymbol{\alpha}_c, \mathbf{v}, \boldsymbol{\tau}_c, \boldsymbol{\nu}_c$	SoMAN minimization

[a] $\mathbf{f}_\ell$ is a d-dimensional vector of unknown off-the-grid poles, $[\boldsymbol{\alpha}]_\ell$ is complex amplitude, $\widetilde{n}$ is a linear indexing sequence of the multidimensional measurement, and L is the number of frequencies.
[b] Vector $\mathbf{g}$ is the transmit signal, $[\boldsymbol{\tau}]_\ell$ is the unknown off-the-grid pole.
[c] This problem assumes a total of L unknown signals of which the ℓ-th signal and the n-th sample is $[\mathbf{g}_\ell]_n$.
[d] $\mathbf{G}$ denotes the sensing Gabor matrix and $\mathbf{a}(\mathbf{r})$ are the atoms containing 2-D where $\mathbf{r}_\ell = [[\boldsymbol{\tau}]_\ell, [\boldsymbol{\nu}]_\ell]$, $[\boldsymbol{\tau}]_\ell$ denote time-delay and $[\boldsymbol{\nu}]_\ell$ Doppler-velocity.
[e] $\mathbf{G}_n$ is the sensing Gabor matrix.
[f] Matrices $\mathbf{B}_\ell, \mathbf{A}_\ell$ model the unknown channels. The unknown signal vectors are $\mathbf{h}_\ell$ and $\mathbf{x}_\ell$.
[g] L_i is the number of off-the-grid parameters of i-th signal $[\mathbf{g}]_{i,n}$.
[h] See Section 6.2 for details. Briefly, $\mathbf{b}_n$ and $\mathbf{d}_{\widetilde{n}}$ are columns of a representation basis; $\mathbf{u}$ and $\mathbf{v}$ are low-dimensional representations of the radar and communications signal, respectively. Note that, from (55), the index $\tilde{n}$ in the communications signal component $\mathbf{d}_{\widetilde{n}}$ depends on the pulse index p. The radar signal component $\mathbf{b}_n$ only depends on the discrete-time index n.

posed and, in general, structural constraints are imposed on signals to derive recovery algorithms with provable performance guarantees. The underlying ideas in these techniques are based on compressed sensing and low-rank matrix recovery, wherein signals lie in the low-dimensional random subspace and/or in high signal-to-noise ratio (SNR) regime Lee et al. (2016); Li et al. (2019b); Ahmed et al. (2013); Kuo et al. (2019); Ahmed and Demanet (2018).

6.1. Prior Art

Unlike prior works on spectral coexistence (see, e.g., Chiriyath et al. (2017); Mishra et al. (2019) and references therein), we examine the overlaid radar-communications signal as the ill-posed DBD problem. Previously, Farshchian and Selesnick (2016) studied *dual deconvolution*

problem which assumed that the radar transmit signal and communications channel were known. Our approach toward a more challenging DBD problem is inspired by some recent works Chi (2016); Yang et al. (2016a) that have analyzed the basic BD for off-the-grid sparse scenarios. The radar and communications channels are usually sparse Mishra and Eldar (2019), more so at higher frequency bands, and their parameters are continuous-valued Mishra et al. (2015). Hence, sparse reconstruction in off-the-grid or continuous parameter domain through techniques based on atomic norm minimization (ANM) Tang et al. (2013) is appropriate for our application. In BD, sparsity and subspace constraints over the signals have shown great ability in reducing the search space to obtain unique solutions Li et al. (2017); Zhang et al. (2019).

Our work has close connections with a rich heritage of research in several related signal processing problems; Table 8 compares our DBD problem with some important prior works. Our approach is based on the semidefinite program (SDP) derived for ANM-based spectral super-resolution (SR) of a high-dimensional signal in Xu et al. (2014). This formulation has previously been extended to bi-variate ANM for estimating delay and Doppler frequencies of targets from received radar signal samples Heckel et al. (2016). The ANM-based recovery was also employed in the 1-D BD formulation of Chi (2016). In general, the absence of knowledge about the transmit signal requires one to define a structural assumption about the signal, e.g., a low-dimensional subspace representation based on incoherence and isotropy properties Candès and Plan (2011). This structural assumption was also used in Yang et al. (2016b), which considered blind SR with multiple modulating unknown waveforms. A 2-D ANM-based blind SR with unknown transmit signals was addressed in Suliman and Dai (2021) and extended to 3-D in Suliman and Dai (2019).

The ANM has also been useful for line spectrum denoising Li et al. (2019a). However, none of the aforementioned works deal with mixed radar-communications signal structures.

A closely related problem is that of *blind demixing* (BdM) where the sampled signal is a mixture of several BD problems Ling and Strohmer (2017); Li and Chi (2019), all of which have a *similar* structure. Per Table 8, for $I = 1$ and 2, the BdM reduces to single and double BDs (with uniform structure in each convolution), respectively. However, our DBD problem has a different structure in each convolution operation (e.g., the communications messages change with each transmission). Further, the recovery guarantees previously derived for BdM do not consider special signal structures of overlaid radar-communications. The recovery algorithms in some BdM formulations Ling and Strohmer (2017) are based on nuclear norm minimization over a set of rank-one *lifted* matrices. In Dong et al. (2018), the BdM recovery problem had multiple non-convex rank-one constraints to deal with multiple low-latency communications systems. These BdM studies do not employ an atomic decomposition over the transfer function and are not well-suited for the estimation of off-the-grid parameters.

Our Contributions. Preliminary results of this work appeared in our conference publication Vargas et al. (2022), where only radar and communications channel parameters were estimated using SoMAN minimization. In this thesis, we apply our methods also to estimate the communications messages and radar waveforms, provide detailed theoretical guarantees, include additional numerical validations, address multiple emitter scenarios, and consider unsynchronized and noisy signals. Our main contributions in this thesis are:

1) Solving spectral coexistence as a structured continuous-valued DBD. Contrary to the above-

mentioned BD, BdM, and SR problems, our DBD is based on practical radar-communications scenarios that lend a special structure to the signal as follows. Both communications and radar channels are sparse and their parameters are continuous-valued; the radar channel is a function of unknown target parameters such as range-time and Doppler velocities. A similar delay-Doppler channel is considered for communications paths. Whereas the radar transmits a train of pulses, the communications transmitter emits a multi-carrier signal. Our DBD formulation incorporates all these restrictions/structures.

2) Novel ANM-based recovery. The continuous-valued nature of radar and communications channel parameters requires using ANM-based approaches. In particular, we cast the recovery problem as the minimization of the sum of multivariate atomic norms (SoMAN). Previously, the sum of 1-D atomic norms (SoAN) was minimized to solve a joint (non-blind) demixing (dM) and SR problem in Li and Chi (2019). However, this 1-D SoAN formulation is non-blind and hence does not trivially extend to our multi-variable DBD. Further, the guarantees in the 1-D SoAN are provided for the same structure of multiple signals while, in our formulation, the structures of radar and communications signals are not identical. Moreover, in the sequel, we also generalize SoMAN to include synchronization errors for scenarios in which radar and communications systems have different clocks.

3) Strong theoretical guarantees. Our method is backed up by strong theoretical performance guarantees. We prove that the minimum number of samples required for perfect recovery scale logarithmically with the maximum of the radar targets and communications paths. This is counter-intuitive because the recovery in radar and communications receivers generally depends on the

number of unknown scatterers and paths.

4) Generalization to practical scenarios. We consider several practical issues to generalize our approach (see Supplementary Material). We show that, in the presence of noise, our formulation is applicable by adding a regularization term to the dual problem. We address the unsynchronized transmission of radar and communications signals by multiplying the atomic set by a lag-dependent diagonal matrix. We also show that the n-tuple BD problem comprising multiple sources of radar and communications signals is addressed by minimizing the addition of multiple SoMANs. Finally, we handle the practical case of an unequal duration of radar pulse and communications symbol by formulating the SoMAN problem to compute multiple coefficient vectors for communications signals. Note that our recent follow-up work in Jacome et al. (2022) also considers the multi-antenna DBD problem.

5) Extensive numerical experiments: We comprehensively validated our SoMAN approach for several scenarios, including randomly distributed targets, unequal numbers of radar targets and communications paths, and closely-spaced parameters. To validate the provided theoretical guarantees numerically, We provide statistical performance by varying system parameters and validate the theoretical result predicted by Theorem 2. Finally, we refer the reader to the Supplementary Material that compiles several numerical results for all the practical scenarios mentioned above.

Notations. Throughout this chapter, we reserve boldface lowercase, boldface uppercase, and calligraphic letters for vectors, matrices, and index sets, respectively. The notation $[\mathbf{x}]_i$ indicates the i-th entry of the vector $\mathbf{x}$ and $[\mathbf{X}]_i$ the i-th row of $\mathbf{X}$. We denote the transpose, conjugate, and Hermitian by $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$, respectively. The identity matrix of size $N \times N$ is $\mathbf{I}_N$. $||\cdot||_p$

is the ℓ_p norm. The notation $\mathrm{Tr}\{\cdot\}$ is the trace of the matrix, $|\cdot|$ is the cardinality of a set, $\mathrm{supp}(\dot{)}$ is the support set of its argument, $\mathbb{E}[\cdot]$ is the statistical expectation function, and $\mathbb{P}$ denotes the probability. The functions max and min output the maximum and minimum values of their arguments, respectively. The sign function is defined as $\mathrm{sign}(c) = \frac{c}{|c|}$. The function $\mathrm{diag}(\cdot)$ outputs a diagonal matrix with the input vector along its main diagonal. The block diagonal matrix with diagonal matrix elements $\mathbf{X}_1, \ldots, \mathbf{X}_P$ is $\mathbf{X} = \mathrm{blockdiag}[\mathbf{X}_1, \ldots, \mathbf{X}_P]$.

6.2. Signal Model

Consider a pulse Doppler radar that transmits a train of P_r uniformly-spaced pulses $s(t)$ with a pulse repetition interval (PRI) T_r; its reciprocal $1/T_r$ is the pulse repetition frequency (PRF). The transmit signal of the radar is

$$x_r(t) = \sum_{p=0}^{P_r-1} s(t - pT_r),\ 0 \le t \le P_r T_r. \tag{37}$$

The entire duration of P_r pulses is known as the coherent processing interval (CPI). The pulse $s(t)$ is a time-limited baseband function, whose continuous-time Fourier transform (CTFT) is $S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} \mathrm{d}t$. It is common to assume that most of the radar signal's energy lies within the frequencies $\pm B/2$, where B denotes the effective signal bandwidth, i.e., $s(t) \approx \int_{-B/2}^{B/2} S(f) e^{j2\pi ft} \mathrm{d}f$.

The radar channel or *target scene* consists of L non-fluctuating point-targets, according to the Swerling-I target model Skolnik (2008). The transmit signal is reflected back by the targets toward the radar receiver. The unknown target parameter vectors are $\boldsymbol{\alpha}_r \in \mathbb{C}^L$, $\overline{\boldsymbol{\tau}}_r \in \mathbb{R}^L$ and $\overline{\boldsymbol{\nu}}_r \in \mathbb{R}^L$, where the ℓ-th target is characterized by: time delay $[\overline{\tau}_r]_\ell$, which is linearly proportional to the

target's range *i.e.* $[\overline{\tau}_r]_\ell = 2d_\ell/c$ where c is the speed of light and d_ℓ is the target range; Doppler frequency $[\overline{\nu}_r]_\ell$, proportional to the target's radial velocity *i.e.* $[\overline{\nu}_r]_\ell = 4\pi f_c v_\ell/c$ where f_c is the carrier frequency and v_ℓ is the radial velocity; and complex amplitude $[\alpha_r]_\ell$ that models the path loss and target reflectivity. The target locations are defined with respect to the polar coordinate system of the radar and their range and Doppler are assumed to lie in the unambiguous time-frequency region, i.e., the time delays (τ) are no longer than the PRI, and Doppler frequencies (ν) are up to the PRF. The radar channel is the product of delta functions as

$$h_r(\nu,\tau) = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell \, \delta(\nu - [\overline{\boldsymbol{\nu}}_r]_\ell)\delta(\tau - [\overline{\boldsymbol{\tau}}_r]_\ell). \tag{38}$$

The delay-Doppler representation of the channel is obtained by taking the Fourier transform along the time-axis of the time-delay representation $h_r(t,\tau)$ as

$$h_r(t,\tau) = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell \, \delta(\tau - [\overline{\boldsymbol{\tau}}_r]_\ell) e^{-\mathrm{j}2\pi[\overline{\boldsymbol{\nu}}_r]_\ell t}. \tag{39}$$

For communications, the transmit signal $x_c(t)$ comprises P_c messages modeled as orthogonal frequency-division multiplexing (OFDM) signals with K equi-bandwidth sub-carriers which together occupy a bandwidth of B with a symbol duration of T_c and have a frequency separation of Δf Cimini (1985). Hence, the m-th message is $x_m(t) = \sum_{k=0}^{K-1} [\mathbf{g}_m]_k e^{\mathrm{j}2\pi k\Delta f t}$, where $[\mathbf{g}_m]_k$ is m-th complex symbol modulated onto the k-th sub-carrier frequency $f_k = k\Delta f$. The communications

transmitted signal is

$$x_c(t) = \sum_{m=0}^{P_c-1} x_m(t - mT_c),\ 0 \le t \le P_c T_c. \tag{40}$$

The communications channel comprises Q propagation paths characterized by their attenuation coefficients, delay, and frequency shifts, respectively, encapsulated in the parameter vectors $\boldsymbol{\alpha}_c \in \mathbb{C}^Q$, $\overline{\boldsymbol{\tau}}_c \in \mathbb{R}^Q$, and $\overline{\boldsymbol{\nu}} \in \mathbb{R}^Q$. The delay-Doppler representation of the communications channel is

$$h_c(\nu, \tau) = \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q \delta(\nu - [\overline{\nu}_c]_q)\delta(\tau - [\overline{\tau}_c]_q), \tag{41}$$

from which we obtain the time-delay representation as

$$h_c(t, \tau) = \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q \delta(\tau - [\overline{\tau}_c]_q) e^{-\mathrm{j}2\pi[\overline{\nu}_c]_q t}. \tag{42}$$

Overlaid Radar-Communications Receiver. In a spectral coexistence scenario, both radar and communications operate in the same spectrum (Fig. 32). At the receiver, the baseband signal is a superposition of radar and communications signals as

$$\begin{aligned} y(t) &= x_r(t) * h_r(t) + x_c(t) * h_c(t), \\ &= \int_{-\infty}^{\infty} h_r(t, \tau) x_r(t - \tau) + h_c(t, \tau) x_c(t - \tau) d\tau, \end{aligned} \tag{43}$$

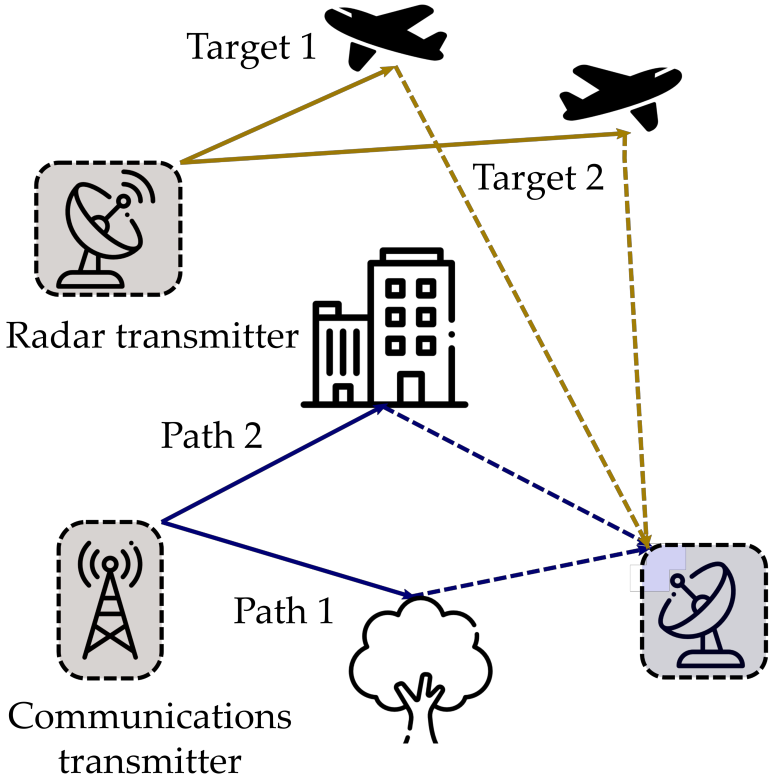

Figure 32. In a spectral coexistence scenario, a joint receiver admits the overlaid radar and communications signals.

where $0 \leq t \leq \max(P_r T_r, P_c T_c)$. Substituting the radar and communications channel models of, respectively, (39) and (42) in (43) yields

$$y(t) = \sum_{\ell=0}^{L-1} [\alpha_r]_\ell x_r(t - [\overline{\tau}_r]_\ell) e^{-\mathrm{j}2\pi[\overline{\nu}_r]_\ell t} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q x_c(t - [\overline{\tau}_c]_q) e^{-\mathrm{j}2\pi[\overline{\nu}_c]_q t}. \tag{44}$$

Substituting expressions of $x_r(t)$ and $x_c(t)$ into (44) leads to

$$\begin{aligned} y(t) &= \sum_{p=0}^{P_r-1}\sum_{\ell=0}^{L-1}[\alpha_r]_\ell s(t-pT_r-[\overline{\tau}_r]_\ell)e^{-\mathrm{j}2\pi[\overline{\nu}_r]_\ell t} \\ &+ \sum_{m=0}^{P_c-1}\sum_{q=0}^{Q-1}[\boldsymbol{\alpha}_c]_q x_m(t-mT_c-[\overline{\tau}_c]_q)e^{-\mathrm{j}2\pi[\overline{\nu}_c]_q t} \qquad (45) \\ &\approx \sum_{p=0}^{P_r-1}\sum_{\ell=0}^{L-1}[\boldsymbol{\alpha}_r]_\ell s(t-pT_r-[\overline{\boldsymbol{\tau}}_r]_\ell)e^{-\mathrm{j}2\pi[\overline{\boldsymbol{\nu}}_r]_\ell pT_r} \\ &+ \sum_{m=0}^{P_c-1}\sum_{q=0}^{Q-1}[\boldsymbol{\alpha}_c]_q x_m(t-mT_c-[\overline{\boldsymbol{\tau}}_c]_q)e^{-\mathrm{j}2\pi[\overline{\boldsymbol{\nu}}_c]_q mT_c}, \qquad (46) \end{aligned}$$

where the last approximation results from the assumptions $[\overline{\boldsymbol{\nu}}_r]_\ell T_r \ll 1$ and $[\overline{\boldsymbol{\nu}}_c]_q T_c \ll 1$ Zheng and Wang (2017); Muns et al. (2019) such that the phase rotation between each radar pulse or OFDM message in the communications is assumed constant during a CPI. Then, for every $P < t < P+1$, the values $[\overline{\nu}_r]_\ell tT$ and $[\overline{\nu}_c]_\ell tT$ remain constant. The change occurs only during the next pulse. Thus, we replace tT with the index p.

For the sake of simplicity, assume that the number of pulses equals the number of messages, i.e., $P_r = P_c = P$ and the PRI T_r is the same as the message duration T_c *i.e.* $T_r = T_c = T$ (Fig. 33). We also show in Appendix A that our formulation is applicable to the case when $T_r \neq T_c$. Following Fig. 33, the received signal in (46) becomes

$$y(t) = \sum_{p=0}^{P-1}\widetilde{y}_p(t), \qquad (47)$$

where

$$\widetilde{y}_p(t) = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell s(t - pT - [\overline{\boldsymbol{\tau}}_r]_\ell) e^{-\mathrm{j}2\pi[\overline{\nu}_r]_\ell pT} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q x_p(t - pT - [\overline{\tau}_c]_q) e^{-\mathrm{j}2\pi[\overline{\nu}_c]_q pT},\ 0 \le t \le PT. \tag{48}$$

Express the measurements in terms of shifted signals $y_p(t) = \widetilde{y}_p(t + pT)$, where the signals $\widetilde{y}_p(t + pT)$ are time-aligned with $y_0(t)$. Then, the signal $y_0(t)$ and the shifted signals contain the same set of parameters.

The CTFT of $y_p(t)$ produces

$$Y_p(f) = \int_{pT}^{pT+T} e^{-\mathrm{j}2\pi ft} \left(\sum_{\ell=0}^{L-1} [\alpha_r]_\ell s(t - [\overline{\boldsymbol{\tau}}_r]_\ell) e^{-\mathrm{j}2\pi[\overline{\nu}_r]_\ell pT} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q x_p(t - [\overline{\tau}_c]_q) e^{-\mathrm{j}2\pi[\overline{\nu}_c]_q pT} \right) dt, \tag{49}$$

which is a sum of the Fourier transforms of the p-th received symbol of the communications signal and the p-th received pulse of the radar system. In the integral above, computing the radar term and expanding $x_p(t)$ in the communications part yields

$$Y_p(f) = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell S(f) e^{-\mathrm{j}2\pi f[\overline{\tau}_r]_\ell} e^{-\mathrm{j}2\pi[\overline{\nu}_r]_\ell pT} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q e^{-\mathrm{j}2\pi f[\overline{\tau}_c]_q} e^{-\mathrm{j}2\pi[\overline{\nu}_c]_q pT} \sum_{k=0}^{K-1} [\mathbf{g}_p]_k \times \int_{pT}^{pT+T} e^{-\mathrm{j}2\pi ft} e^{\mathrm{j}2\pi k\Delta ft} dt. \tag{50}$$

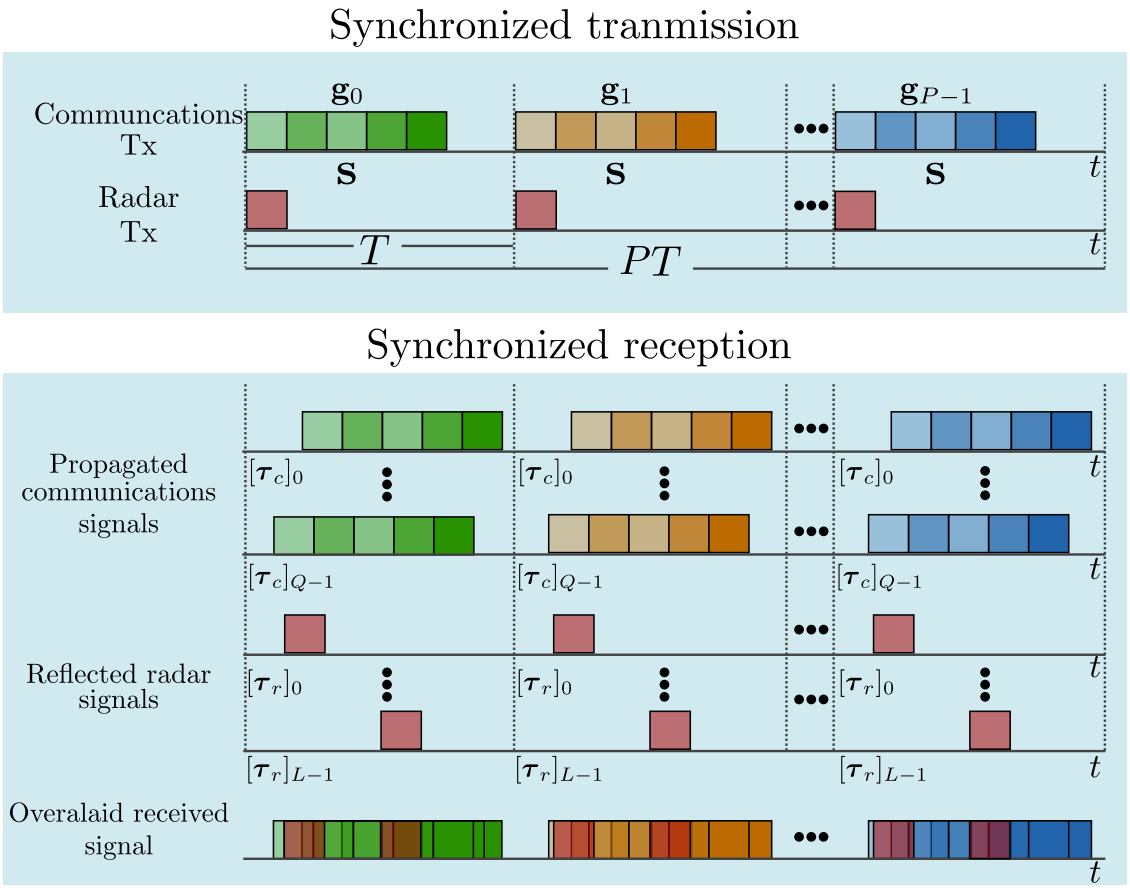

Figure 33. Sequence of transmission and reception of radar and communications signals. In synchronized transmission, the radar transmits each pulse simultaneously as the communications transmitter sends out messages. The common receiver of both systems simultaneously admits radar signals reflected off L targets and communications messages through Q paths. For the unsynchronized transmission, see Fig. 1 in the supplementary material.

Sampling (50) uniformly at the frequencies $f_n = \frac{Bn}{M} = n\Delta f$, with $n = -N,\ldots,N$, $M = 2N+1$, and assuming that $M = K$, i.e., sampling in the frequency domain at the OFDM separation

frequency Δf Zheng and Wang (2017) gives

$$Y_p(f_n) = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell S(f_n) e^{-\mathrm{j}2\pi n \Delta f [\overline{\tau}_r]_\ell} e^{-\mathrm{j}2\pi [\overline{\nu}_r]_\ell pT} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q e^{-\mathrm{j}2\pi n \Delta f [\overline{\tau}_c]_q} e^{-\mathrm{j}2\pi [\overline{\boldsymbol{\nu}}_c]_q pT} \sum_{k=0}^{K-1} [\mathbf{g}_p]_k \int_{pT}^{pT+T} e^{\mathrm{j}2\pi \Delta f (k-n)t} dt. \tag{51}$$

Using the property $\int_0^T e^{-\mathrm{j}2\pi\Delta f(k-n)t} dt = 0$ with $k \neq n$ and concatenating the values $Y_p(f_n)$ in the vector $\mathbf{y}_p$, i.e., $[\mathbf{y}_p]_{n+N} = Y_p(f_n)$ for $n = -N, \cdots, N$, the discrete values of the Fourier transform (50) are

$$[\mathbf{y}_p]_{n+N} = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell [\mathbf{s}]_{n+N} e^{-\mathrm{j}2\pi (n[\boldsymbol{\tau}_r]_\ell + p[\boldsymbol{\nu}_r]_\ell)} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q [\mathbf{g}_p]_{n+N} e^{-\mathrm{j}2\pi (n[\boldsymbol{\tau}_c]_q + p[\boldsymbol{\nu}_c]_q)}, \tag{52}$$

where $[\boldsymbol{\tau}_r]_\ell = \frac{[\overline{\boldsymbol{\tau}}_r]_\ell}{T} \in [0,1]$ and $[\boldsymbol{\tau}_c]_\ell = \frac{[\overline{\boldsymbol{\tau}}_c]_\ell}{T} \in [0,1]$ are the normalized delays; $[\boldsymbol{\nu}_r]_\ell = \frac{[\overline{\boldsymbol{\nu}}_r]_\ell}{\Delta f} \in [0,1]$ and $[\boldsymbol{\nu}_c]_\ell = \frac{[\overline{\boldsymbol{\nu}}_c]_\ell}{\Delta f} \in [0,1]$ are the normalized Doppler frequencies; the absolute values of $[\boldsymbol{\alpha}_r]_\ell$ and $[\boldsymbol{\alpha}_c]_q$ are unit-norm after normalizing the signal by its magnitude, i.e. $|[\boldsymbol{\alpha}_r]_\ell| = |[\boldsymbol{\alpha}_c]_q| = 1$; and $[\mathbf{s}]_{n+N} = S(f_n)$.

We collect all discrete Fourier transform samples in the measurement vector

$$\mathbf{y} = \left[\mathbf{y}_0^T, \cdots, \mathbf{y}_{P-1}^T\right]^T. \tag{53}$$

Then, all samples of P pulses (and communications messages) are

$$[\mathbf{y}]_{\widetilde{n}} = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell [\mathbf{s}]_n e^{-\mathrm{j}2\pi(n[\boldsymbol{\tau}_r]_\ell + p[\boldsymbol{\nu}_r]_\ell)} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q [\mathbf{g}]_{\widetilde{n}} e^{-\mathrm{j}2\pi(n[\boldsymbol{\tau}_c]_q + p[\boldsymbol{\nu}_c]_q)}, \tag{54}$$

where

$$\widetilde{n} = n + N + Mp, \tag{55}$$

is a linear indexing sequence of the multidimensional measurement with $n = -N, \ldots, N$ and $p = 0, \ldots, P-1$, and $\mathbf{g} = [\mathbf{g}_0^T, ..., \mathbf{g}_{P-1}^T]$. Note that $[\mathbf{y}]_{\widetilde{n}} = [\mathbf{y}_p]_{n+N}$ and $[\mathbf{g}]_{\widetilde{n}} = [\mathbf{g}_p]_{n+N}$.

Remark 1. In general, a cyclic prefix (CP) of length K_{cp} is added to each modulated OFDM symbol for mitigating inter-symbol interference (ISI). The OFDM symbol duration then becomes $T_{\mathrm{cp}} + T_c$, where T_{cp} denote the CP duration. Usually, T_{cp} is chosen such that the maximum delay $[\boldsymbol{\tau}_c]_{q_{\max}} < T_{\mathrm{cp}} < T_c$. While we did not consider CP in the communications model above, our approach could be easily modified to include CP by changing the expression $x_m(t) = \sum_{k=0}^{K-1} [\mathbf{g}_m]_k e^{\mathrm{j}2\pi k\Delta f t}$ to $x_m(t) = \sum_{k=K-K_{\mathrm{cp}}+1}^{K-1} [\mathbf{g}_m]_k e^{\mathrm{j}2\pi k\Delta f t}$. The consequence of this change is that the communications messages $\mathbf{g}_m$ is of a larger dimension. This requires a different low-dimensional subspace representation described in the next subsection. This case also resembles the unequal PRI and symbol duration scenario.

Representation in low-dimensional space. Our goal is to estimate the sets of parameters $[\boldsymbol{\alpha}_r]_\ell, [\boldsymbol{\tau}_r]_\ell, [\boldsymbol{\nu}_r]_\ell, [\boldsymbol{\alpha}_c]_q, [\boldsymbol{\tau}_c]_q$ and $[\boldsymbol{\nu}_c]_q$, without knowing the radar pulse signal samples $\mathbf{s}$ and the

communications symbols $\{\mathbf{g}_p\}_{p=0}^{P-1}$. This is an ill-posed problem because of a large number of unknown variables and fewer measurements. We now represent the unknown transmit radar and communications signals $\mathbf{s}$ and $\mathbf{g}_p$ in a low-dimensional subspace spanned by the columns of known random matrices $\mathbf{B} \in \mathbb{C}^{M\times J}$ and $\mathbf{D}_p \in \mathbb{C}^{M\times J}$, respectively. Therefore, $\mathbf{s} = \mathbf{B}\mathbf{u}$ and $\mathbf{g}_p = \mathbf{D}_p\mathbf{v}_p$, where $\mathbf{u} \in \mathbb{C}^J$ and $\mathbf{v}_p \in \mathbb{C}^J$ are the unknown low-dimensional coefficient vector of the radar and communications signal respectively. This reduces the number of unknown variables because $J \ll M$. The basis matrices are

$$\mathbf{B} = [\mathbf{b}_{-N}, \ldots, \mathbf{b}_N]^H, \mathbf{b}_n \in \mathbb{C}^J, \tag{56a}$$

$$\mathbf{D}_p = \left[[\mathbf{D}_p]_{-N}, \cdots, [\mathbf{D}_p]_N\right]^H, \ [\mathbf{D}_p]_n \in \mathbb{C}^J. \tag{56b}$$

Rewrite the set of symbols $\mathbf{g}$ compactly as

$$\mathbf{g} = \mathbf{D}\mathbf{v}, \quad \mathbf{D} = [\mathbf{d}_0, \cdots, \mathbf{d}_{MP-1}]^H = \text{blockdiag}\left[\mathbf{D}_0, \cdots, \mathbf{D}_{P-1}\right]. \tag{57}$$

The low-dimensional representation of $\mathbf{s}$ and $\mathbf{g}$ is inspired by radar and communications applications. For instance, this structure has been validated previously for a sensing-based SR of complex exponentials that are modulated with unknown waveforms Yang et al. (2016a). In the case of wireless communications, this assumption has been verified for multipath channels with unknown modulation Ahmed et al. (2013) and multi-user communications Luo and Giannakis (2006). Using

the basis decompositions (56a)-(56b) and (57), the signal model in (54) becomes

$$[\mathbf{y}]_{\widetilde{n}} = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell \mathbf{b}_n^H \mathbf{u} e^{-\mathrm{j}2\pi(n[\boldsymbol{\tau}_r]_\ell + p[\boldsymbol{\nu}_r]_\ell)} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q \mathbf{d}_{\widetilde{n}}^H \mathbf{v} e^{-\mathrm{j}2\pi(n[\boldsymbol{\tau}_c]_q + p[\boldsymbol{\nu}_c]_q)}. \tag{58}$$

Denote the channel vectors as

$$\mathbf{h}_r = \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell \mathbf{a}(\mathbf{r}_\ell), \tag{59}$$

and

$$\mathbf{h}_c = \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q \mathbf{a}(\mathbf{c}_q), \tag{60}$$

where $\mathbf{r}_\ell = [[\boldsymbol{\tau}_r]_\ell, [\boldsymbol{\nu}_r]_\ell]^T$ and $\mathbf{c}_q = [[\boldsymbol{\tau}_c]_q, [\boldsymbol{\nu}_c]_q]^T$ such that

$$\begin{aligned} &\mathbf{a}(\mathbf{r}_\ell) \\ &= \mathbf{a}([[\boldsymbol{\tau}_r]_\ell, [\boldsymbol{\nu}_r]_\ell]^T) \\ &= \left[e^{\mathrm{j}2\pi([\boldsymbol{\tau}_r]_\ell(-N) + [\boldsymbol{\nu}_r]_\ell(0))}, \ldots, e^{\mathrm{j}2\pi([\boldsymbol{\tau}_r]_\ell(N) + [\boldsymbol{\nu}_r]_\ell(P-1))}\right] \in \mathbb{C}^{MP}, \end{aligned} \tag{61}$$

and

$$\begin{aligned}
&\mathbf{a}(\mathbf{c}_q) \\
&= \mathbf{a}(\left[[\boldsymbol{\tau}_c]_q, [\boldsymbol{\nu}_c]_q\right]^T) \\
&= \left[e^{\mathrm{j}2\pi([\boldsymbol{\tau}_c]_q(-N)+[\boldsymbol{\nu}_c]_q(0))}, \ldots, e^{\mathrm{j}2\pi([\boldsymbol{\tau}_c]_q(N)+[\boldsymbol{\nu}_c]_q(P-1))}\right] \in \mathbb{C}^{MP}.
\end{aligned} \tag{62}$$

Then, the model in (58) reduces to

$$[\mathbf{y}]_{\widetilde{n}} = \mathbf{h}_r^H \mathbf{e}_{\widetilde{n}} \mathbf{b}_n^H \mathbf{u} + \mathbf{h}_c^H \mathbf{e}_{\widetilde{n}} \mathbf{d}_{\widetilde{n}}^H \mathbf{v}, \tag{63}$$

or equivalently

$$\begin{aligned}
[\mathbf{y}]_{\widetilde{n}} &= \langle \mathbf{Z}_r, \mathbf{G}_{\widetilde{n}}^H \rangle + \langle \mathbf{Z}_c, \mathbf{A}_{\widetilde{n}}^H \rangle, \\
&= \mathrm{Tr}(\mathbf{G}_{\widetilde{n}} \mathbf{Z}_r) + \mathrm{Tr}(\mathbf{A}_{\widetilde{n}} \mathbf{Z}_c),
\end{aligned} \tag{64}$$

with the *waveform-channel matrices*

$$\mathbf{Z}_r = \mathbf{u}\mathbf{h}_r^H \in \mathbb{C}^{J \times MP}, \tag{65}$$

$$\mathbf{Z}_c = \mathbf{v}\mathbf{h}_c^H \in \mathbb{C}^{PJ \times MP}, \tag{66}$$

$\mathbf{G}_{\widetilde{n}} = \mathbf{e}_{\widetilde{n}} \mathbf{b}_n^H \in \mathbb{C}^{MP \times J}$ and $\mathbf{A}_{\widetilde{n}} = \mathbf{e}_{\widetilde{n}} \mathbf{d}_{\widetilde{n}}^H \in \mathbb{C}^{MP \times PJ}$, where $\mathbf{e}_{\widetilde{n}}$ is the v-th canonical vector of $\mathbb{R}^{MP}$.

Define the linear operators $\aleph_r : \mathbb{C}^{J\times MP} \to \mathbb{C}^{MP}$ and $\aleph_c : \mathbb{C}^{PJ\times MP} \to \mathbb{C}^{MP}$ such that

$$[\aleph_r(\mathbf{Z}_r)]_{\widetilde{n}} = \mathrm{Tr}(\mathbf{G}_{\widetilde{n}}\mathbf{Z}_r),$$

and

$$[\aleph_c(\mathbf{Z}_c)]_{\widetilde{n}} = \mathrm{Tr}(\mathbf{A}_{\widetilde{n}}\mathbf{Z}_c).$$

The measurement vector now becomes

$$\mathbf{y} = \aleph_r(\mathbf{Z}_r) + \aleph_c(\mathbf{Z}_c), \tag{67}$$

where the matrices $\mathbf{Z}_r$ and $\mathbf{Z}_c$ encapsulate all the unknown variables to be estimated, i.e., channel vectors $\mathbf{h}_r$ and $\mathbf{h}_c$ along with the signal coefficients $\mathbf{u}$ and $\mathbf{g}$. This representation of the unknown variables in 1-rank matrix is similar to the *lifting* trick previously employed in other ill-posed problems such as BD Ahmed et al. (2013); Chi (2016) and phase retrieval Candes et al. (2013). Given that $\mathbf{Z}_r = \mathbf{u}\mathbf{h}_r^H$ ($\mathbf{Z}_c = \mathbf{v}\mathbf{h}_c^H$), one could estimate the vectors $\mathbf{u}$ and $\mathbf{h}_r$ ($\mathbf{v}$ and $\mathbf{h}_c$) up to a scaling factor as the left and right singular vectors of $\mathbf{Z}_r$ ($\mathbf{Z}_c$).

6.3. Parameter Recovery via SoMAN minimization

We model the radar and communications trails of the received signal in (67) as positive linear combinations of constituent *atoms*, which are lifted rank-one matrices $\mathbf{ua}(\mathbf{r})^H$ and $\mathbf{va}(\mathbf{c})^H$

belonging to the atomic sets

$$\mathscr{A}_r = \left\{ \mathbf{ua}(\mathbf{r})^H : \mathbf{r} \in [0,1)^2, ||\mathbf{u}||_2 = 1 \right\} \subset \mathbb{C}^{J \times MP}, \tag{68}$$

and

$$\mathscr{A}_c = \left\{ \mathbf{va}(\mathbf{c})^H : \mathbf{c} \in [0,1)^2, ||\mathbf{v}||_2 = 1 \right\} \subset \mathbb{C}^{PJ \times MP}, \tag{69}$$

respectively. These atoms are basic units for synthesizing our signal with sparse channels. This definition leads to the following formulation of *atomic norms* $||\mathbf{Z}_r||_{\mathscr{A}_r}$ and $||\mathbf{Z}_c||_{\mathscr{A}_c}$ - a sparsity-enforcing analog of ℓ_1 norm for, respectively, general atomic sets $\mathscr{A}_r$ and $\mathscr{A}_c$:

$$||\mathbf{Z}_r||_{\mathscr{A}_r} = \inf_{\substack{[\alpha_r]_\ell \in \mathbb{C}, \boldsymbol{r}_\ell \in [0,1]^2 \\ ||\mathbf{u}||_2 = 1}} \left\{ \sum_\ell |[\alpha_r]_\ell| \Big| \mathbf{Z}_r = \sum_\ell [\alpha_r]_\ell \mathbf{ua}(\mathbf{r}_\ell)^H \right\}, \tag{70}$$

and

$$||\mathbf{Z}_c||_{\mathscr{A}_c} = \inf_{\substack{[\boldsymbol{\alpha}_c]_q \in \mathbb{C}, \boldsymbol{c}_q \in [0,1]^2 \\ ||\mathbf{v}||_2 = 1}} \left\{ \sum_q |[\boldsymbol{\alpha}_c]_q| \Big| \mathbf{Z}_c = \sum_q [\boldsymbol{\alpha}_c]_q \mathbf{va}(\mathbf{c}_q)^H \right\}. \tag{71}$$

To estimate the unknown matrices $\mathbf{Z}_r$ and $\mathbf{Z}_c$, we minimize the SoMAN $||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\mathbf{Z}_c||_{\mathscr{A}_c}$ among all possible matrices $\mathbf{Z}_r$ and $\mathbf{Z}_c$ leading to the same observed samples as $\mathbf{y}$. The *primal*

convex optimization problem for our DBD then becomes

$$\underset{\mathbf{Z}_r,\mathbf{Z}_c}{\text{minimize}} ||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\mathbf{Z}_c||_{\mathscr{A}_c} \text{ subject to } \mathbf{y} = \aleph_r(\mathbf{Z}_r) + \aleph_c(\mathbf{Z}_c). \tag{72}$$

A semidefinite characterization of minimization of SoMAN in (72) does not directly result from the primal problem as is the case with the standard atomic norm Mishra et al. (2015); Xu et al. (2014). Since the primal problem has only equality constraints, Slater's condition is satisfied, and strong duality holds. This implies, solving the dual problem also yields an exact solution to the primal problem. Therefore, we now analyze the *dual* problem of (72) and derive a new semidefinite program for SoMAN minimization using theories of positive-hyperoctant trigonometric polynomials (PhTP).

Dual problem. The dual problem of (72) is obtained via standard Lagrangian analysis. The Lagrangian function of problem (72) is

$$\mathscr{L}(\mathbf{Z}_r, \mathbf{Z}_c, \mathbf{q}) = ||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\mathbf{Z}_c||_{\mathscr{A}_c} + \langle \mathbf{q}, \mathbf{y} - \aleph_r(\mathbf{Z}_r) - \aleph_c(\mathbf{Z}_c) \rangle, \tag{73}$$

where $\mathbf{q}$ is the dual variable. Define the dual function $g(\mathbf{q})$ as

$$\begin{aligned} g(\mathbf{q}) &= \inf_{\mathbf{Z}_r,\mathbf{Z}_c} \mathscr{L}(\mathbf{Z}_r, \mathbf{Z}_c, \mathbf{q}) \\ &= \langle \mathbf{q}, \mathbf{y} \rangle + \inf_{\mathbf{Z}_r,\mathbf{Z}_c} \left(||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\mathbf{Z}_c||_{\mathscr{A}_c} \right) - \langle \mathbf{q}, \aleph_r(\mathbf{Z}_r) \rangle - \langle \mathbf{q}, \aleph_c(\mathbf{Z}_c) \rangle \big) \\ &= \langle \mathbf{q}, \mathbf{y} \rangle - \sup_{\mathbf{Z}_r} \left(\langle \aleph_r^*(\mathbf{q}), \mathbf{Z}_r \rangle - ||\mathbf{Z}_r||_{\mathscr{A}_r} \right) - \sup_{\mathbf{Z}_c} \left(\langle \aleph_c^*(\mathbf{q}), \mathbf{Z}_c \rangle - ||\mathbf{Z}_c||_{\mathscr{A}_c} \right), \end{aligned} \tag{74}$$

where $\aleph_r^* : \mathbb{C}^{MP} \to \mathbb{C}^{J\times MP}$ and $\aleph_c^* : \mathbb{C}^{MP} \to \mathbb{C}^{PJ\times MP}$ are the adjoint operators of $\aleph_r$ and $\aleph_c$, respectively, such that $\aleph_r^*(\boldsymbol{q}) = \sum_{p=0}^{P-1}\sum_{n=-N}^{N}[\mathbf{q}]_{\widetilde{n}}\mathbf{G}_{\widetilde{n}}^H$ and $\aleph_c^*(\mathbf{q}) = \sum_{p=0}^{P-1}\sum_{n=-N}^{N}[\mathbf{q}]_{\widetilde{n}}\mathbf{A}_{\widetilde{n}}^H$. The supremum values in (74) correspond to the convex conjugate function of the atomic norms $||\mathbf{Z}_r||_{\mathscr{A}_r}$ and $||\mathbf{Z}_c||_{\mathscr{A}_c}$.

Using the expansion of the dual function $g(\mathbf{q})$ in (74), the dual problem of (72) is

$$\begin{aligned} &\underset{\mathbf{q}}{\text{maximize}} \langle \mathbf{q}, \mathbf{y}\rangle_{\mathbb{R}} \\ &\text{subject to } \| \aleph_r^\star(\mathbf{q})\|_{\mathscr{A}_r}^\star \leq 1 \\ &\qquad\qquad \| \aleph_c^\star(\mathbf{q})\|_{\mathscr{A}_c}^\star \leq 1, \end{aligned} \tag{75}$$

where $\|\cdot\|^*$ represents the dual norm. This dual norm is defined as

$$||\mathbf{Z}||_{\mathscr{A}}^* = \sup_{||\mathbf{U}||_{\mathscr{A}}\leq 1} \langle \mathbf{U}, \mathbf{Z}\rangle, \tag{76}$$

where the indicator function

$$h^*(\mathbf{Z}) = \begin{cases} 0 & \text{if } ||\mathbf{Z}||_{\mathscr{A}}^* \leq 1, \\ \infty & \text{otherwise,} \end{cases}$$

is the conjugate function of the dual norm.

We identify that the constraints in (75) are positive-hyperoctant trigonometric polynomials (PhTP) in $\mathbf{r} \in [0,1]^2$ and $\mathbf{c} \in [0,1]^2$. Using the definition of dual-norm (76), rewrite these constraints

in (75) as

$$\|\,\aleph_r^*(\mathbf{q})\|_{\mathscr{A}_r}^* = \sup_{\mathbf{r}\in[0,1]^2,\|\mathbf{u}\|=1} \left|\langle \mathbf{u},\, \aleph_r^*(\mathbf{q})^H \mathbf{a}(\mathbf{r})\rangle\right| = \left\|\,\aleph_r^*(\mathbf{q})^H \mathbf{a}(\mathbf{r})\right\|_2,$$

$$\|\,\aleph_c^*(\mathbf{q})\|_{\mathscr{A}_c}^* = \sup_{\mathbf{c}\in[0,1]^2,\|\mathbf{v}\|=1} \left|\langle \mathbf{v},\, \aleph_c^*(\mathbf{q})^H \mathbf{a}(\mathbf{c})\rangle\right| = \left\|\,\aleph_c^*(\mathbf{q})^H \mathbf{a}(\mathbf{c})\right\|_2.$$

In the expressions above, define the bi-variate (delay and Doppler) vector polynomials in $\mathbf{r}$ and $\mathbf{c}$ as

$$\mathbf{f}_r(\mathbf{r}) = \aleph_r^*(\mathbf{q})^H \mathbf{a}(\mathbf{r}) = \sum_{p=0}^{P-1} \sum_{n=-N}^{N} [\mathbf{q}]_{\widetilde{n}} \mathbf{G}_{\widetilde{n}}^H \mathbf{a}(\mathbf{r}) \in \mathbb{C}^J, \tag{77}$$

$$\mathbf{f}_c(\mathbf{c}) = \aleph_c^*(\mathbf{q})^H \mathbf{a}(\mathbf{c}) = \sum_{p=0}^{P-1} \sum_{n=-N}^{N} [\mathbf{q}]_{\widetilde{n}} \mathbf{A}_{\widetilde{n}}^H \mathbf{a}(\mathbf{c}) \in \mathbb{C}^{JP}. \tag{78}$$

Evidently, the inequality constraints in (75) are equivalent to the Euclidean norms of the polynomials $\mathbf{f}_r(\mathbf{r})$ and $\mathbf{f}_c(\mathbf{c})$ being upper bounded by unity. It follows from the Bounded Real Lemma (BRL) Dumitrescu (2007) that the bound on the magnitude of a multivariate trigonometric polynomial is characterized by a linear matrix inequality (LMI). More precisely, given a bi-variate PhTP $R(\boldsymbol{\lambda}) = \sum_{\mathbf{t}} k_{\mathbf{t}} e^{-\mathrm{j}2\pi \boldsymbol{\lambda}^T \mathbf{t}}$, where $\boldsymbol{\lambda} \in [0,1]^2, \mathbf{t} = [t_1, t_2], 0 \le t_1 \le L_1, 0 \le t_2 \le L_2$, with integers L_1 and L_2. The inequality $|R(\boldsymbol{\lambda})| < 1$ implies that there exists a positive semidefinite *Gram* matrix $\mathbf{K}$ such

that (Dumitrescu, 2007, p. 135)

$$\delta_{\mathbf{n}} = \mathrm{Tr}\left(\mathbf{\Theta}_{\mathbf{n}}\mathbf{K}\right), \begin{bmatrix} \mathbf{K} & \mathbf{k}^H \\ \mathbf{k} & \mathbf{I}_J \end{bmatrix} \succeq 0, \tag{79}$$

where $\mathbf{n} = [n_1, n_2]$, $0 < n_1 < m_1$, $-m_2 < n_2 < m_2$, $\mathbf{k}$ is the vector containing all the polynomial coefficients,

$$\delta_{\mathbf{n}} = \begin{cases} 1, \; n_1 = n_2 = 0, \\ 0, \; \text{otherwise}, \end{cases} \tag{80}$$

and $\mathbf{\Theta}_{\mathbf{n}} = \mathbf{\Theta}_{n_2} \otimes \mathbf{\Theta}_{n_1}$, where $\mathbf{\Theta}_{n_1}$, $\mathbf{\Theta}_{n_2}$ are two elementary Toeplitz matrices with ones in the n_1-th and n_2-th diagonal respectively and zero elsewhere.

In the case of uni-variate trigonometric polynomials where the size of the *Gram* matrix is fixed. However, for the multi-variate case, the size of $\mathbf{K}$ depends on the sum-of-squares relaxation degree vector $\mathbf{m} = [m_1, m_2]$, where $m_1 \geq L_1$ and $m_2 \geq L_2$. A higher relaxation degree leads to a better approximation but it also leads to a higher computational complexity. In practice, this degree may be chosen to be the minimum possible value, i.e., $m_1 = L_1$ and $m_2 = L_2$ and still yield the optimal result Xu et al. (2014); Heckel et al. (2016). Furthermore, it is obvious that when such a positive semidefinite matrix $\mathbf{K}$ exists, $|R(\boldsymbol{\lambda})| < 1$. Following the SDP characterization of PhTP shown heretofore, we convert the constraints of the dual optimization problem (75) to the

following LMIs

$$\begin{bmatrix} \mathbf{K} & \widehat{\mathbf{K}}_r^H \\ \widehat{\mathbf{K}}_r & \mathbf{I}_J \end{bmatrix} \succeq 0, \begin{bmatrix} \mathbf{K} & \widehat{\mathbf{K}}_c^H \\ \widehat{\mathbf{K}}_c & \mathbf{I}_{JP} \end{bmatrix} \succeq 0, \mathbf{K} \succeq 0,$$

where $\widehat{\mathbf{K}}_r = \sum_{p=0}^{P-1} \sum_{n=-N}^{N} [\mathbf{q}]_{\widetilde{n}} \mathbf{G}_{\widetilde{n}}^H \in \mathbb{C}^{MP \times J}$ and $\widehat{\mathbf{K}}_c = \sum_{p=0}^{P-1} \sum_{n=-N}^{N} [\mathbf{q}]_{\widetilde{n}} \mathbf{A}_{\widetilde{n}}^H \in \mathbb{C}^{MP \times PJ}$ are the coefficients of the PhTPs. This leads to the SDP of (75) as

$$\begin{aligned} \underset{\mathbf{q},\mathbf{K}}{\text{maximize}} \quad & \langle \mathbf{q}, \mathbf{y} \rangle_{\mathbb{R}} \\ \text{subject to } & \mathbf{K} \succeq 0 \\ & \begin{bmatrix} \mathbf{K} & \widehat{\mathbf{K}}_r^H \\ \widehat{\mathbf{K}}_r & \mathbf{I}_J \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} \mathbf{K} & \widehat{\mathbf{K}}_c^H \\ \widehat{\mathbf{K}}_c & \mathbf{I}_{JP} \end{bmatrix} \succeq 0 \\ & \text{Tr}(\mathbf{\Theta}_\mathbf{n} \mathbf{K}) = \delta_\mathbf{n}. \end{aligned} \tag{81}$$

This SDP is solved using an off-the-shelf solver such as CVX Grant et al. (2009).

Remark 2. We highlight some key differences in our SDP with respect to some related prior works that apply ANM to other (non-BD) applications. For instance, the SDP in Xu et al. (2014), the coefficient of the polynomial is just the dual variable. In the non-blind SR problem of Heckel

et al. (2016), the coefficient of the polynomial in the LMI is defined as a product of the inverse Fourier matrix basis, the Gabor sensing matrix, and the dual variable. For blind 2-D SR of Suliman and Dai (2021), the coefficients are defined in terms of a low-dimensional subspace representation. Furthermore, these ANM approaches have only one LMI while SoMAN has as many LMIs as the number of atomic norms. A multitude of LMIs occur in the SDP for ANM applied to super-resolution problem in Mishra et al. (2015) but it employed only one PhTP that was also univariate.

After the dual variable is obtained by solving the SDP in (81), we build the PhTPs $\mathbf{f}_r(\mathbf{r})$, $\mathbf{f}_c(\mathbf{c})$. The set of delay and Doppler parameters for radar $\widehat{\mathscr{R}} = \{\widehat{\mathbf{r}}_\ell\}_{\ell=0}^{L-1}$ and communications $\widehat{\mathscr{C}} = \{\widehat{\mathbf{r}}_\ell\}_{\ell=0}^{Q-1}$ are localized in the respective delay-Doppler planes when, respectively, the PhTPs $\|\mathbf{f}_r(\widehat{\mathbf{r}}_\ell)\|$ and $\|\mathbf{f}_c(\widehat{\mathbf{r}}_q)\|1$ assume a maximum modulus of unity. This procedure yields channel information for both systems. Thereafter, to extract the radar and communications transmit waveform, a system of linear equations needs to be solved.

6.3.1. Dual Certificate. The primal problem (72) has only equality constraints. Hence, Slater's condition is satisfied, and strong duality holds. This allows us to state the following Proposition 1, which presents the dual-certificate of support for the optimizer of (72).

Proposition 1. Consider the signal $\mathbf{y}$ in (67). The atomic sets $\mathscr{A}_r$ and $\mathscr{A}_c$ are as defined in (68) and (69), respectively. Denote $\mathscr{R} = \{\mathbf{r}_\ell\}_{\ell=0}^{L-1}$ and $\mathscr{C} = \{\mathbf{c}_q\}_{q=0}^{Q-1}$. The pair $\{\mathbf{Z}_r^\star, \mathbf{Z}_c^\star\}$ is the unique solution of (72) if there exist two bi-variate PhTPs defined as in (77)-(78), one each in $\{\boldsymbol{\tau}_r,\ \boldsymbol{\nu}_r\}$

and $\{\boldsymbol{\tau}_c, \boldsymbol{\nu}_c\}$, with complex coefficients $\mathbf{q}$ such that

$$\mathbf{f}_r(\mathbf{r}_\ell) = \text{sign}([\boldsymbol{\alpha}_r]_\ell)\mathbf{u},\ \forall \mathbf{r}_\ell \in \mathscr{R}, \tag{82a}$$

$$\mathbf{f}_c(\mathbf{c}_q) = \text{sign}([\boldsymbol{\alpha}_c]_q)\mathbf{v},\ \forall \mathbf{c}_q \in \mathscr{C}, \tag{82b}$$

$$\|\mathbf{f}_r(\mathbf{r})\|_2^2 < 1,\ \forall \mathbf{r} \in [0,1]^2 \setminus \mathscr{R}, \tag{82c}$$

$$\|\mathbf{f}_c(\mathbf{c})\|_2^2 < 1,\ \forall \mathbf{c} \in [0,1]^2 \setminus \mathscr{C}. \tag{82d}$$

Proof. The variable $\mathbf{q}$ is dual feasible. This gives

$$\begin{aligned}
&\langle \mathbf{q}, \mathbf{y} \rangle_{\mathbb{R}} = \langle \mathbf{q}, \aleph_{\mathbf{r}}(\mathbf{Z}_{\mathbf{r}}) \rangle_{\mathbb{R}} + \langle \aleph_c(\mathbf{Z}_c) \rangle_{\mathbb{R}} \\
&= \langle \aleph_r^*(\mathbf{q}), \mathbf{Z}_r \rangle_{\mathbb{R}} + \langle \aleph_c^*(\mathbf{q}), \mathbf{Z}_c \rangle_{\mathbb{R}} \\
&= \langle \aleph_r^*(\mathbf{q}), \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell \mathbf{u}\mathbf{a}(\mathbf{r}_\ell)^H \rangle_{\mathbb{R}} + \langle \aleph_c^*(\mathbf{q}), \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q \mathbf{v}\mathbf{a}(\mathbf{c}_q)^H \rangle_{\mathbb{R}} \\
&= \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell^* \langle \aleph_r^*(\mathbf{q}), \mathbf{u}\mathbf{a}(\mathbf{r}_\ell)^H \rangle_{\mathbb{R}} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q^* \langle \aleph_c^*(\mathbf{q}), \mathbf{v}\mathbf{a}(\mathbf{c}_q)^H \rangle_{\mathbb{R}} \\
&= \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell^* \langle \mathbf{f}_r(\mathbf{r}_\ell), \mathbf{u} \rangle_{\mathbb{R}} + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q^* \langle \mathbf{f}_c(\mathbf{c}_q), \mathbf{v} \rangle_{\mathbb{R}} \\
&= \sum_{\ell=0}^{L-1} [\boldsymbol{\alpha}_r]_\ell^* \text{sign}([\boldsymbol{\alpha}_r]_\ell) + \sum_{q=0}^{Q-1} [\boldsymbol{\alpha}_c]_q^* \text{sign}([\boldsymbol{\alpha}_c]_q) \\
&= \sum_{\ell=0}^{L-1} |[\boldsymbol{\alpha}_r]_\ell| + \sum_{q=0}^{Q-1} |[\boldsymbol{\alpha}_c]_q| \\
&\geq ||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\mathbf{Z}_c||_{\mathscr{A}_c}.
\end{aligned} \tag{83}$$

On the other hand, based on the Hölder inequality we have

$$\begin{aligned}\langle \mathbf{q}, \mathbf{y} \rangle_{\mathbb{R}} &= \langle \aleph_r^*(\mathbf{q}), \mathbf{Z}_r \rangle_{\mathbb{R}} + \langle \aleph_c^*(\mathbf{q}), \mathbf{Z}_c \rangle_{\mathbb{R}} \\ &\leq ||\aleph_r^*(\mathbf{q})||_{\mathscr{A}_r}^* ||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\aleph_c^*(\mathbf{q})||_{\mathscr{A}_c}^* ||\mathbf{Z}_c||_{\mathscr{A}_c} \\ &\leq ||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\mathbf{Z}_c||_{\mathscr{A}_c}, \end{aligned} \tag{84}$$

where the first inequality follows from the Cauchy-Schwartz inequality and the last follows from (82a)-(82d). From (83) and (84), $\langle \mathbf{q}, \mathbf{y} \rangle_{\mathbb{R}} = ||\mathbf{Z}_r||_{\mathscr{A}_r} + ||\mathbf{Z}_c||_{\mathscr{A}_c}$ so that the pair $\{\mathbf{Z}_r, \mathbf{Z}_c\}$ is primal optimal and $\mathbf{q}$ is dual optimal based on strong duality. ■

6.3.2. Performance guarantees. Recall the following useful properties.

Definition 1 (Isotropy Candès and Plan (2011))**.** The distribution $\mathscr{F}$ is isotropic if given a random vector $\mathbf{x} \in \mathbb{C}^N \sim \mathscr{F}$ satisfies the following condition:

$$\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_N. \tag{85}$$

which means that the columns have unit variance and are uncorrelated.

Definition 2 (Incoherence Candès and Plan (2011))**.** The distribution $\mathscr{F}$ is incoherent if given a vector $\mathbf{x} = [x_0, \ldots, x_{N-1}] \sim \mathscr{F}$ and given coherence parameter μ it satisfies that:

$$\max_{0<i<N-1} |x_i|^2 \leq \mu. \tag{86}$$

Following these definitions, we now state our main result.

Theorem 2. Assume the columns of **B** and **D** in, respectively, (56a) and (57) are drawn from a distribution that satisfies the isotropy and incoherence properties. Further, if $\|\mathbf{v}\|_2^2 = \|\mathbf{u}\|_2^2 = 1$ and $|\alpha_{\mathbf{i}}| = 1$, then there exists a numerical constant C such that

$$\begin{aligned} MP \geq C\mu \max(L,Q) J \log^2\left(\frac{MPJ}{\delta}\right) \\ \times \max\left\{\log\left(\frac{MPQJ}{\delta}\right), \log\left(\frac{MPLJ}{\delta}\right)\right\}, \end{aligned} \tag{87}$$

is sufficient to guarantee that, with probability at least $1-\delta$, the pair $\{\mathbf{Z}_r, \mathbf{Z}_c\}$ is perfectly recovered from MP measurements of the observation vector $\mathbf{y}$ in (67).

Proof. See Appendix A. ■

7. Conclusions and Future Work

This doctoral dissertation introduced a new coding strategy dubbed Time-Multiplexed Coded Aperture (TMCA) and demonstrated that it improves optical codifications in three different applications: light-field imaging, compressive hyperspectral imaging, and spectral-depth imaging. TMCA improves traditional optical coding strategies using amplitude or only-phase CA without the need for supplementary optical elements. The main principle of the proposed TMCA codification strategy consists of allowing the CA to vary in time and synchronize with a coded shutter. In the studied applications, this dissertation demonstrates that the TMCA yields three novel codifications: a colored coded aperture, a CA with angle-sensitivity response, and a spatially variant response for a given depth in a scene. This codification cannot be obtained without the synergistic interaction of the CA with the coded shutter in time. The coded shutter can be implemented electronically, employing dedicated sensors or a sequence of multiple snapshots subsequently averaged together. Furthermore, leveraging the advances in deep learning, the TMCA codes are optimized jointly with the recovery algorithm using an end-to-end optimization approach. Through comprehensive synthetic and real experiments, this dissertation validates that the optimized data-driven TMCA achieves largely superior reconstruction quality compared with traditional CA approaches. The main limitations of the proposed approach are the assumption of a static scene and the low light efficiency because of the sensor codification which can be detrimental in low-light scenarios or high-speed imaging. However, additional constraints on light throughput of the coded shutter and CAs can be included in the optimization, to further engineer the proposed systems to be real-time

and thus enable high-quality videos on those applications. Furthermore, this thesis assumes perfect state switching of coded aperture and shutters for the discrete-time slots and they are perfectly synchronized. While real evaluation in controlled scenarios validates our proposed approach, it is expected a little degradation of performance when considering non-zero switching times. Overcoming these challenges represents a promising direction for further investigation. The proposed TMCA would naturally extend to a broader range of applications that leverage coded apertures and improve imaging quality. This is also one of our future objectives.

Finally, exploring extended ideas that came out in the fulfillment of the objectives, this thesis investigated the joint radar-communications processing from a dual-blind deconvolution (DBD) perspective. The proposed approach leverages the sparse structure of the channels and recovers the parameters of interest in this highly ill-posed problem by minimizing the sum of multivariate atomic norms objective. This research employed positive-hyperoctant trigonometric polynomials theories to formulate a semidefinite programming problem and provide theoretical analysis in the minimum number of samples required for near-perfect recovery. Numerical experiments showed numerical perfect recovery of radar and communications parameters in several different scenarios. Future work in this area includes considering a multiple-antenna receiver or the investigation of a continuous-wave transmission model for radar and a common waveform for co-designing radar-communications systems.

Bibliography

Abed-Meraim, K., Qiu, W., and Hua, Y. (1997). Blind system identification. *Proceedings of the IEEE*, 85(8):1310–1322.

Ahmed, A. and Demanet, L. (2018). Leveraging diversity and sparsity in blind deconvolution. *IEEE Transactions on Information Theory*, 64(6):3975–4000.

Ahmed, A., Recht, B., and Romberg, J. (2013). Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732.

Altmann, Y., Ren, X., McCarthy, A., Buller, G. S., and McLaughlin, S. (2016). Lidar waveform-based analysis of depth images constructed using sparse single-photon data. *IEEE Transactions on Image Processing*, 25(5):1935–1946.

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

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

Arguello, H., Pinilla, S., Peng, Y., Ikoma, H., Bacca, J., and Wetzstein, G. (2021). Shift-variant color-coded diffractive spectral imaging system. *Optica*, 8(11):1424–1434.

Ayers, G. and Dainty, J. C. (1988). Iterative blind deconvolution method and its applications. *Optics Letters*, 13(7):547–549.

Backman, V., Wallace, M. B., Perelman, L., Arendt, J., Gurjar, R., Müller, M., Zhang, Q., Zonios, G., Kline, E., McGillican, T., et al. (2000). Detection of preinvasive cancer cells. *Nature*, 406(6791):35.

Bacon, R., Adam, G., Baranne, A., Courtes, G., Dubet, D., Dubois, J., Emsellem, E., Ferruit, P., Georgelin, Y., Monnet, G., et al. (1995). 3d spectrography at high spatial resolution. i. concept and realization of the integral field spectrograph tiger. *Astronomy and Astrophysics Supplement Series*, 113:347.

Baek, S.-H., Ikoma, H., Jeon, D. S., Li, Y., Heidrich, W., Wetzstein, G., and Kim, M. H. (2021). Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660.

Baek, S.-H., Kim, I., Gutierrez, D., and Kim, M. H. (2017). Compact single-shot hyperspectral imaging using a prism. *ACM Transactions on Graphics (TOG)*, 36(6):1–12.

Bică, M. and Koivunen, V. (2018). Radar waveform optimization for target parameter estimation in cooperative radar-communications systems. *IEEE Transactions on Aerospace and Electronic Systems*, 55(5):2314–2326.

Bioucas-Dias, J. M. and Valadao, G. (2007). Phase unwrapping via graph cuts. *IEEE Transactions on Image processing*, 16(3):698–709.

Biswas, J. and Veloso, M. (2012). Depth camera based indoor mobile robot localization and

navigation. In *2012 IEEE International Conference on Robotics and Automation*, pages 1697–1702. IEEE.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122.

Candès, E. J. and Plan, Y. (2011). A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254.

Candes, E. J., Strohmer, T., and Voroninski, V. (2013). Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274.

Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30.

Chang, J. and Wetzstein, G. (2019). Deep optics for monocular depth estimation and 3d object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10193–10202.

Chi, Y. (2016). Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):782–794.

Chiriyath, A. R., Paul, B., and Bliss, D. W. (2017). Radar-communications convergence: Co-

existence, cooperation, and co-design. *IEEE Transactions on Cognitive Communications and Networking*, 3(1):1–12.

Choi, I., Jeon, D. S., Nam, G., Gutierrez, D., and Kim, M. H. (2017). High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics (TOG)*, 36(6):1–13.

Cimini, L. (1985). Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Transactions on Communications*, 33(7):665–675.

Correa, C. V., Arguello, H., and Arce, G. R. (2015). Snapshot colored compressive spectral imager. *JOSA A*, 32(10):1754–1763.

Correa, C. V., Hinojosa, C. A., Arce, G. R., and Arguello, H. (2016). Multiple snapshot colored compressive spectral imager. *Optical Engineering*, 56(4):041309.

Delalieux, S., Auwerkerken, A., Verstraeten, W. W., Somers, B., Valcke, R., Lhermitte, S., Keulemans, J., and Coppin, P. (2009). Hyperspectral reflectance and fluorescence imaging to detect scab induced stress in apple leaves. *Remote Sensing*, 1(4):858–874.

Dokhanchi, S. H., Mysore, B. S., Mishra, K. V., and Ottersten, B. (2019). A mmWave automotive joint radar-communications system. *IEEE Transactions on Aerospace and Electronic Systems*, 55(3):1241–1260.

Dong, J., Yang, K., and Shi, Y. (2018). Blind demixing for low-latency communication. *IEEE Transactions on Wireless Communications*, 18(2):897–911.

Duggal, G., Vishwakarma, S., Mishra, K. V., and Ram, S. S. (2020). Doppler-resilient 802.11ad-based ultrashort range automotive joint radar-communications system. *IEEE Transactions on Aerospace and Electronic Systems*, 56(5):4035–4048.

Dumitrescu, B. (2007). *Positive trigonometric polynomials and signal processing applications*. Springer.

Eigen, D., Puhrsch, C., and Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, pages 2366–2374.

Elbir, A. M., Mishra, K. V., and Chatzinotas, S. (2021). Terahertz-band joint ultra-massive MIMO radar-communications: Model-based and model-free hybrid beamforming. *IEEE Journal of Special Topics in Signal Processing*, 15(6):1468–1483.

Farshchian, M. and Selesnick, I. (2016). A dual-deconvolution algorithm for radar and communication black-space spectrum sharing. In *IEEE International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing*, pages 6–10.

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

Fu, H., Gong, M., Wang, C., Batmanghelich, K., and Tao, D. (2018). Deep ordinal regression network for monocular depth estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2002–2011.

Gao, L. and Wang, L. V. (2016). A review of snapshot multidimensional optical imaging: measuring photon tags in parallel. *Physics reports*, 616:1–37.

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

Godard, C., Mac Aodha, O., and Brostow, G. J. (2017). Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 270–279.

Goodman, J. W. (2005). *Introduction to Fourier optics*. Roberts and Company Publishers.

Gowen, A., O'Donnell, C., Cullen, P., Downey, G., and Frias, J. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, 18(12):590–598.

Grant, M., Boyd, S., and Ye, Y. (2009). CVX: MATLAB software for disciplined convex programming.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M., et al. (1998). Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 65(3):227–248.

Guo, M., Hou, J., Jin, J., Chen, J., and Chau, L.-P. (2020). Deep spatial-angular regularization

for compressive light field reconstruction over coded apertures. In *European Conference on Computer Vision*, pages 278–294. Springer.

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

Haim, H., Elmalem, S., Giryes, R., Bronstein, A. M., and Marom, E. (2018). Depth estimation from a single image using deep learned phase coded mask. *IEEE Transactions on Computational Imaging*, 4(3):298–310.

Hansard, M., Lee, S., Choi, O., and Horaud, R. P. (2012). *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media.

Heckel, R., Morgenshtern, V. I., and Soltanolkotabi, M. (2016). Super-resolution radar. *Information and Inference: A Journal of the IMA*, 5(1):22–75.

Heist, S., Zhang, C., Reichwald, K., Kühmstedt, P., Notni, G., and Tünnermann, A. (2018). 5d hyperspectral imaging: fast and accurate measurement of surface shape and spectral characteristics using structured light. *Optics express*, 26(18):23366–23379.

Henderson, P. and Ferrari, V. (2019). Learning single-image 3d reconstruction by generative modelling of shape, pose and shading. *International Journal of Computer Vision*, pages 1–20.

Herrala, E., Okkonen, J., Hyvarinen, T., Aikio, M., and Lammasniemi, J. (1994). Imaging spectrometer for process industry applications. In *Proc. SPIE*, volume 2248, pages 33–40.

Hoiem, D., Efros, A. A., and Hebert, M. (2007). Recovering surface layout from an image. *International Journal of Computer Vision*, 75(1):151–172.

Honauer, K., Johannsen, O., Kondermann, D., and Goldluecke, B. (2016). A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision*, pages 19–34. Springer.

Horisaki, R., Xiao, X., Tanida, J., and Javidi, B. (2013). Feasibility study for compressive multi-dimensional integral imaging. *Optics express*, 21(4):4263–4279.

Hsu, S. M. and Burke, H.-H. K. (2003). Multisensor fusion with hyperspectral imaging data: detection and classification. *Lincoln Laboratory Journal*, 14(1).

Ikoma, H., Nguyen, C. M., Metzler, C. A., Peng, Y., and Wetzstein, G. (2021). Depth from defocus with learned optics for imaging and occlusion-aware depth estimation. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE.

Jacome, R., Mishra, K. V., Vargas, E., Sadler, B. M., and Arguello, H. (2022). Multi-dimensional dual-blind deconvolution approach toward joint radar-communications. In *IEEE International Workshop on Signal Processing Advances in Wireless Communication*, pages 1–5.

Jalal, A., Kamal, S., and Azurdia-Meza, C. A. (2019). Depth maps-based human segmentation and action recognition using full-body plus body color cues via recognizer engine. *Journal of Electrical Engineering & Technology*, 14(1):455–461.

Jefferies, S. M. and Christou, J. C. (1993). Restoration of astronomical images by iterative blind deconvolution. *The Astrophysical Journal*, 415:862.

Jeon, D. S., Baek, S.-H., Yi, S., Fu, Q., Dun, X., Heidrich, W., and Kim, M. H. (2019). Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Transactions on Graphics (TOG)*, 38(4):1–13.

Jeon, H.-G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.-W., and Kweon, I. S. (2018). Depth from a light field image with learning-based matching costs. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):297–310.

Jeon, H.-G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.-W., and So Kweon, I. (2015). Accurate depth map estimation from a lenslet light field camera. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1547–1555.

Kalantari, N. K., Wang, T.-C., and Ramamoorthi, R. (2016). Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–10.

Kester, R. T., Bedard, N., Gao, L., and Tkaczyk, T. S. (2011). Real-time snapshot hyperspectral imaging endoscope. *Journal of biomedical optics*, 16(5):056005–056005.

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

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kruse, F. A., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., and Goetz, A. (1993). The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163.

Kuo, H.-W., Lau, Y., Zhang, Y., and Wright, J. (2019). Geometry and symmetry in short-and-sparse deconvolution. In *International Conference on Machine Learning*, pages 3570–3580.

Lee, K., Li, Y., Junge, M., and Bresler, Y. (2016). Blind recovery of sparse signals from subsampled convolution. *IEEE Transactions on Information Theory*, 63(2):802–821.

Levin, A., Fergus, R., Durand, F., and Freeman, W. T. (2007). Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26(3):70–es.

Levoy, M. and Hanrahan, P. (1996). Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42.

Li, S., Wakin, M. B., and Tang, G. (2019a). Atomic norm denoising for complex exponentials with unknown waveform modulations. *IEEE Transactions on Information Theory*, 66(6):3893–3913.

Li, X., Ling, S., Strohmer, T., and Wei, K. (2019b). Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Applied and Computational Harmonic Analysis*, 47(3):893–934.

Li, Y. and Chi, Y. (2019). Stable separation and super-resolution of mixture models. *Applied and Computational Harmonic Analysis*, 46(1):1–39.

Li, Y., Lee, K., and Bresler, Y. (2017). Identifiability and stability in blind deconvolution under minimal assumptions. *IEEE Transactions on Information Theory*, 63(7):4619–4633.

Lin, X., Suo, J., Wetzstein, G., Dai, Q., and Raskar, R. (2013). Coded focal stack photography. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE.

Ling, S. and Strohmer, T. (2017). Blind deconvolution meets blind demixing: Algorithms and performance bounds. *IEEE Transactions on Information Theory*, 63(7):4497–4520.

Liu, H., Bruning, B., Garnett, T., and Berger, B. (2020). Hyperspectral imaging and 3d technologies for plant phenotyping: From satellite to close-range sensing. *Computers and Electronics in Agriculture*, 175:105621.

Liu, J., Mishra, K. V., and Saquib, M. (2022). Co-designing statistical MIMO radar and in-band full-duplex multi-user MIMO communications. *IEEE Transactions on Aerospace and Electronic Systems*. in press.

Luo, X. and Giannakis, G. B. (2006). Low-complexity blind synchronization and demodulation for (ultra-) wideband multi-user ad hoc access. *IEEE Transactions on Wireless communications*, 5(7):1930–1941.

Martel, J. N., Mueller, L., Carey, S. J., Dudek, P., and Wetzstein, G. (2020). Neural sensors:

Learning pixel exposures for hdr imaging and video compressive sensing with programmable sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Martínez, E., Vargas, E., and Arguello, H. (2022). Fast disparity estimation from a single compressed light field measurement. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 1991–1995. IEEE.

Marwah, K., Wetzstein, G., Bando, Y., and Raskar, R. (2013). Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Transactions on Graphics (TOG)*, 32(4):1–12.

Metzler, C. A., Ikoma, H., Peng, Y., and Wetzstein, G. (2019). Deep optics for single-shot high-dynamic-range imaging. *arXiv preprint arXiv:1908.00620*.

Metzler, C. A., Ikoma, H., Peng, Y., and Wetzstein, G. (2020). Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1375–1385.

Middleton, E. M., Ungar, S. G., Mandl, D. J., Ong, L., Frye, S. W., Campbell, P. E., Landis, D. R., Young, J. P., and Pollack, N. H. (2013). The Earth Observing One (EO-1) satellite mission: Over a decade in space. *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, 6(2):243–256.

Mishra, K. V., Bhavani Shankar, M. R., Koivunen, V., Ottersten, B., and Vorobyov, S. A. (2019). Toward millimeter wave joint radar-communications: A signal processing perspective. *IEEE Signal Processing Magazine*, 36(5):100–114.

Mishra, K. V., Cho, M., Kruger, A., and Xu, W. (2015). Spectral super-resolution with prior knowledge. *IEEE Transactions on Signal Processing*, 63(20):5342–5357.

Mishra, K. V. and Eldar, Y. C. (2019). Sub-Nyquist radar: Principles and prototypes. In Maio, A. D., Eldar, Y. C., and Haimovich, A., editors, *Compressed Sensing in Radar Signal Processing*, pages 1–48. Cambridge University Press.

Morris, H. R., Hoyt, C. C., and Treado, P. J. (1994). Imaging spectrometers for fluorescence and raman microscopy: acousto-optic and liquid crystal tunable filters. *Applied spectroscopy*, 48(7):857–866.

Muns, G. R., Mishra, K. V., Guerra, C. B., Eldar, Y. C., and Chowdhury, K. R. (2019). Beam alignment and tracking for autonomous vehicular communication using IEEE 802.11 ad-based radar. In *IEEE Conference on Computer Communications Workshops - Hot Topics in Social and Mobile Connected Smart Objects*, pages 535–540.

Nehme, E., Freedman, D., Gordon, R., Ferdman, B., Weiss, L. E., Alalouf, O., Naor, T., Orange, R., Michaeli, T., and Shechtman, Y. (2020). Deepstorm3d: dense 3d localization microscopy and psf design by deep learning. *Nature Methods*, 17(7):734–740.

Okamoto, T. and Yamaguchi, I. (1991). Simultaneous acquisition of spectral image information. *Optics letters*, 16(16):1277–1279.

Panasyuk, S. V., Yang, S., Faller, D. V., Ngo, D., Lew, R. A., Freeman, J. E., and Rogers, A. E.

(2007). Medical hyperspectral imaging to facilitate residual tumor identification during surgery. *Cancer biology & therapy*, 6(3):439–446.

Paschalidou, D., Ulusoy, A. O., and Geiger, A. (2019). Superquadrics revisited: Learning 3d shape parsing beyond cuboids. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10344–10353.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch. In *NIPS-W*.

Paul, B., Chiriyath, A. R., and Bliss, D. W. (2016). Survey of RF communications and sensing convergence research. *IEEE Access*, 5:252–270.

Radosevich, A. J., Bouchard, M. B., Burgess, S. A., Stolper, R., Chen, B., and Hillman, E. M. (2008). Hyperspectral in-vivo two-photon microscopy of intrinsic fluorophores. In *Biomedical Optics*, page BWG7. Optica Publishing Group.

Rueda-Chacon, H., Florez, J. F., Lau, D. L., and Arce, G. R. (2019). Snapshot compressive tof+ spectral imaging via optimized color-coded apertures. *IEEE transactions on pattern analysis and machine intelligence*.

Schaepman, M. E., Ustin, S. L., Plaza, A. J., Painter, T. H., Verrelst, J., and Liang, S. (2009). Earth system science related imaging spectroscopy-An assessment. *Remote Sens. Environ.*, 113(123–137).

Schreiber, N. F., Genzel, R., Bouché, N., Cresci, G., Davies, R., Buschkamp, P., Shapiro, K., Tacconi, L., Hicks, E., Genel, S., et al. (2009). The SINS survey: SINFONI integral field spectroscopy of z 2 star-forming galaxies. *The Astrophysical Journal*, 706(2):1364.

Sedighi, S., Mishra, K. V., Shankar, M. B., and Ottersten, B. (2021). Localization with one-bit passive radars in narrowband internet-of-things using multivariate polynomial optimization. *IEEE Transactions on Signal Processing*, 69:2525–2540.

Shaw, G. A. and Burke, H.-H. K. (2003a). Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14(1):3–28.

Shaw, G. A. and Burke, H. K. (2003b). Spectral imaging for remote sensing. *Lincoln laboratory journal*, 14(1):3–28.

Shi, J., Jiang, X., and Guillemot, C. (2019). A framework for learning depth from a flexible subset of dense and sparse light field views. *IEEE Transactions on Image Processing*, 28(12):5867–5880.

Sitzmann, V., Diamond, S., Peng, Y., Dun, X., Boyd, S., Heidrich, W., Heide, F., and Wetzstein, G. (2018). End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, 37(4):1–13.

Skolnik, M. I. (2008). *Radar handbook*. McGraw-Hill, third edition.

Song, S., Lichtenberg, S. P., and Xiao, J. (2015). Sun rgb-d: A rgb-d scene understanding bench-

mark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 567–576.

Suliman, M. A. and Dai, W. (2019). Exact three-dimensional estimation in blind super-resolution via convex optimization. In *Annual Conference on Information Sciences and Systems*, pages 1–9.

Suliman, M. A. and Dai, W. (2021). Mathematical theory of atomic norm denoising in blind two-dimensional super-resolution. *IEEE Transactions on Signal Processing*, 69:1681–1696.

Suliman, M. A. and Dai, W. (2022). Blind two-dimensional super-resolution and its performance guarantee. *IEEE Transactions on Signal Processing*, 70:2844–2858.

Sun, M.-J., Edgar, M. P., Gibson, G. M., Sun, B., Radwell, N., Lamb, R., and Padgett, M. J. (2016). Single-pixel three-dimensional imaging with time-based depth resolution. *Nature communications*, 7(1):1–6.

Tang, G., Bhaskar, B. N., Shah, P., and Recht, B. (2013). Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490.

Vargas, E., Mishra, K. V., Jacome, R., Sadler, B. M., and Arguello, H. (2022). Joint radar-communications processing from a dual-blind deconvolution perspective. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5622–5626.

Wagadarikar, A., John, R., Willett, R., and Brady, D. (2008). Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51.

Wang, L., Xiong, Z., Shi, G., Zeng, W., and Wu, F. (2016). Simultaneous depth and spectral imaging with a cross-modal stereo system. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(3):812–817.

Wang, L., Zhang, T., Fu, Y., and Huang, H. (2018). Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Transactions on Image Processing*, 28(5):2257–2270.

Wang, T.-C., Efros, A. A., and Ramamoorthi, R. (2015). Occlusion-aware depth estimation using light-field cameras. In *Proceedings of the IEEE international conference on computer vision*, pages 3487–3495.

Wang, Z. and Bovik, A. C. (2002). A universal image quality index. *IEEE signal processing letters*, 9(3):81–84.

Wen, R., Chng, C.-B., and Chui, C.-K. (2017). Augmented reality guidance with multimodality imaging data and depth-perceived interaction for robot-assisted surgery. *Robotics*, 6(2):13.

Wetzstein, G., Ozcan, A., Gigan, S., Fan, S., Englund, D., Soljačić, M., Denz, C., Miller, D. A., and Psaltis, D. (2020). Inference in artificial intelligence with deep optics and photonics. *Nature*, 588(7836):39–47.

Wu, J., Xiong, B., Lin, X., He, J., Suo, J., and Dai, Q. (2016). Snapshot hyperspectral volumetric microscopy. *Scientific reports*, 6(1):1–10.

Wu, L., Mishra, K. V., Bhavani Shankar, M. R., and Ottersten, B. (2022). Resource allocation in heterogeneously-distributed joint radar-communications under asynchronous Bayesian tracking framework. *IEEE Journal on Selected Areas in Communications*, 40(7):2026–2042.

Wu, Y., Boominathan, V., Chen, H., Sankaranarayanan, A., and Veeraraghavan, A. (2019). Phase-cam3d?learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE.

Xu, W., Cai, J.-F., Mishra, K. V., Cho, M., and Kruger, A. (2014). Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies. In *IEEE Information Theory and Applications Workshop*, pages 1–4.

Yang, D., Tang, G., and Wakin, M. B. (2016a). Super-resolution of complex exponentials from modulations with unknown waveforms. *IEEE Transactions on Information Theory*, 62(10):5809–5830.

Yang, D., Tang, G., and Wakin, M. B. (2016b). Super-resolution of complex exponentials from modulations with unknown waveforms. *IEEE Transactions on Information Theory*, 62(10):5809–5830.

Yao, M., Xiong, Z., Wang, L., Liu, D., and Chen, X. (2019). Spectral-depth imaging with deep learning based reconstruction. *Optics Express*, 27(26):38312–38325.

Zhang, L., Curless, B., and Seitz, S. M. (2002). Rapid shape acquisition using color structured

light and multi-pass dynamic programming. In *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*, pages 24–36. IEEE.

Zhang, Y., Kuo, H.-W., and Wright, J. (2019). Structured local optima in sparse blind deconvolution. *IEEE Transactions on Information Theory*, 66(1):419–452.

Zheng, L. and Wang, X. (2017). Super-resolution delay-Doppler estimation for OFDM passive radar. *IEEE Transactions on Signal Processing*, 65(9):2197–2210.

Zhu, J., Wang, L., Yang, R., and Davis, J. (2008). Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

Appendices

Appendix A. Dual Blind Deconvolution for Joint Radar-Communications Systems

Vargas, E., Mishra, K. V., Jacome, R., Sadler, B. M., & Arguello, H. (2023). Dual-blind deconvolution for overlaid radar-communications systems. IEEE Journal on Selected Areas in Information Theory. https://doi.org/10.1109/JSAIT.2023.3287823