Coded aperture design for adaptive compressive spectral imaging

Nelson Eduardo Díaz Díaz

Doctoral thesis to qualify for the Doctor of Engineering degree, electronic area

Advisor

PhD. Henry Arguello Fuentes

Department of Systems Engineering, and Informatics

Universidad Industrial de Santander

School of Physical-Mechanical Engineering

Department of Electrical, Electronics, and Telecommunication Engineering

Bucaramanga

2020

Dedication

I dedicate my dissertation to,

my parents,

my brothers,

Thank you for your unconditional support.

Acknowledgements

A gratefully acknowledge to the High Dimensional Signal Processing (HDSP) research group for providing the scientific support and the necessary equipment for the development of this research work. Especially. thanks to the director of the group and advisor, Ph.D. Henry Arguello Fuentes for the teaching and training as a doctor in engineering and as a human being.

I also would like to thank Universidad Industrial de Santander for the training, teaching and its support with the project the STIC-AmSud project 18-STIC-05. HyperMed - Image reconstruction from compressed measurements: application to hyperspectral and medical imaging. Moreover, a recognize to Colciencias with the grant national doctor 727 de 2015. Last but not the less, thanks Télécommunications spatiales et aéronautiques (TéSA, acronymic in french) for allowing work in the lab in Toulouse, France.

Table of Contents

List of figures

List of tables

Resumen

Título: Diseño de códigos de apertura para un sistema de muestreo comprimido de imágenes espectrales [1]

Autor: NELSON EDUARDO DÍAZ DÍAZ [**]

Palabras Clave: Imágenes espectrales comprimidas, códigos de apertura, adquisición adaptativa de imágenes, adqusición de imágenes digitales.

Descripción: La adquisición de imágenes espectrales de manera tradicional es un proceso lento debido a que se requiere escanear la escena bajo estudio. En contraste, obtener imágenes espectrales comprimidas (*compressive spectral imaging*, CSI por sus siglas en inglés) ha atraído un gran interés porque reduce el número de muestras respecto a los sistemas de escaneo para capturar el cubo de datos, esto mejora la velocidad de muestreo. CSI mide proyecciones lineales de una escena, y luego un algoritmo de reconstrucción estima la escena subyacente. Una arquitectura ejemplo de CSI es el sistema de adquisición espectral comprimido de única captura con apertura de color (*Colored coded aperture compressive spectral imager*, C-CASSI, por sus siglas en inglés) que emplea un arreglo de filtros como códigos de apertura para codificar espacial y espectralmente la luz que entra en el sistema. Dichos códigos de apertura son equivalentes a la matriz de muestreo del sistema, su función es obtener medidas de la señal. Al día de hoy los trabajos relacionados con C-CASSI han usado códigos de apertura no adaptados. Específicamente, el muestreo no adaptado ignora la información a priori acerca de la señal para diseñar los patrones de codificación. Por lo tanto, este trabajo propone tres métodos diferentes para diseñar de forma adaptativa la apertura codificada de colores aplicada a diferentes problemas, incluyendo, (1) mejora de la calidad de la reconstrucción de la imagen de una escena estática, (2) incremento de la precisión en la clasificación usando mediciones espectrales comprimidas, y (3) diseño de un enfoque de estimación de movimiento utilizando una representación escasa en escenas dinámicas (ultrasonido y video multiespectral), en el caso de video multispectral es usado para diseñar aperturas codificadas de forma adaptada para mejorar la calidad de la reconstrucción de la imagen en escenas dinámicas.

---

[1] Tesis doctoral

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

Abstract

Title: Coded aperture design for adaptive compressive spectral imaging [*]

Author: NELSON EDUARDO DÍAZ DÍAZ [**]

Keywords: Compressive spectral imaging, adaptive imaging, coded aperture imaging, computational imaging.

Description: Sensing a spectral data cube has been traditionally a time-consuming task since it requires a scanning process. In contrast, compressive spectral imaging (CSI) has attracted widespread interest since it requires fewer samples than scanning systems to acquire the data cube, thus improving the sensing speed. CSI captures linear projections of the scene, and then a reconstruction algorithm estimates the underlying scene. One notable CSI architecture is the colored coded aperture snapshot spectral imager (C-CASSI), which employs pixelated filter arrays as the coding patterns, to spatially and spectrally encode the incoming light. Up to date works on C-CASSI have used non-adaptive colored-coded apertures. Non-adaptive sampling ignores prior information about the signal to design the coding patterns. Therefore, this work proposes four different methods to adaptively design the colored coded aperture applied to different problems, including, (1) quality improvement of image reconstructions of a static scene, (2) Improvement of classification accuracy using compressive spectral measurements, and (3) a motion estimation approach using sparse representation in dynamic scenes, (4) to utilize the motion estimation to adaptively design coded apertures to improve quality of image reconstruction in dynamic scenes.

---

[*] Doctoral Thesis

[**] Department of Electrical, Electronics, and Telecommunication Engineering. Department of Electrical, Electronics, and Telecommunication Engineering. Advisor: Henry Arguello Fuentes.

Introduction

Traditional spectral imaging acquisition. Spectral imaging (SI) collects 3D spatio-spectral information of a scene, referred to as data cube, with two spatial dimensions and one spectral dimension. Each spatial position along the spectral axis is known as a spectral signature, which can be used to determine the components present in the scene. SI has attracted widespread interest due to the numerous applications, for example, in cultural heritage for non-invasive characterization and identification of the deterioration state of materials (France, 2011), in medicine for the diagnosis of oral cancer (Bedard et al., 2013), and in food safety to detect pathogens (Coffey, 2015). Traditionally, SI systems rely on scanning techniques to collect the data cube (Sellar and Boreman, 2005a,b). For instance, push-broom is a line scanning methodology that moves across a spatial dimension to construct the data cube; by concatenating 2D spatio-spectral slices, whisk-broom is a point-scanning method that moves across the two spatial dimensions capturing each spectrum at a time; and the tunable spectral filter is a wavelength scanning methodology that constructs the data cube by capturing each spectral band at a time. Figure 1 depicts an example of a datacube, and the different scanning techniques.

Traditional snapshot spectral imaging. Collecting a data cube using conventional scanning techniques is a time-consuming effort, non-suitable for varying dynamic targets. In contrast, snapshot systems are faster than scanning systems because they capture all the information about the scene at once. Hagen in (Hagen and Kudenov, 2013) traces the advances in the evolution of snapshot spectral imaging technologies. He identifies the 13 most representative snapshot spec-

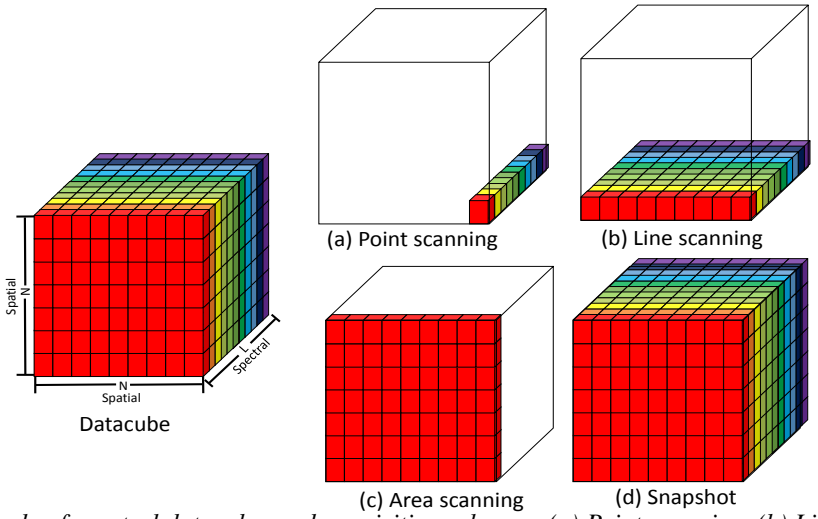


Figura 1 An example of spectral datacube, and acquisition schemes. (a) Point scanning. (b) Line scanning. (c) Area scanning. (d) Snapshot.

tral imaging technologies. Among them, the integral field spectrometry (IFS) has three variations: the first, mirror arrays IFS-M, the second, fibers array IFS-F, and the third, lenslet array IFS-L, (Bowen, 1938; Kapany, 1968; Courtès, 1960). The technology is called integral field because each measurement is attained by integrating the light over a region of the field. The disadvantage of IFS is its low spatial resolution. The fourth system in this list is the Multispectral beam splitting (MSBS) which splits the light into different color bands; however, it is limited to five or six spectral bands (Stoffels et al., 1978). The fifth is the Computer tomography imaging spectrometer (CTIS) is a compact architecture, but its major disadvantage is the difficulty in manufacturing the kinoform dispersing elements (Okamoto and Yamaguchi, 1991). Another technology, the Multiaperture filtered camera (MAFC) is a simple approach for multispectral imaging, but it requires a specific light distribution in order to work correctly (Hirai et al., 1994). Another technology is the Tunable echelle imager (TEI) which gathers individual images of the spectral channels at the detector, and the trade-off is that the system throughput is quite low (Baldry and Bland-Hawthorn, 2000).

Another technology is the Spectrally resolving detector arrays (SRDA) in which the number of spectral filters increases concerning previous architectures. It is exceptionally compact since the filtering occurs in the detection layer. SRDA has two main drawbacks; the first is that the system can not easily adjust the spectral resolution, and the second is the loss in light throughput (Bilbro, 2001). Another usual technology is the Image-replicating imaging spectrometer (IRIS), which is a generalization of the tunable filter in which the polarizer is replaced by Wollaston beam-splitting polarizers. By incorporating N Wollaston polarizers, the system can view 2^N scenes simultaneously. However, IRIS has a low spectral resolution, which is limited to up to eight spectral bands (Harvey and Fletcher-Holmes, 2003). A recent SI system is the image mapping spectrometry (IMS) in which a microfacet mirror is placed at the image plane. The pattern in the detector resembles the scene through a picket fence. The drawback of IMS is that it requires high precision for cutting the image mapper (Gao et al., 2009).

Another classical approach is the Snapshot hyperspectral imaging Fourier transform spectrometer (SHIFT) based on two birefringent Normarski prisms behind a $N \times M$ lenslet array, forming $N \times M$ sub-images in the detector array. SHIFT is easy to fabricate, but it is affected by the parallax effect (Kudenov and Dereniak, 2012). Another method is the multispectral sagnac interferometer (MSI) that measures the interference fringes in the detector array as a superposition coincident amplitude-modulated spatial carrier frequencies (Kudenov et al., 2010). The advantage of MSI is its simple postprocessing and spatial coregistration between bands. In contrast, the drawback of MSI is the correspondence of the spectral bands and the grating's order and only one dimension of the Fourier space can be used to code the spatial and spectral information.

Snapshot compressive spectral imaging. In 2008 a remarkable approach was presented, the coded aperture snapshot spectral imaging (CASSI) which is the first imager based on compressive spectral imaging (Wagadarikar et al., 2008), CASSI is detailed in section 2.3.1. The most outstanding architectures based on CASSI are shown in section 2.3. The first variation uses spectral filters to encode the scene, and it depicts in section 2.3.2. Another variation of CASSI without dispersive element shows in section 2.3.3. In addition, a video architecture based on CASSI is shown in section 2.3.5.

CASSI is a snapshot system that captures compressive projections of the data cube and then exploits CS concepts to estimate the data cube (Arce et al., 2014),(Arguello and Arce, 2013),(Mejia and Arguello, 2016),(Wagadarikar et al., 2008), (Wu et al., 2011a). In their cutting-edge paper published in 2006, Candes, Tao, and Donoho demonstrate that it is possible to recover a s-sparse n-dimensional signal from $s\log(n)$ randomly chosen projections. For that reason, the number of compressive measurements of CASSI is fewer than that of conventional snapshot architectures. The main elements in CASSI are the coded aperture, the dispersive element, and the detector. In CASSI, the coded aperture modulates spatially the incoming light of the scene, then the enconded light is smeared by the dispersive element. The modulated and smeared intensity is then multiplexed in the Focal Plane Array (FPA) forming the compressive measurements.

Traditionally, CASSI uses a block-unblock coded aperture which blocks or transmits the full spectral signature in a given pixel. Recently, the colored CASSI system (C-CASSI) has gained attention due to its flexibility to filter the incoming light per pixel (Arguello and Arce, 2014; Rueda et al., 2015; Hinojosa et al., 2016) with the use of a colored-coded aperture (CCA). A CCA is a 2D

pixelated filter array, which not only encodes the scene spatially but also spectrally. In comparison with block-unblock coded apertures, CCA permits the selection of portions of the spectrum that traverse the system. A CCA is equivalent to a 3D block-unblock coded aperture, where each slice of the 3D coding pattern represents the filtering applied to each spectral band. Figure 2 shows the equivalence between a 3D block-unblock coded aperture and the colored coded aperture. A vector of block-unblock elements represents each spectral filter, and the size of the vector denotes the spectral resolution. Conventional C-CASSI uses low pass, bandpass, and high pass filters to modulate the incoming light.

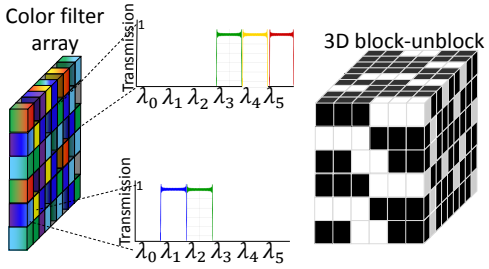


Figura 2 Equivalence between 2D colored coded aperture and 3D block-unblock coded aperture. The colored filters array is composed of spectral filter that modulate the incoming light in the system.

Adaptive compressive spectral imaging. Up-to-date work on C-CASSI has focused on non-adaptive measurements (Arguello and Arce, 2014; Rueda et al., 2015). Specifically, adaptive measurements refer to the use of prior knowledge of the data cube in the design of the entries of the CCA. This means that during the acquisition $\mathbf{y}^{\ell} = \mathbf{H}^{\ell}\mathbf{f}^{\ell} + \boldsymbol{\omega}$ of K snapshots, the compressive measurements $\mathbf{y}^{\ell}$ are attained using the sensing matrix $\mathbf{H}^{\ell}$ which samples the signal $\mathbf{f}$, the samples are contaminated with additive noise. The measurement $\mathbf{y}^{\ell}$ is acquired avoiding to oversample previous captured compressive measurements, $\mathbf{y}^{1}, \cdots, \mathbf{y}^{\ell}$ which are sequentially measured

according to the structure of the scene, but where the $\mathbf{y}^{\ell}$-th snapshot depends on the previously gathered measurements $\mathbf{y}^{\ell-1}$. Adaptive sensing has been used in the past in areas such as high dynamic range imaging to improve the quantization resolution and to avoid saturation (Diaz et al., 2015). Between the adaptive methodologies in SI, there are three different approaches. The first approach is dictionary learning (Yang et al., 2016), and consists on learning a dictionary from the data, then computing a singular value decomposition from the dictionary, and finally use a small number of left singular vectors as the measurement matrix $\mathbf{H}$, to collect the projection of the scene. The reconstruction is then performed using a conventional CS signal recovery method. It shows that adaptive sampling using dictionary learning performs better than conventional CS methods, according to (Yang et al., 2016). The adaptive sampling in the Wavelet domain developed by Averbuch et al. (Averbuch et al., 2012) shows that it is possible to sample the Wavelet coefficients with a Digital Micromirror Device (DMD) for the single-pixel architecture. This method acquires low-resolution measurements, and then, based on the previous measurements and using the properties of the Wavelet structure, the algorithm adaptively extracts the significant information corresponding to the edges of the image. The adaptive sampling based on the Bayes method relies on the inversion of the compressive measurements $\mathbf{y} = \mathbf{H}\boldsymbol{\Psi}^{-1}\boldsymbol{\theta} + \boldsymbol{\omega}$ from a Bayesian perspective. The prior belief relies on the idea that $\boldsymbol{\theta}$ is a sparse signal that might be represented in some basis $\boldsymbol{\Psi}$. The objective is to provide a posterior belief of the entries of $\boldsymbol{\theta}$ (Ji et al., 2008). Specifically, Bayesian compressive sensing estimates the covariance matrix using as prior information the measurement matrix and compressive measurements, with that information the next sensing matrix is inferred.

The fundamental characteristic of adaptive measurements is that they are more outstandingly robust to Gaussian noise than traditional non-adaptive measurements (Malloy and Nowak, 2014). Furthermore, the intuition in adaptive sampling is always to use previous knowledge to adjust tunable parameters related to the CS acquisition system, such as the covariance matrix or a low-resolution reconstruction. The advantages of adaptive sampling include coding according to the scene, robustness to noise, and better quality of image reconstruction in comparison with conventional non-adaptive sampling systems.

Open challenges. Three challenges are identified in this dissertation:

1. **To improve the quality of image reconstruction in static scenes in C-CASSI**: In CSI architectures the coded aperture can be random (Gehm et al., 2007), or designed (Arguello and Arce, 2013), (Arguello and Arce, 2014). However, both approaches are non-adaptive, which means there is no assumption about the underlying image. Recently, Malloy suggests that the adaptive approach would be beneficial in the sampling process (Malloy and Nowak, 2014). In particular, the algorithms based on adaptive sampling succeeded in recovering the signal at the lowest signal to noise ratio (SNR). The adaptive approach might be used in scenarios with noise, such as CSI in order to improve the quality of image reconstruction.

2. **The classification accuracy in the 3D-CASSI and C-CASSI**: Traditionally, the compressive measurements of CSI imaging architectures lose the spatial context of the scene of interest. This issue occurs with compressive measurements obtained from random patterns (Gehm et al., 2007), and using optimal designed patterns (Arguello and Arce, 2013), (Argue-

llo and Arce, 2014). Not preserving the spatial context of the scene interest harms the classification performed from the compressive measurements. A recent approach for SI highlights the importance of preserving the spatial context of SI scenes (Zhao et al., 2016), (Falco et al., 2015). The use of spatial information can be used to improve classification in CSI.

3. **The quality of image reconstruction in dynamic scenes in C-CASSI**: Surveillance videos have redundant information because some parts of the scene keep static while others change with time. The compression of videos with such characteristic can be improved if a system can identify the static and dynamic part of the scene. This process can be performed adaptively during the sampling of the video estimating the motion. Recently, motion estimation methods have been proposed for cardiac ultrasound images (Ouzir et al., 2018), (Ouzir et al., 2019). These cardiac motion estimation approaches might extend to video CSI.

Chapters 3, 4, 5 and 6 of this dissertation address the three proposed challenges by adaptively designing the coded apertures for each particular case i.e. to improve the quality of image reconstruction in static scenes. To enhance classification accuracy, and to increase the quality of image reconstruction in dynamic scenes.

General objective

To adaptively design coding patterns to improve the quality of image reconstruction of a compressive spectral imaging architecture.

Specific objectives.

1. To simulate the coding patterns of compressive spectral imaging architectures of state of the

art, such as CASSI, C-CASSI, and SSCSI.

2. To develop a mathematical model of the adaptive coding patterns and its effect on the quality of image reconstruction in compressive spectral imaging architectures.

3. To evaluate the designed adaptive coding patterns in compressive spectral imaging architectures and compare them against non-adaptive and optimized patterns.

4. To implement a DMD-based optical prototype implementation in the lab of the designed adaptive coding patterns of one state-of-the-art CSI architecture.

Contribution. The approaches proposed in this dissertation tackle each challenge making assumptions about the signal to improve the quality of image reconstruction or enhance the accuracy in the classification. In each approach, the coded aperture is designed adaptively using some prior information.

The first approach proposes an algorithm to adaptively design the coded aperture of the C-CASSI to improve the quality of image reconstruction in static scenes. The coded aperture is designed adaptively by reconstructing a low-resolution spectral image from the compressive measurements. The approach performs a thresholding of the second derivative of each spectral signature to compute the coded aperture at each snapshot [C1,C4], [J2] (specific objectives 1,1,1,1).

The second approach proposes an algorithm to adaptively design the coded apertures of a compressive spectral imaging architecture of two arms. The simulation is performed using the 3D-CASSI, and C-CASSI. The first arm captures compressive measurements of high spatial resolution and low spectral resolution (multispectral measurements). The second arm samples compressive

measurements of low spatial resolution and high spectral resolution (hyperspectral measurements). The motivation of the adaptive scheme is to improve the classification accuracy. The proposed approach preserves contextual spatial information. The adaptive design of complementary coded apertures samples the consecutive bands to reduce the shifting induced by the dispersive element [J3] (specific objectives 1,1,1).

The third approach proposes an algorithm to adaptively design the coded aperture to improve the quality of spectral images. The approach is based on a motion estimation algorithm used in ultrasound images (UI) [C2]. The method has three terms, one corresponding to the data fidelity, other to promote the smoothness of the motion, and the last one to attain a sparse representation of the UI. The proposed approach uses the convolutional sparse representation to model images UI. This approach is also used in Compressive Spectral Images (CSI) to estimate the motion in spectral images reconstructed from compressive spectral measurements of the 3D-CASSI [C3]. Via the motion estimation, the filter array is designed in order to reduce the samples of the moving object using a mobile complementary blue noise coded aperture [J2] (specific objectives 1,1,1).

Dissertation outline

This work is devoted to employing the adaptive approach in different topics of CSI. In particular, the dissertation is divided into five chapters: Chapter 2 gives a brief overview of the most representative CSI architectures. Specifically, the mathematical model of C-CASSI and the non-adaptive matrix representation of the sampling phenomenon. Chapter 3 proposes an adaptive algorithm denominated gradient thresholding algorithm (GTA) to adaptively design the array of color filters adaptively. Simulation results show that the proposed adaptive C-CASSI improves

the quality of image reconstruction in comparison with the non-adaptive C-CASSI in up to 3 dB in terms of PSNR. Moreover, adaptive C-CASSI is implemented in the laboratory to attain real experimental data. The adaptive versus the conventional non-adaptive methodologies are compared in this. Chapter 4 are compared in this paper presents a dual-arm CSI architecture to perform classification using the compressive measurements. The architecture has two sensors which collect compressive measurements with different resolution. The first arm acquires high spatial resolution and low spectral resolution. The second arm measures low spatial resolution and high spectral resolution. Simulation results show that 99.5 % in the overall accuracy. In chapter 5 a cardiac motion estimation algorithm based on convolutional sparse coding is proposed for images of ultrasound. The algorithm is comparable in terms of mean point error with an algorithm of state of the art and also overcomes three of the algorithms of state of the art. In chapter 6 the proposed motion estimation algorithm of the previous chapter is used to design the colored coded apertures for a dynamic scene.

1. Objetivos

General objective

To adaptively design coding patterns to improve the quality of image reconstruction of a compressive spectral imaging architecture.

Specific objectives

To simulate the coding patterns of compressive spectral imaging architectures of state of the art, such as CASSI, C-CASSI, and SSCSI ;

To develop a mathematical model of the adaptive coding patterns and its effect on the quality of image reconstruction in compressive spectral imaging architectures ;

To evaluate the designed adaptive coding patterns in compressive spectral imaging architectures and compare them against non-adaptive and optimized patterns

To implement a DMD-based optical prototype implementation in the lab of the designed adaptive coding patterns of one state-of-the-art CSI architecture .

2. Theoretical background: Compressive spectral imaging

2.1. Compressive sensing

Compressive sensing (CS) can recover a signal with just a few projections. More specifically, the compressive measurements are given by $\mathbf{y} = \mathbf{\Phi}\boldsymbol{\theta} + \boldsymbol{\omega}$ where $\mathbf{y} \in \mathbb{R}^v$, $\mathbf{\Phi} \in \mathbb{R}^{v \times n}$ is the the sensing matrix denoted by $\mathbf{\Phi} = \mathbf{H}\mathbf{\Psi}$ where $\mathbf{H} \in \mathbb{R}^{v \times n}$ is the measurement matrix[1], $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$ is the representation basis, $\boldsymbol{\theta} \in \mathbb{R}^n$ is a n dimensional signal and $\boldsymbol{\omega} \in \mathbb{R}^v$ is the Gaussian noise. Typically, the signals in nature are not sparse, but they can be sparse in some representation basis. In more detail, $\mathbf{f}$ can be s-sparse in some basis $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$, such that $\mathbf{f} = \mathbf{\Psi}^{-1}\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a n-dimensional, s-sparse signal, with, $s \ll n$. Figure 3 depicts the general CS scheme. The underlying signal $\mathbf{f}$ can be recovered from the projections $\mathbf{y}$, by solving the convex optimization problem

$$\hat{\mathbf{f}} = \mathbf{\Psi}^{-1}(\underset{\boldsymbol{\theta}}{\operatorname{argmín}} \|\mathbf{y} - \mathbf{\Phi}\boldsymbol{\theta}\|_2 + \tau\|\boldsymbol{\theta}\|_1), \tag{1}$$

where $\|.\|_2$ is the l_2 norm, $\|.\|_1$ is the l_1 norm, and τ is a regularization parameter.

---

[1] Throughout this proposal, the terms coding pattern, coded aperture, and coded aperture pattern are used interchangeably to refer to measurement matrix.

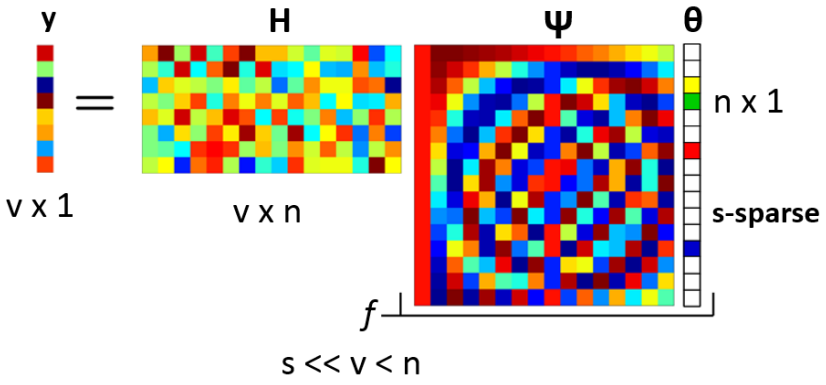


Figura 3 Scheme of compressive sensing.

Traditionally, CS does not rely on any assumption about the underlying signal, which we call non-adaptive CS (Candes et al., 2006), (Donoho, 2006), (Baraniuk, 2007), (Romberg, 2008), (Romberg, 2009), (Wang and Arce, 2010), (Tsaig et al., 2012). However, nowadays, there is a better understanding about the advantages of adaptive schemes (Malloy and Nowak, 2014). Two aspects are important, the number of compressive measurements, v, and the minimum magnitude of sparse signal, θ_{min}, correctly recover a signal. According to Table 1, adaptive and non-adaptive schemes require the same number of compressive measurements, i.e. $s\log(n)$ for a s-sparse, n-dimensional signal. In non-adaptive CS, to attain a correct reconstruction, the SNR must grow according to $\log(n)$ (Wainwright, 2009a; Aeron et al., 2010; Wainwright, 2009b). That is, the smallest non-zero entry of $\boldsymbol{\theta}$ must exceed the constant $\sqrt{C\frac{n}{M}\log(n)}$, where $\sum_{i=1}^{v}\|\phi_{\mathbf{i}}\|_2^2 = M$. In contrast, to attain a correct recovery in adaptive CS, the SNR must grow according to $\log(s)$ (Castro, 2012). That is, the smallest non-zero entry of $\boldsymbol{\theta}$ must exceed the constant $\sqrt{C\frac{n}{M}\log(s)}$. It means that adaptive CS is more robust to noise than non-adaptive CS.

Table 1
Asymptotic requirements for exact support recovery (Malloy and Nowak, 2014).

	non-adaptive	**adaptive**
direct	1. $v \geq n$ 2. $\theta_{min} > \sqrt{2\frac{n}{M}\log n}$	1. $v > n$ 2. $\theta_{min} > \sqrt{2\frac{n}{M}\log s}$
CS	1. $v \geq s\log n$ 2. $\theta_{min} > \sqrt{C\frac{n}{M}\log n}$	1. $v \geq s\log n$ 2. $\theta_{min} > \sqrt{C\frac{n}{M}\log s}$

2.2. Compressive spectral imaging

Compressive spectral imaging (CSI) denotes novel snapshot methods (Cao et al., 2016), which utilize the advantages of under-sampling techniques and constrained reconstruction algorithms to deal with the high dimensionality in the spatial, spectral and temporal dimensions. CSI exploits the theory of CS to spectral imaging. The key idea is that the spectral image can be modeled as the linear combination between a representation basis $\mathbf{\Psi}$ and a sparse vector $\boldsymbol{\theta}$ the signal has to be incoherent with the basis. The figure 4 illustrates the sampling process of the multispectral image $\mathbf{f}$ using sensing matrix $\mathbf{H}$, which attain the compressive measurements $\mathbf{y}$.

The sensing matrix represents the spatial encoder or coded aperture of the CSI system. The coded aperture can be implemented using piezo-electric (Kittle et al., 2010), (Wu et al., 2011b), spatial light modulator (SLM) (Zhu et al., 2013), a digital micromirror device (DMD) (Wu et al., 2011b). The DMD as is depicted in figure 5 is an array that contains thousand of microscopic mirrors to vary the structure of the coded aperture. Each micromirror can be rotated $+10/-10$

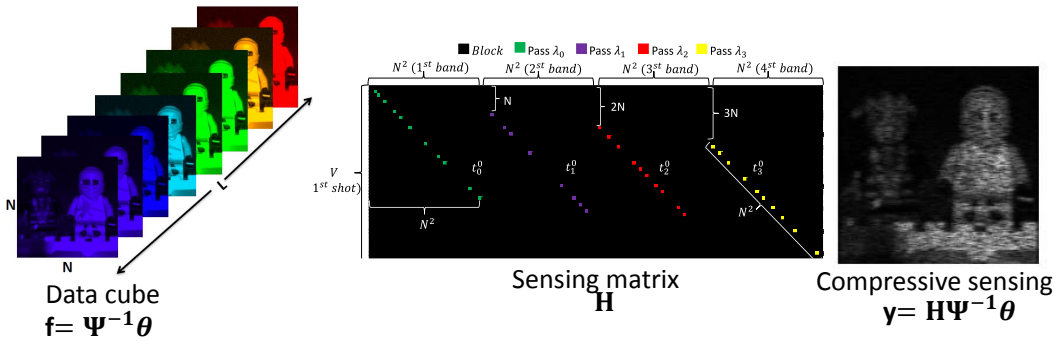


Figura 4 Acquisition of compressive measurements. The signal is sampled using the sensing matrix resulting in the compressive measurements.

degrees or equivalently activated/deactivated state. The activated state reflects the light in the direction of the pick-up sensor, and the deactivated state in another direction outside the sampling path, generating a dark pixel in the pick-up sensor. In this way, the DMD allows modulating the light intensity, through the programming of its micro-mirrors. Figure 6 depicts two coded apertures with different transmittance. The transmittance corresponds to the amount of light that transfer the coded aperture.

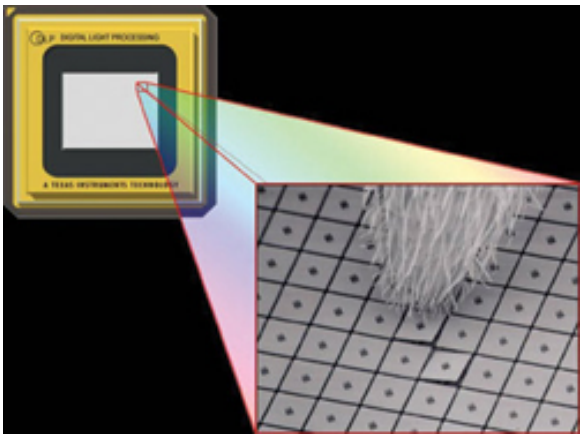


Figura 5 Light modulator device. The image on the right compares the size of an ant's leg with respect to the tiny micro mirrors of a DMD.

Figura 6 Two examples of coded apertures with 75% (left), and 25% (right) of transmittance, respectively.

2.3. Compressive spectral imaging architectures

Since the CASSI invention in 2008 (Wagadarikar et al., 2008), new modifications have been proposed. Two main systems are described in this section. The first exploits the snapshot potential, the second replaces the block-unblock coded aperture with an array of optical filters. In 2010, a modification to CASSI was introduced, which accepts multiple snapshots with the advantage of increasing the quality of image reconstruction (Kittle et al., 2010), (Arguello and Arce, 2011),(Arguello et al., 2012), (Arce et al., 2014). Additionally, in 2012 the traditional block-unblock coded aperture was replaced by an array of optical filters with the benefit of a more flexible modulation design. In addition to CASSI, other CSI architectures have been developed. For instance, compressive hyperspectral imaging by separable spatial and spectral operators (CHISSS) (August et al., 2013), Dual-coded compressive hyperspectral imaging (DCSI) (Lin et al., 2014b), and the spatial-spectral encoded compressive hyperspectral imager (SSCSI) (Lin et al., 2014a), snapshot colored compressive spectral imager (SCCSI) (Correa et al., 2015). Moreover, other CSI architectures are detailed in (Wang et al., 2015b), (Yuan et al., 2015), (Golub et al., 2016), (Starling et al., 2016).

In the following sections, the most relevant CSI architectures are detailed in chronological order. For each CSI system based on CASSI, the sketch is presented, the continuous, discrete model and the multishot model. The coded aperture snapshot spectral imaging (CASSI) is detailed in section 2.3.1 and colored coded aperture snapshot spectral imaging (C-CASSI) is described in section 2.3.2. Section 2.3.3 and 2.3.5 presents the 3D coded aperture snapshot spectral imaging (3D-CASSI) and the Video colored coded aperture snapshot spectral imaging (Video C-CASSI), respectively. Although, less mathematical detail is put in 3D-CASSI and Video C-CASSI because it is straightforward to model the 3D-CASSI as C-CASSI without shifting due to the lack of a dispersive element. The video C-CASSI can be modeled similarly to the C-CASSI multi-snapshot. A clear understanding of the CSI based on CASSI is important to allow the implementation of the adaptive approach.

2.3.1. Coded aperture snapshot spectral imaging (CASSI). CASSI is a single snapshot architecture that captures 2D compressive projections of the scene (Wagadarikar et al., 2008). Figure 7 depicts a scheme of CASSI, which includes an objective lens that collects the light from the scene and directs it towards the coded aperture, that spatially modulates the incoming light. Then, the encoded light is dispersed by the prism. The spatio-spectral multiplexed and sheared light is integrated in the detector array. The registered intensities in the detector are called compressive measurements. The system can be also modeled as $\mathbf{y} = \mathbf{H}\boldsymbol{\Psi}^{-1}\boldsymbol{\theta}$ where $\mathbf{y} \in \mathbb{R}^{v}$, $\mathbf{H} \in \mathbb{R}^{v \times n}$ is the measurement matrix, with $v = N \times (N + L - 1)$, $n = N \times N \times L$, and $\boldsymbol{\Psi}^{-1} \in \mathbb{R}^{n \times n}$ is the representation basis and $\boldsymbol{\theta} \in \mathbb{R}^{n}$ is the s-sparse signal where $s \ll v < n$.

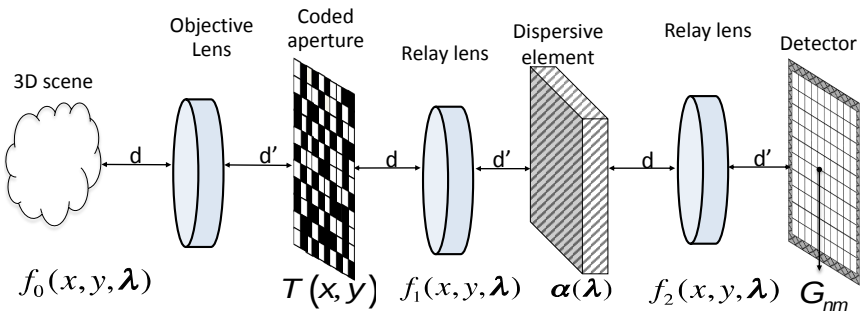


Figura 7 Sketch of the CASSI architecture.

The continuous model of CASSI is given by

$$G(x,y) = \int T(x+\alpha(\lambda-\lambda_c),y) f_0(x+\alpha(\lambda-\lambda_c),y,\lambda) d\lambda, \tag{2}$$

where $G(x,y)$ are the continous measurements, denoting the sum of encoded images $T(x,y)$ denotes the 2D coded aperture, $f_0(x,y,\lambda)$ represents the spatio-spectral information and $\alpha(\lambda-\lambda_c)$ the dispersion induced by the dispersive element. The pitch of the detector is represented by Δ. The discrete measurements are given by:

$$G_{p,q} = \int\int G(x,y) \text{rect}\left(\frac{x}{\Delta}-p, \frac{y}{\Delta}-q\right) dxdy, \tag{3}$$

where $G_{p,q}$ samples the continuous measurements $G(x,y)$, which corresponds the sum of encoded

spectral bands, the sampling uses the function rectangular rect(.) to measure the projections. Then, each entry of the discrete measurement is denoted by:

$$Y_{p,q} = \sum_{k=0}^{L-1} T_{(p,q+k)} F_{p,q+k,k} + \omega_{p,q}, \tag{4}$$

where $T_{p,q}$ denotes the block-unblock coded aperture, $F_{p,q,j}$ represents the discrete representation of the spectral source $f_0(x,y;\lambda)$, and $\omega_{p,q}$ is the inherent noise associated to the system.

2.3.2. Colored coded aperture snapshot spectral imaging system (C-CASSI).

Figure 8 shows a scheme of C-CASSI. The colored coded aperture $T(x,y,\lambda)$ modulates the incoming light $F(x,y,\lambda)$, where (x,y) represent the spatial coordinates and λ the wavelength (Arguello and Arce, 2013). The coded density is then spread out by the dispersive element and captured in the sensor, as

$$Y(x,y) = \int\int\int T(x',y',\lambda) F(x',y',\lambda) h(x' - S(\lambda) - x, y' - y) dx' dy' d\lambda, \tag{5}$$

where $h(x' - S(\lambda) - x, y' - y)$ denotes the optical impulse response of the system, with $S(\lambda)$ being the dispersion generated by the prism. Each discrete voxel of the data cube, denoted as F_{ijk}, can be represented from the continuous source as

$$F_{ijk} = \int_{\lambda_k}^{\lambda_{k+1}} \int_{j\Delta}^{(j+1)\Delta} \int_{i\Delta}^{(i+1)\Delta} F(x,y,\lambda) dx dy d\lambda = c_{ijk} \cdot F(x_i, y_j, \gamma_k), \tag{6}$$

for $i, j = 0, \ldots, N-1$, $k = 0, \ldots, L-1$ where c_{ijk} denotes the quadrature weight, and (x_i, y_j, γ_k) index the coordinates of the $(i, j, k)^{\text{th}}$ voxel, where i and j respresent the indices of the spatial coordinates and k the spectral axis. Let Δ be the pixel pitch of the detector and let λ_k the spectral discretization. The range of the k^{th} spectral band is given by$[\lambda_k, \lambda_{k+1}]$, where λ_k is the solution to the equation $S(\lambda_{k+1}) - S(\lambda_k) = \Delta$.

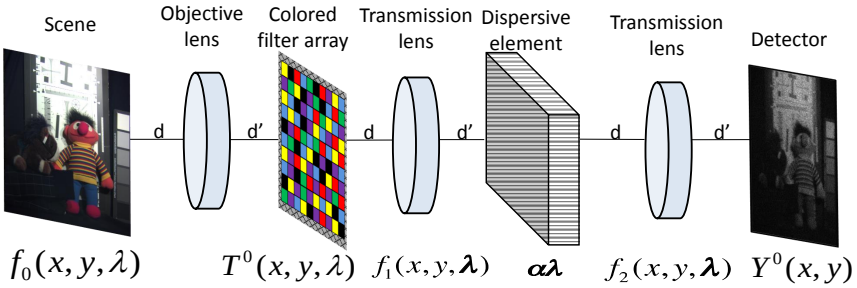


Figura 8 Sketch of the C-CASSI architecture.

Let $T^{\ell}_{i,j,k} \in \{0,1\}$ be the discretization of the colored coded aperture $T^{\ell}(x,y,\lambda)$ used in the ℓ-th snapshot, $\ell = 0, \ldots, K-1$, such that

$$T^{\ell}(x,y,\lambda) = \sum_{ijk} T^{\ell}_{ijk} \text{rect}\left(\frac{x}{\Delta} - i - \frac{1}{2}, \frac{y}{\Delta} - j - \frac{1}{2}, \frac{\lambda}{\sigma_k} - k - \frac{1}{2}\right), \tag{7}$$

where Δ is the side length of the coded aperture, and $\sigma_k = \lambda_{k+1} - \lambda_k$. Notice that the index i is in the direction orthogonal to dispersion, and j is along the axis of dispersion. As shown in Eq. 7 the colored coded aperture can be regarded as a 3D structure T^{ℓ}_{ijk}, whereas, the traditional block-unblock coded aperture is a 2D structure. The ℓ^{th} discretized compressive measurement in C-CASSI can be written as

$$Y^{\ell}_{i,l} = \sum_{k=0}^{L-1} F_{i,(l+k),k} T^{\ell}_{i,(l+k),k} + \omega_{i,l}, \tag{8}$$

where $Y_{i,l}^{\ell}$ is the intensity in the $(i,l)^{\text{th}}$ sensor pixel, where $l \in \{0,\cdots,N+L-1\}$ and $\omega_{i,l}$ is the Gaussian noise in the $(i,l)^{\text{th}}$ position, $i = 0,\cdots,N-1$. Notice that index l is along the axis dispersion. The measurements in the ℓ_{th}snapshot can be arranged as $\mathbf{Y}^{\ell} \in \mathbb{R}^{N\times(N+L-1)}$. Similarly, the underlying scene can be denoted as $\mathbf{F} \in \mathbb{R}^{N\times N\times L}$, and $\mathbf{T} \in \mathbb{R}^{N\times N\times L}$ is a binary three dimensional array modeling the colored coded aperture. The compressive measurements in C-CASSI can then be expressed in matrix form as

$$\mathbf{y}^{\ell} = \mathbf{H}^{\ell}\mathbf{f} + \boldsymbol{\omega}, \tag{9}$$

where $\mathbf{H}^{\ell}$ stands for the ℓ^{th} sensing matrix, $\mathbf{f}$ denotes the vectorization of the underlying data cube $\mathbf{F}$, and $\boldsymbol{\omega}$ is the vectorization of the Gaussian noise. The vectorization of the matrix $\mathbf{F}$ is given by $(\mathbf{f}_k)_p^{\ell} = \mathbf{F}^{\ell}_{(p-rN)rk}$, for $p = 0,\cdots,N^2-1$, $r = \lfloor p/N \rfloor$. Equation 9 can be rewritten as,

$$\mathbf{y} = \overbrace{\begin{bmatrix} \text{diag}(\mathbf{H}_0^{\ell}) & \text{diag}(\mathbf{0}^0_{N(1)\times N^2}) & \dots & \text{diag}(\mathbf{0}^0_{N(L-1)\times N^2}) \\ & \text{diag}(\mathbf{H}_1^{\ell}) & \dots & \\ & & \ddots & \\ \text{diag}(\mathbf{0}^0_{N(L-1)\times N^2}) & \text{diag}(\mathbf{0}^0_{N(L-2)\times N^2}) & \dots & diag(\mathbf{H}_{L-1}^{\ell-1}) \end{bmatrix}}^{\mathbf{H}^{\ell}} \begin{bmatrix} \mathbf{f}_0 \\ \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_{L-1} \end{bmatrix} + \boldsymbol{\omega}, \tag{10}$$

where $\text{diag}(\mathbf{H}_k^{\ell})$ is a $N^2 \times N^2$ diagonal matrix whose entries are the elements of the ℓ^{th} vectorized colored coded aperture of the k^{th} spectral band.

Figure 9 depicts an example of the sensing matrix $\mathbf{H} = [(\mathbf{H}^0)^T, (\mathbf{H}^1)^T, \dots (\mathbf{H}^{K-1})^T]^T$, for $K = 2$, $N = 6$, and $L = 4$. Notice that $\mathbf{H}$ corresponds to a sparse matrix since it is mostly populated

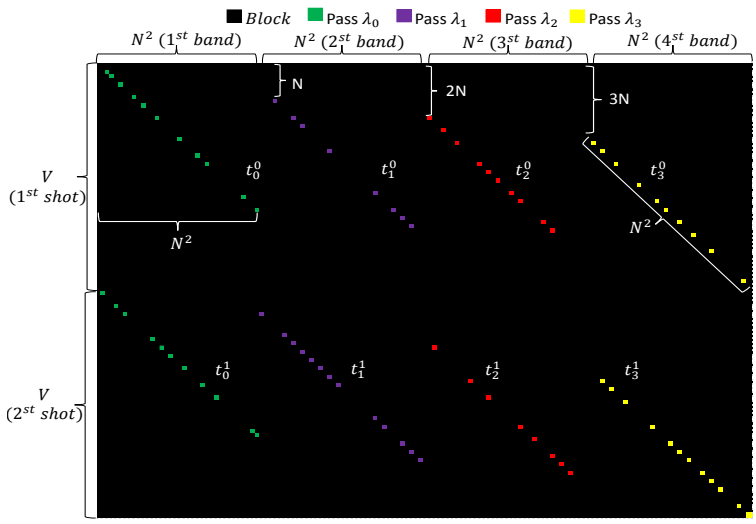


Figura 9 Sensing matrix **H** *where* $K = 2$, $N = 6$ *and* $L = 4$, *and* $V = N(N+L-1) = 54$. *Black squares represent a zero-value element (blocking light). Green, magenta, red, and yellow squares represent one-value element (unblocking light) in the corresponding spectral band* λ_0, λ_1, λ_2, *and* λ_3.

with zeros, thus, its very large dimensions are easily manageable. t_k^ℓ is a vector that represents the non-zero elements in the k^{th} spectral band, and the ℓ^{th} snapshot. The black squares represent blocking (zero-valued) elements, the green, magenta, red, and yellow squares represent transmissive (one-valued) elements at the respective wavelengths. C-CASSI architecture is utilised in chapter 3.

2.3.3. 3D coded aperture snapshot spectral imaging system (3D-CASSI). The 3D-CASSI is proposed in (Cao et al., 2016). The 3D-CASSI preserves the filter array which modulates spatial and spectrally the incoming light. This architecture does not have a dispersive element, such as prism or grating. The advantage of the architecture is that the spatial edges are better conserved. The 2D compressive measurements can be represented by

$$Y_{i,j}^{\ell} = \sum_{k=0}^{L-1} F_{i,j,k} T_{i,j,k}^{\ell} + \omega_{i,j}, \tag{11}$$

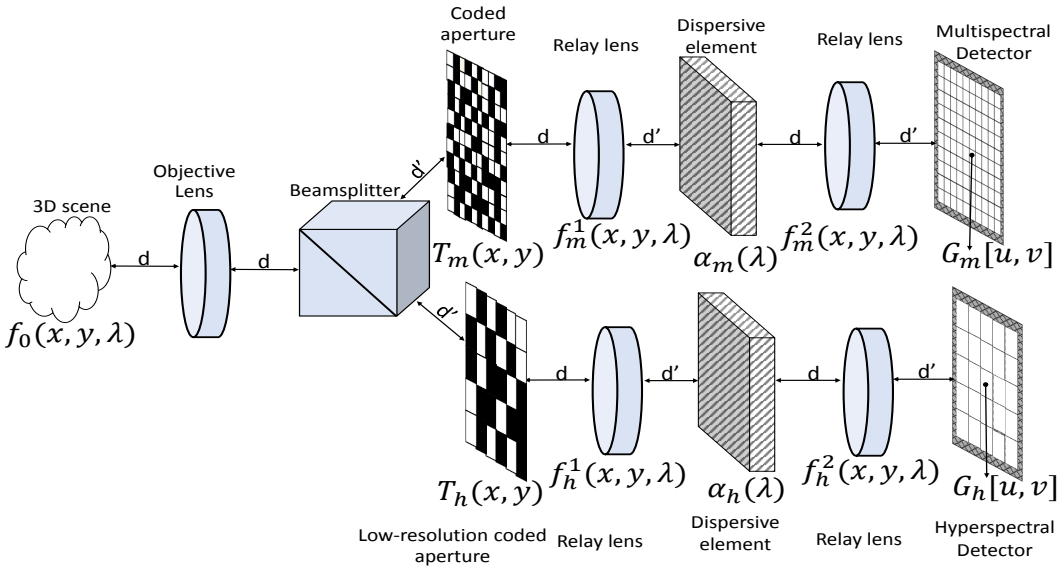


Figura 10 Dual-arm architecture with two detector of different spatial resolution.

where $F^{\ell}_{i,j,k}$ is the discretization of the colored coded aperture, being (i,j) the spatial indexes, k represents the spectral coordinate. The matrix T denotes the discretization of the colored coded aperture, where each entry of $T^{\ell}_{i,j,k} \in \{0,1\}$; $\boldsymbol{\omega}_{i,j}$ is the Gaussian additive noise which depends of the optical system.

2.3.4. Dual-arm architecture. The Fig. 10 depicts the dual-arm architecture utilise to capture the multispectral and hyperspectral compressive measurements. The mathematical model involves two C-CASSI (see section 2.3.2) with different spatial resolution. Dual-arm architecture is utilised in chapter 4.

2.3.5. Video colored coded aperture snapshot spectral imaging system (Video C-CASSI). Video colored coded aperture snapshot spectral imaging system (Video C-CASSI) acquires dynamic scenes at a particular frame rate. It is composed by an objective lens, a temporal colored coded aperture (T-CCA), a relay lens, a dispersive element and the focal plane array (FPA)

or detector. Then, the adapted system changes the CCA in C-CASSI by the T-CCA. Temporal colored coded apertures are time-varying CCA, this means each frame is spatio-spectrally modulated by a different pattern (Correa-Pugliese et al., 2016; León-López et al., 2019). The discrete compressive measurements for the d^{th} frame on the detector can be written as

$$Y^t_{i,l} = \sum_{k=0}^{L-1} F^t_{i,(l+k),k} T^t_{i,(l+k),k} + \omega_{i,l}, \tag{12}$$

where $Y^t_{i,l}$ are the elements in the matrix $\mathbf{Y}^t \in \mathbb{R}^{N\times(M+L-1)}$, that represents the attained measurements of frame t; $M \times (N+L-1)$ is the dimension of the detector and $\omega_{i,l}$ is the Gaussian noise of the sensing system at the $(i,l)^{\text{th}}$ position on the detector. The physical sensing phenomenon of the dynamic spectral source in the video C-CASSI system is depicted in Fig. 11. There, the optical elements are represented by their effect on the discretized datacube. It is possible to rewrite the acquisition process in matrix notation as $\mathbf{y} = \mathbf{H}\mathbf{f} + \boldsymbol{\omega}$, and its expansion as

$$\overbrace{\begin{bmatrix} \mathbf{y}^0 \\ \vdots \\ \mathbf{y}^t \\ \vdots \\ \mathbf{y}^{D-1} \end{bmatrix}}^{\mathbf{y}} = \overbrace{\begin{bmatrix} \mathbf{H}^0 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & & & \vdots \\ \mathbf{0} & & \mathbf{H}^t & & \mathbf{0} \\ \vdots & & & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{H}^{D-1} \end{bmatrix}}^{\mathbf{H}} \overbrace{\begin{bmatrix} \boldsymbol{f}^0 \\ \vdots \\ \boldsymbol{f}^t \\ \vdots \\ \boldsymbol{f}^{D-1} \end{bmatrix}}^{\boldsymbol{f}} + \boldsymbol{\omega}, \tag{13}$$

where $\mathbf{y}^t \in \mathbb{R}^{N\cdot(M+L-1)}$ is a vector representing the compressive measurements $\mathbf{Y}^t$ for the t^{th} frame such that $\mathbf{y}^t = \mathbf{H}^t\boldsymbol{f}^t + \boldsymbol{\omega}$; $\boldsymbol{f}^t \in \mathbb{R}^{NML}$ denotes the vectorization of the underlying datacube $\mathbf{F}^t \in \mathbb{R}^{N\times M\times L}$ at frame t; $\boldsymbol{\omega}$ is the vectorization of the Gaussian noise and $\mathbf{H}^t$ stands for the t^{th}

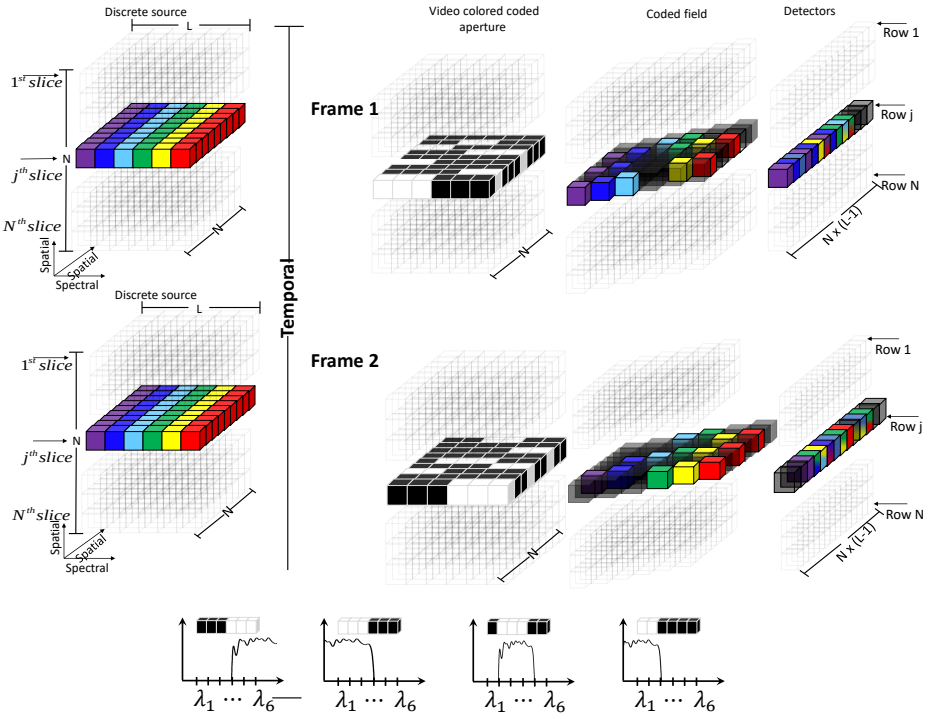


Figura 11 Illustration of the optical course for a dynamic spectral scene in the video C-CASSI architecture (León-López et al., 2019). At each frame the system encodes the incoming light with a different CCA.

C-CASSI sensing matrix whose structure, as it is defined in (Arguello and Arce, 2014), models the physical phenomenon of the employed coded aperture and the shifting produced by the dispersive element. Video C-CASSI architecture is utilised in chapter 6.

3. Adaptive filtering design for compressive spectral imaging

3.1. Introduction

This chapter presents a gradient thresholding algorithm (GTA) to adaptively compute colored-coded apertures to improve the quality of image reconstruction in C-CASSI. The adaptive C-CASSI implies to include a processing unit to perform computations with the compressive measurements and attaining the subsequently coded aperture. The intuition behind the adaptive sampling strategy is to design the CCA according to the most significant spectral signatures present in the scene, i.e., matched filter. For that reason, first, the algorithm estimates a low-resolution reconstruction of the spectral data cube, which is then interpolated to compute the second derivative of the spectral signatures. The second derivative of each spectral signature is then thresholded to detect the local maximum of each spectral signature. The transmittance of the CCA in the matched filter is established to be higher in the region of local maximum than in the local minimum sampling better the high intensities of the spectral signature. The intuition here is that adaptively sensing the matched filter, improves the quality of image reconstruction.

This chapter is divided into four sections: Section 3.2 gives a brief overview of the methods, specifically, the mathematical model of the adaptive C-CASSI, the non-adaptive matrix representation of the sampling phenomenon, the proposed adaptive system, and the gradient thresholding algorithm (GTA). In Section 3.3, simulation results show that the proposed adaptive C-CASSI outperforms the non-adaptive C-CASSI by up to 3 dB in terms of peak signal-to-noise ratio (PSNR). In Section 3.4, an implementation of the adaptive C-CASSI along with reconstructions of real

experimental data are shown, comparing the adaptive versus the conventional non-adaptive methodologies. Conclusions are presented in section 3.5.

3.2. Adaptive colored coded aperture design

The proposed method designs the CCA used in the next projection of the adaptive C-CASSI. The goal is to perform a matched filter of the reconstructed data cube, with the intuition that adaptive sensing the high intensities of the spectral signature improves the quality of image reconstruction.

3.2.1. Adaptive scheme. Figure 12 depicts the scheme of adaptive C-CASSI. In detail, the optical architecture inside the green dotted box corresponds to the C-CASSI system. The red dashed box denotes the proposed adaptive scheme to adaptively design the colored coded aperture performing an encoding according to the structure of the scene. In particular, the steps of the proposed adaptive system are (I) low-resolution reconstruction, (II) interpolation of the low-resolution data cube, (III) computation of the spectral gradient, (IV) thresholding of the spectral gradient and (V) adjustment of the transmittance of the spectral filter. (VI) resulting filter array.

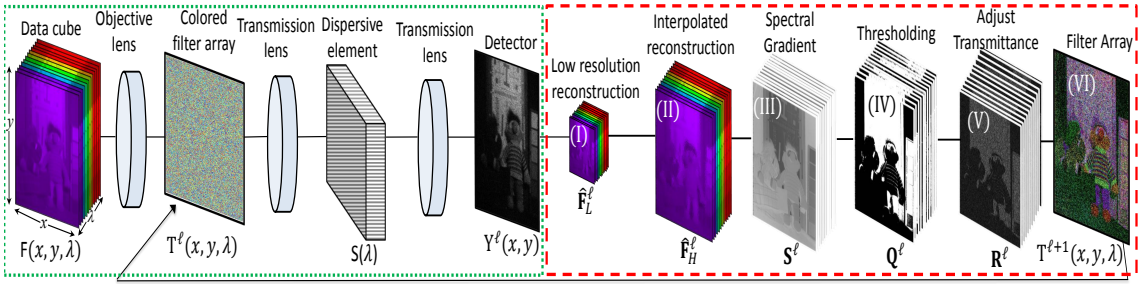


Figura 12 System layout of the adaptive C-CASSI. The scheme depicts two snapshots of C-CASSI. In the first snapshot, the colored coded aperture $\mathbf{T}^{\ell}$ *is generated randomly. In the second snapshot, the colored coded aperture* $T^{\ell+1}(x,y,\lambda)$ *is adaptively designed, using as prior information a low-resolution reconstruction* $\mathbf{F}_L^{\ell}$ (I)*, the reconstruction is interpolated* $\mathbf{F}_H^{\ell}$ (II)*, then the gradient of each spectral signatures of the interpolated data cube is computed* $\mathbf{S}^{\ell}$ (III)*, the gradient is thresholded* $\mathbf{Q}^{\ell}$ (IV)*, and the transmittance is adjusted* $T^{\ell+1}(x,y,\lambda)$ (V)*. This process repeats for the subsequent snapshots.*

3.2.2. Low-resolution reconstruction. The first step, denoted as (I) in Fig. 12, is the computation of a low resolution data cube from $\mathbf{y}^0$ and $\mathbf{H}^0$. This low-resolution estimation is required, since it is difficult to extract information about the scene directly from the compressed measurements, and a full resolution data cube reconstruction demands higher computational complexity. Instead of this step, some other works extract borders by employing side-information from a second sensor placed aside the sensing device (Galvis et al., 2017). Although, this reconstruction is not obtained in real time, this penalty is tolerable for the adaptive design of CCA. This step exploits the low-resolution data cube, which provides prior information about the underlying scene $\mathbf{f}$. Formally, let $\mathbf{f}_L^\ell = \mathbf{D}\mathbf{f}$, be the vectorized low-resolution image where $\mathbf{D}$ denotes the decimation matrix and $\mathbf{f}$ identifies the underlying scene. The low resolution version $\hat{\mathbf{f}}_L^\ell$ is obtained by solving,

$$\hat{\mathbf{f}}_L^\ell = \boldsymbol{\Psi}_L(\underset{\boldsymbol{\theta}_L}{\text{argmín}}\|\mathbf{y}-\mathbf{H}_L^\ell\boldsymbol{\Psi}_L\boldsymbol{\theta}_L\|_2^2+\tau\|\boldsymbol{\theta}_L\|_1+\lambda\|(\mathbf{C}-\mathbf{I})(\boldsymbol{\Psi}_L\boldsymbol{\theta}_L)\|_2^2), \tag{14}$$

where $\mathbf{H}_L^\ell = \mathbf{D}_H\mathbf{H}^\ell$ denotes the decimated sensing matrix, $\boldsymbol{\Psi}_L$ is the representation basis, and $\boldsymbol{\theta}_L$ is the vectorization of a sparse vector for the low resolution reconstruction, $\mathbf{I}$ is the identity matrix, $\mathbf{C}$ is a Gaussian filter to promote smoothness, and τ, λ are regularization constants.

3.2.3. Interpolation of reconstruction. The low resolution data cube $\hat{F}_L^\ell$ is then interpolated to obtain a high-resolution data cube $\hat{F}_H^0$, as shown in Fig. 12 (II). In particular, this step performs the interpolation of the low resolution data cube by using $\hat{\mathbf{f}}_H^\ell \leftarrow \mathbf{P}\hat{\mathbf{f}}_L^\ell$ where $\mathbf{P}$ is a bilinear interpolator attaining a high-resolution data cube $\hat{\mathbf{f}}_H^\ell$.

3.2.4. Computation of spectral gradient. In the following step, the spectral gradient $\mathbf{S}^\ell$ is computed, as depicted in Fig. 12 (III) or equivalently $\mathbf{s}^\ell$ which correspond to the vecto-

rization of $\mathbf{S}^{\ell}$. More precisely, $\mathbf{S}^{\ell}$ is the discrete second derivative of each spectral signature, also referred to as the matched filter. The spectral gradient computation of the high-resolution data cube is performed as $\mathbf{s}^{\ell} \leftarrow \mathbf{B}^2\hat{\mathbf{f}}_H^{\ell}$, where ℓ refers to the ℓ^{th} snapshot, $\mathbf{B}^2$ computes the second derivative of the spectral signature, with $\mathbf{B} \in \mathbb{R}^{N^2L \times N^2L}$ being the gradient matrix given by

$$\mathbf{B} = \mathbf{G} \otimes \mathbf{I}, \tag{15}$$

where the matrix $\mathbf{B}$ computes the central differences for all the spectral signatures in the datacube. Similarly, notice that $\mathbf{B}^2$ finds the second derivative of each spectral signature of the high resolution data cube by calculating $\mathbf{B}^2\hat{\mathbf{f}}_H^{\ell}$, $\mathbf{I} \in \mathbb{R}^{N^2 \times N^2}$ is an identity matrix, $\otimes$ is the Kronecker product, and $\mathbf{G} \in \mathbb{R}^{L \times L}$ computes the central differences in the internal data points and the single-side differences along the spectral signature. Each row of the matrix $\mathbf{G}$ computes the pixel-wise central difference of the spectral signature (Mathews and Fink, 1998). The matrix $\mathbf{G}$ is denoted as

$$\mathbf{G} = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -0.5 & 0 & 0.5 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -0.5 & 0 & 0.5 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ddots & -0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{pmatrix}. \tag{16}$$

3.2.5. Thresholding of the gradient. The next step is to compute the thresholding of the gradient to identify the maximum $\mathbf{Q}^\ell$ and minimum local spectral intensities regions $1-\mathbf{Q}^\ell$, where $\mathbf{Q}^\ell$ is a binary cube with ones in the maximum spectrum, and $1-\mathbf{Q}^\ell$ is a binary cube with ones in the minimum spectrum, i.e. the negative of $\mathbf{Q}^\ell$. The thresholding of the gradient is performed according to the logical operation $\mathbf{q}^\ell \leftarrow (\mathbf{s}^\ell \preceq \mathbf{0})$, where $\mathbf{q}^\ell$ is the vectorization of $\mathbf{Q}$, $\preceq$ is the element-wise inequality, $\mathbf{s}^\ell$ is the computed gradient, and $\mathbf{0}$ is an all zero vector. The thresholding $\mathbf{s}^\ell \preceq \mathbf{0}$ splits the spectral intensity profile in a maximum local spectral intensity region $\mathbf{q}^\ell$ and a minimum local spectral intensity region $\mathbf{1}-\mathbf{q}^\ell$, which corresponds to the complement of $\mathbf{q}^\ell$. This approach creates a matched filter across each spectral profile by sampling with different transmittances the high and low spectral intensities.

3.2.6. Adjust of transmittance. The transmittance is the percentage of light that hits the FPA, and is set according to the maximum and minimum spectral intensities, which denoted as $\mathbf{R}^\ell$ shown in Fig. 12 (V) sets a higher transmittance η_d in the local maximum intensity region $\mathbf{Q}^\ell$, and a lower transmittance η_u in the local minimum intensity region $\mathbf{1}-\mathbf{Q}^\ell$, where, $\eta_d \geq \eta_u$. A vectorization version of the adaptive colored coded aperture is computed according to $\mathbf{r}^\ell \leftarrow \mathbf{q}^\ell \odot \mathbf{r}_d^\ell + (\mathbf{1}-\mathbf{q}^\ell) \odot \mathbf{r}_u^\ell$ where $\mathbf{r}_d^\ell$ and $\mathbf{r}_u^\ell$ are vectors with binary entries and transmittance functions $\eta_d[\ell]$, and $\eta_u[\ell]$, respectively, dependent on the number of snapshots as

$$\mathbf{r}_d^\ell \sim \begin{cases} \text{Ber}(\eta_d[\ell]=0.5), & \text{if } \ell = 0 \\ \text{Ber}(\eta_d[\ell]=2/(\ell+3)), & \text{otherwise,} \end{cases} \tag{17}$$

$$\mathbf{r}_u^\ell \sim \begin{cases} \text{Ber}(\eta_u[\ell] = 0.5), & \text{if } \ell = 0 \\ \text{Ber}(\eta_u[\ell] = 1/(\ell+1)), & \text{otherwise.} \end{cases} \tag{18}$$

This step represents the adjustment of the transmittance, which emphasizes the intensities of the spectral signature $\mathbf{S}^\ell$, where Ber(.) denotes the Bernoulli distribution. The intuition behind these transmittance functions builds on a state-of-the-art transmittance criterion for colored coded apertures (Correa et al., 2016). In particular, $\mathbf{r}_u^\ell$ is a transmittance criterion for colored coded apertures for $\ell > 0$, defined as the inverse of the number of snapshots. This criterion reduces the transmittance of the colored coded aperture as the number of snapshots increases. In order to promote the high intensities spectral signature, two different transmittances are used in this algorithms, i.e. $\mathbf{r}_u^\ell \leq \mathbf{r}_d^\ell$. The values of these transmittance functions are close to the theoretical value found in (Correa et al., 2016), but in this work, they were found through experimentation. In particular, we set the transmittance $\mathbf{r}_d^\ell$ to be higher than $\mathbf{r}_u^\ell$ because we want to sample more the local maximum spectral intensity region $\mathbf{q}^\ell$. This in turn, permits to sample more the higher intensities to sharp the profiles of the spectral bands (Gonzalez and Woods, 2006). $\mathbf{R}^\ell$ then enhances the high intensities spectral signature of the reconstructed data cube by increasing the transmittance in the abrupt spectral intensity changes of the interpolated data cube. The transmittance parameters of $\mathbf{r}_d^\ell$, $\mathbf{r}_u^\ell$ were found after exhaustive search in the range between 0.1 to 0.9.

According to Eq. 17, for the first snapshot, the sampling vector $\mathbf{r}_d \sim \text{Ber}(\eta_d[\ell] = 0.5)$ has a transmittance function $\eta_d[\ell] = 0.5$ for the local maximum spectral intensity region $\mathbf{q}^\ell$. When the number of snapshots is $\ell > 0$, the sampling vector is given by $\mathbf{r}_d \sim \text{Ber}(\eta_d[\ell] = 2/(\ell+3))$ with

transmittance function denoted by $\eta_d[\ell] = 2/(\ell+3)$. Also, as stated in Eq. 18, for the first snapshot, the sampling vector $\mathbf{r}_u \sim \text{Ber}(\eta_u[\ell] = 0.5)$ has a transmittance $\eta_u[\ell] = 0.5$ for the local minimum spectral intensity $\mathbf{1} - \mathbf{q}^\ell$. When the number of snapshots is $\ell > 0$ the sampling vector is $\mathbf{r}_u \sim \text{Ber}(\eta_u[\ell] = 1/(\ell+1))$, whose transmittance function is $\eta_u[\ell] = 1/(\ell+1)$. The intuition behind the transmittance adjustment comes from Gonzalez and Woods, who classified the edge models according to intensity profiles (Gonzalez and Woods, 2006). In general, it is known that digital images are blurred due to the focusing mechanism, and are contaminated with noise because of the electronic components. The two most accurate edge models are the intensity ramp profile and the roof edge profile (Gonzalez and Woods, 2006). The transmittance adjustment is designed based on these edge models. Particularly, the proposed design increases the transmittance in some areas of $\mathbf{H}$ and decreases it in others. Let H_{mn} denote each entry of $\mathbf{H}$, if $\text{H}_{mn} = 1$ the light gets into the system, and, if $\text{H}_{mn} = 0$ the light is filtered out from reaching the FPA, where $m = 0, \ldots, K(N+L-1)-1$; $n = 0, \ldots, N^2L-1$.

3.2.7. Designed filter array. Figure 12 (VI) shows the computed consecutive filter array which is denoted by $T^{\ell+1}(x, y, \lambda)$. To attain the filter array it necessary to rearrange the vector version of $\mathbf{r}^\ell$ to its matrix version $\mathbf{H}^{\ell+1}$. In more detail, the vector $\mathbf{r}^\ell$ is rearranged as the 2D k^{th} plane according to $(\mathbf{r}_k^\ell)_l \leftarrow \mathbf{r}_n^\ell$. Afterwards, in step 13, the vectorization of the colored coded aperture $\mathbf{r}^\ell$ is stacked in the sensing matrix $\mathbf{H}^\ell$ according to $(H_m^{\ell+1})_n \leftarrow (r_{k_n}^{\ell_m})_{m-\ell_m v-k_n N}$, $k_n = \lfloor n/N^2 \rfloor$, $\ell_m = \lfloor m/V \rfloor$, $\ell_m \in \{0, \ldots, K-1\}$. In step 16, the adaptive snapshot is captured as $\mathbf{y}^{\ell+1} \leftarrow \mathbf{H}^{\ell+1}\mathbf{f}$. In step 19, the output the procedure is the approximation $\hat{\mathbf{f}}$. At each snapshot, the prior information is computed because the quality of the low-resolution data cube increases with

each additional snapshot.

Examples of the coding patterns designed by the proposed algorithm are depicted in Fig. 13. These coding patterns show the colored coded aperture along a set of axial slices of the 3D block-unblock representation. Specifically, Fig. 13 depicts a comparison between non-adaptive CCA, and adaptive CCA. In detail, the first row depicts the non-adaptive color coded aperture. The second, third, and fourth rows show the adaptive coded apertures for 3 different spectral scenes. The first column depicts the CCAs. The columns $2-4^{\text{th}}$ show 3 out of 12 axial slices of the block-unblock representation. The total number of axial slices is the same as the number of spectral bands, in this case $L = 12$. Notice that each colored coded aperture is highly correlated with the corresponding scene. The scenes are illustrated in Fig. 13. The adaptive filter arrays resemble the spectral components of the corresponding scene.

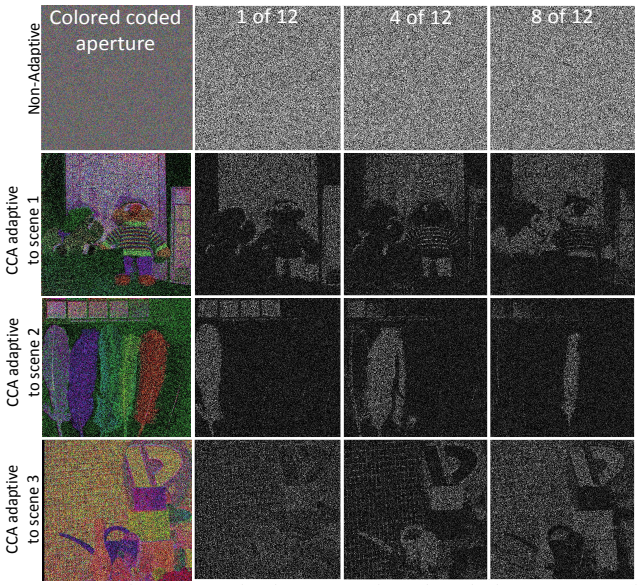


Figura 13 Comparison between non-adaptive and adaptive colored coded apertures when $K = 6$. The top row depicts non-adaptive random patterns. The second, third, and fourth rows show adaptively designed apertures, resulting of the use of the proposed algorithm, highly correlated with the spectral bands of each scene. (First column) CCA (second column) 3 slices of the 3D block-unblock representation.

3.2.8. Gradient thresholding algorithm (GTA). The algorithm 1 corresponds to the gradient thresholding algorithm (GTA), which summarize each of the six steps depicted in Fig. 12. The proposed adaptive scheme in Fig. 12 corresponds to the steps 3 to 7 and 13 of Algorithm 1. The inputs of the GTA algorithm are the initial compressive measurements $\mathbf{y}$ and the initial sensing matrix $\mathbf{H}^0$. GTA exploits a low-resolution reconstruction to provide prior information about the scene, which is shown in step 3. The line 4 of the GTA algorithm performs the interpolation of the low-resolution reconstruction. Step 5 obtains the gradient of each spectral signature. Line 6 computes the thresholding of the spectral gradient. Step 7 of GTA performs the adjust of transmittance. The subsequent sensing matrix which characterizes the coded aperture is attained in steps 13, and 15. The next snapshot is yielded in step 16. The output of the algorithm 1 is the subsequent filter array $T(x,y,\lambda)^{\ell+1}$ shown in Fig. 12 (VI),which is the approximation $\hat{\mathbf{f}}$.

3.3. Simulations

In order to assess the capabilities of the proposed method, it performs a comparison between the proposed adaptive C-CASSI and the non-adaptive C-CASSI. A critical parameter in the simulations is the transmittance which is traditionally set to 50 % in the non-adaptive C-CASSI. In adaptive C-CASSI, the transmittance sets according to two critical regions, local maximum $\mathbf{q}^\ell$ and local minimum $\mathbf{1}-\mathbf{q}^\ell$ spectral intensity regions. This process emphasizes the high intensities spectral signature $\mathbf{s}^\ell$, according to the intensity profile edge model (Gonzalez and Woods, 2006). The GTA algorithm 1 uses transmittance functions that reduce the transmittance of the colored coded apertures as the number of snapshots increases $\eta_d[\ell]$, $\eta_u[\ell]$, where $\eta_d[\ell] \geq \eta_u[\ell]$. Equations 17, and

Algorithm 1 GTA gradient thresholding algorithm.

Input: $\mathbf{y} \leftarrow [\mathbf{y}^0]^T$, $\mathbf{H} \leftarrow [\mathbf{H}^0]^T$
Output: $\hat{\mathbf{f}}$

1: **function** GTA($\mathbf{y}$, $\mathbf{H}$, $\mathbf{B}$)
2: **for** $\ell \leftarrow 0, K-1$ **do**
3: $\hat{\mathbf{f}}_L^\ell \leftarrow \boldsymbol{\Psi}_L(\text{argmín}_{\boldsymbol{\theta}_L}\|\mathbf{y}-\mathbf{H}_L\boldsymbol{\Psi}_L\boldsymbol{\theta}_L\|_2^2 + \quad \tau\|\boldsymbol{\theta}_L\|_1 + \lambda\|(\mathbf{C}-\mathbf{I})(\boldsymbol{\Psi}_L\boldsymbol{\theta}_L)\|_2^2)$ ▷ Low-resolution
4: $\hat{\mathbf{f}}_H^\ell \leftarrow \mathbf{P}\hat{\mathbf{f}}_L^\ell$ ▷ Interpolation
5: $\mathbf{s}^\ell \leftarrow \mathbf{B}^2\hat{\mathbf{f}}_H^\ell$ ▷ Compute gradient
6: $\mathbf{q}^\ell \leftarrow (\mathbf{s}^\ell \preceq \mathbf{0})$ ▷ Thresholding
7: $\mathbf{r}^\ell \leftarrow \mathbf{q}^\ell \odot \mathbf{r}_d^\ell + (\mathbf{1}-\mathbf{q}^\ell) \odot \mathbf{r}_u^\ell$, ▷ Transmittance
8: **for** $n \leftarrow 0, N^2L-1$ **do**
9: $k = \lfloor n/N^2 \rfloor$, $l = n \mod N^2$
10: $(r_k^\ell)_l \leftarrow r_n^\ell$ ▷ Rearrange $\mathbf{r}$
11: **for** $m \leftarrow 0, KV-1$ **do**
12: **if** $m - \ell_m V = n - k_n(N^2-N)$ **then**
13: $(H_m^{\ell+1})_n \leftarrow (r_{k_n}^{\ell_m})_{m-\ell_m v-k_n N}$ ▷ Compute $\mathbf{H}$
14: **else**
15: $(H_m^{\ell+1})_n \leftarrow 0$
16: $\mathbf{y}^{\ell+1} \leftarrow \mathbf{H}^{\ell+1}\mathbf{f}$ ▷ Next snapshot
17: $\mathbf{y} \leftarrow [(\mathbf{y}^0)^T, \ldots (\mathbf{y}^{K-1})^T]^T$
18: $\mathbf{H} \leftarrow [(\mathbf{H}^0)^T, \ldots (\mathbf{H}^{K-1})^T]^T$
19: **return** $\hat{\mathbf{f}} \leftarrow \boldsymbol{\Psi}(\text{argmín}_{\boldsymbol{\theta}}\|\mathbf{y}-\mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}\|_2^2 + \tau\|\boldsymbol{\theta}\|_1)$

18 determine the transmittance in the adaptive approach. The GTA algorithm 1 uses the Gradient Projection for Sparse Reconstruction (GPSR) algorithm (Figueiredo et al., 2007) to solve the step 3 and step 19. The type of GPSR used in the experimentation is the Barzilai-Borwein gradient projection GPSR-BB. For the reconstruction, the basis representation $\boldsymbol{\Psi}$ is set to be the Kronecker product of three basis $\boldsymbol{\Psi} = \boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_3$. The 2D spatial representation basis $\boldsymbol{\Psi}_{2D} = \boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_2$ is the 2D-Wavelet Symmlet 8 basis, and the spectral sparsifier is the 1-D cosine transform basis $\boldsymbol{\Psi}_3$. The input measurements $(\mathbf{y})$ in the GTA algorithm were acquired using Eq. 9, including a Gaussian noise with $SNR = 10$ dB.

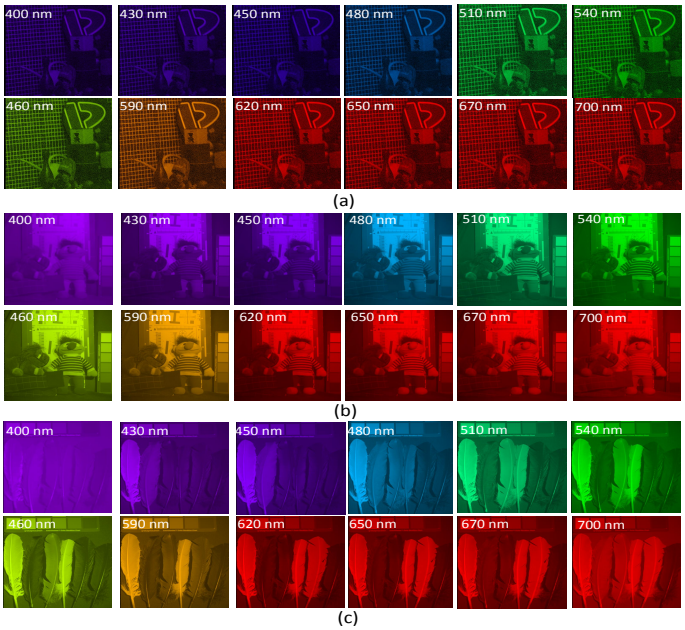


Figura 14 Three spectral scenes are used to test the proposed adaptive approach. These scenes have a resolution of $512 \times 512 \times 12$ *pixels, and they span a spectral range between 400 nm and 700 nm. Figure 14 (a), 14 (b), and 14(c) depict 12 spectral bands with the corresponding central wavelength of database of scene 5 from (Foster et al., 2006), chart and stuffed toys and feathers from (Yasuma et al., 2010).*

The GTA algorithm was evaluated with three scenes shown in Fig. 14. The first database is a portion of the scene 5 from (Foster et al., 2006), (Fig. 14a); the second database is the chart and stuffed toys from (Yasuma et al., 2010), (Fig. 14b); and the third database is the feathers also from (Yasuma et al., 2010), (Fig. 14c). Each spectral data cube has 12 bands that span between 400 nm to 700 nm, and the spatial resolution is 512×512 pixels. Figure 14(a), Figures 14(b), and 14(c) depict 12 spectral bands of scene 5 in (Foster et al., 2006), chart and stuffed toys, and feathers in (Yasuma et al., 2010). Each spectral band has its corresponding central wavelengths.

Figure 15 illustrates a comparison of the averaged PSNR against the number of snapshots for each database. Each subplot compares the quality of image reconstruction for the non-adaptive C-CASSI and the adaptive C-CASSI. Each column of the figure depicts experiments using the

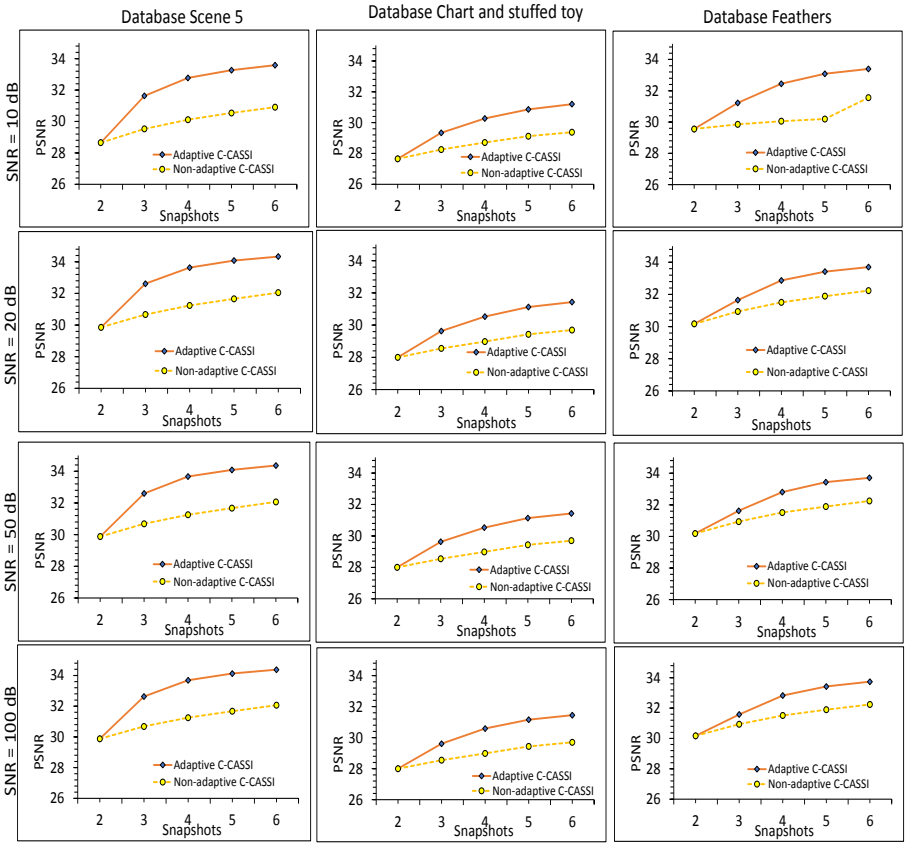


Figura 15 Averaged PSNR against the number of snapshots for each spectral scene (columns of the image), under four different SNR scenarios $10, 20, 50$ *and* 100 *dB in the compressive measurements (rows of the image).*

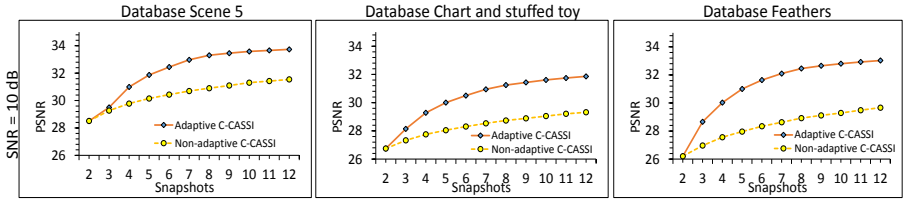


Figura 16 Averaged PSNR against the number of snapshots for each spectral scene with $L = 24$ *spectral bands and capturing* $K = 12$ *snapshots. The simulations include Gaussian noise with SNR* $= 10$ *dB in the compressive measurements.*

Scene 5, Chart and stuffed toy, and Feathers database, respectively. Each row refers to a certain SNR-scenario for each database; the corresponding SNR for each row is 10, 20, 50 and 100 dB, respectively. Note that when the number of snapshots is less than or equal to two, the transmittance

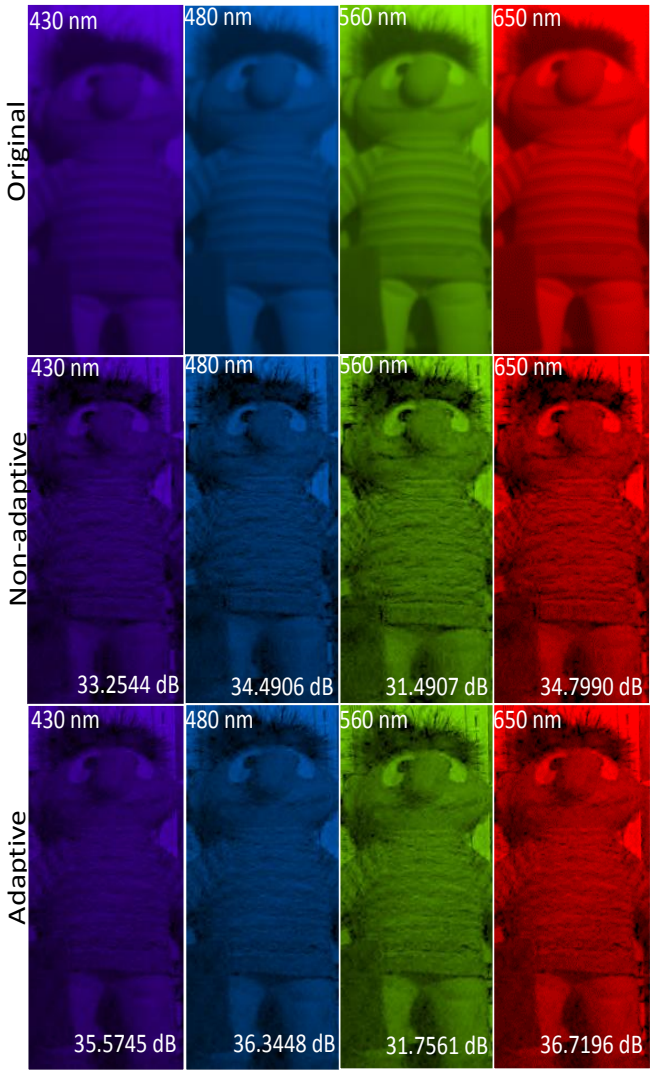


Figura 17 Zoomed reconstruction of a subset of spectral bands using $K = 6$ snapshots for non-adaptive C-CASSI with average PSNR across the spectral bands 33.5 dB with compression ration of 50 *%, and adaptive C-CASSI with average PSNR 35.6 dB and compression ration of* 80 *%. In addition, the PSNR for the specific band is indicated.*

is 50 % for both approaches, therefore at that snapshot level, there is not a significant difference. In contrast, after the second snapshot, the adaptive method overcomes the traditional approach. The simulations reveal that in the presence of Gaussian noise, the proposed adaptive approach outperforms the traditional approach in up to 2 dB. The results highlight the importance of adapting the subsequent colored apertures according to the previous information gathered in the compressive

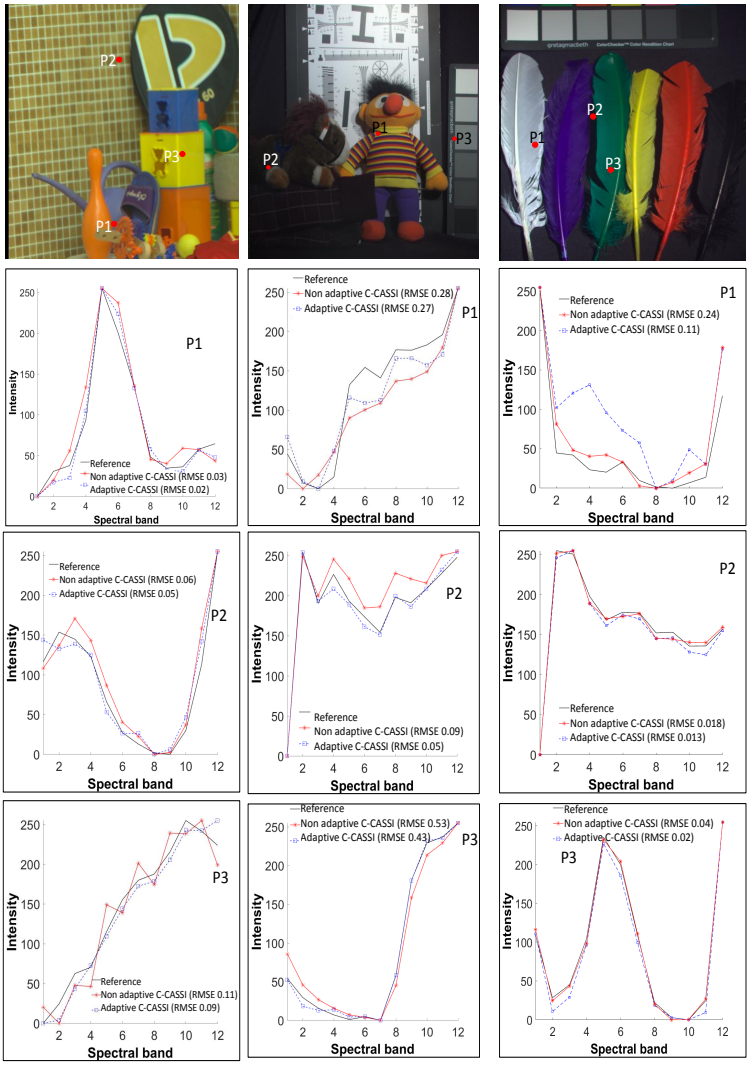


Figura 18 Spectral fidelity analysis at 3 points (P1, P2, P3) for each scene. The reference spectral signature and the reconstructed spectral signatures of the non-adaptive and adaptive C-CASSI are compared. Note that the spectral signatures of the proposed adaptive C-CASSI approach are closer to the reference spectral signature than the traditional non-adaptive C-CASSI.

measurements. A remarkable result from the simulations is that the non-adaptive C-CASSI is less robust to Gaussian noise than the proposed adaptive C-CASSI. In order to show the scalability of the adaptive C-CASSI, Fig. 16 shows the results of the quality of image reconstruction against the number of snapshots when the databases exhibit $L = 24$ spectral bands. The simulations show

that the adaptive C-CASSI overcomes the non-adaptive in up to 3 dBs. In order to compare the traditional method and the proposed adaptive method, the peak signal-to-noise ratio (PSNR) is used to measure the quality of image reconstruction. Figure 17 shows a zoomed version of the toy and the stuff scene. In the first row depicts the original scene. The second row shows the result of the non-adaptive C-CASSI, and the third row presents the result of the proposed adaptive method. Four out of the 12 wavebands shows with their corresponding central wavelength 430 nm, 480 nm, 560 nm and 650 nm along with the attained PSNR. The PSNR of the reconstructed bands with the non-adaptive approach is lower in comparison with the adaptive C-CASSI. The averaged PSNR across the 12 spectral bands is 33.5 dB for the non-adaptive C-CASSI and 35.6 dB for the adaptive C-CASSI. In order to test the fidelity of the spectral reconstructions, three reference spectral signatures denoted as $P1$, $P2$, and $P3$ are compared against the reconstructed spectral signatures of non-adaptive and adaptive C-CASSI in Fig. 18. The root means squared error between the reference and the reconstruction are included in each subfigure to improve the understanding of the results. These results reveal that there exists more similarity between the reference and the reconstructed spectral signatures using adaptive C-CASSI.

3.4. Real measurements

To experimentally prove the advantage of the adaptive design against the traditional system, the C-CASSI was assembled in the laboratory as depicted in Fig. 19, in accordance with (Rueda et al., 2015), using a DMD-based implementation of the colored coded apertures. The optical scheme depicts two arms, the imaging arm, and the integration arm. In the imaging arm; the imaging lens focuses the light into the DMD. Due to the DMD rotation angle of 45° relative to the detector

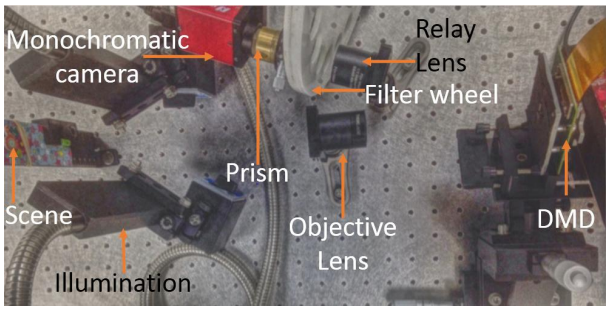


Figura 19 Optical scheme of the DMD filter-wheel based C-CASSI. In the imaging arm, the imaging lens focuses on the scene in the DMD. In the integration arm, the relay lens focuses the encoded light by the DMD, into the sensor, after the filter wheel filters it, and dispersed by the prism. The synchronization control coordinates the change of filters of the filter wheel, the DMD switching, and integration of the sensor.

pixels, the alignment of the optical elements is critical. In order to correct for the inclination of the DMD, the elements in the integration arm of the setup are rotated 45°, including the relay lens, the filter wheel, the prism, and the FPA. The apparatus is made up of (1) an AC254-100-A-ML objective lens (Thorlabs); (2) a DLI4130 DMD (DLInnovations) with spatial resolution of 1024×768 and mirror pitch size of 13.68 μm; (3) a MAP10100100-A relay lens (Thorlabs); (4) an Amici Prism (Shanghai Optics); (5) a monochrome charged-coupled device detector (AVT Stingray F-145B) with spatial resolution 1388×1038 and pitch size of $6.45 \mu m$. The emulation of the CCA using the DMD and the set of optical filters is as follows: each CCA mapped to a 3D array of block-unblock coded apertures paired with the corresponding set of optical filters, as described in (Rueda et al., 2015), thus emulating T_{ijk}^{ℓ} in Eq. 7. Each pair of the filter and block-unblock coded aperture is calibrated to characterize the impulse response of the system, which uses as input a monochromatic light, and a white plate as the target. The compressive measurements are obtained using as a target the real scene instead of the white plate and replacing the monochromatic light with a broadband white light. If each band is set to be encoded independently, a single shot

acquisition with L bands will require L switches in the DMD, and L rotations of the filter wheel. Experimentally, the DMD switching time is ~ 50 μs and the rotation time of the filter wheel is ~ 50 *ms*, being the latter the bottleneck. Therefore, the largest integration time for a single shot is bounded by $50L$ *ms*. Two target scenes were used to evaluate the quality of image reconstruction of the adaptive C-CASSI against non-adaptive C-CASSI. Figure 20 depicts the two targets. In this setup, the 3D block-unblock representation of CCA is loaded in the DMD, and each 2D slice is paired with the corresponding color filter placed in the filter wheel. Afterward, the measurements are obtained by adding 2D measurements. The experimental setup is synchronized, such that the DMD sets the pattern, and the sensor captures the projection. After that, the DMD updates the coded aperture, and the filter wheel rotates. Subsequently, the detector measures the next projection. The first target is the bear-stars scene (left), and the second target is the flower-stars scene (right). The spatial resolution of the target scenes is 128×128. The characterization of the Amici prism resolves $L = 11$ spectral bands. The corresponding wavelength intervals are: $\{423-436\}$; $\{437-448\}$; $\{449-463\}$; $\{464-479\}$; $\{480-499\}$; $\{500-521\}$; $\{522-546\}$; $\{547-577\}$; $\{578-618\}$; $\{619-673\}$; $\{674-700\}$ nanometers. Due to the nonlinear dispersion of the prism, the spectral bandwidths are different. In order to compare the reconstructed spectral signatures using the prism with the spectral signatures obtained with the commercial spectrometer, we select the values in the spectrometer readout corresponding to the central wavelength of the mentioned intervals. Figure 21 shows an RGB representation of the reconstructed data cubes of the bear-stars scene using $K = 4$ snapshots (first row) and the flower-stars scene (second row). Fig. 21 (a) and (c) shows the non-adaptive C-CASSI using random colored-coded aperture, and the proposed

Figura 20 Bear-stars scene (left). Flower-stars scene (right).

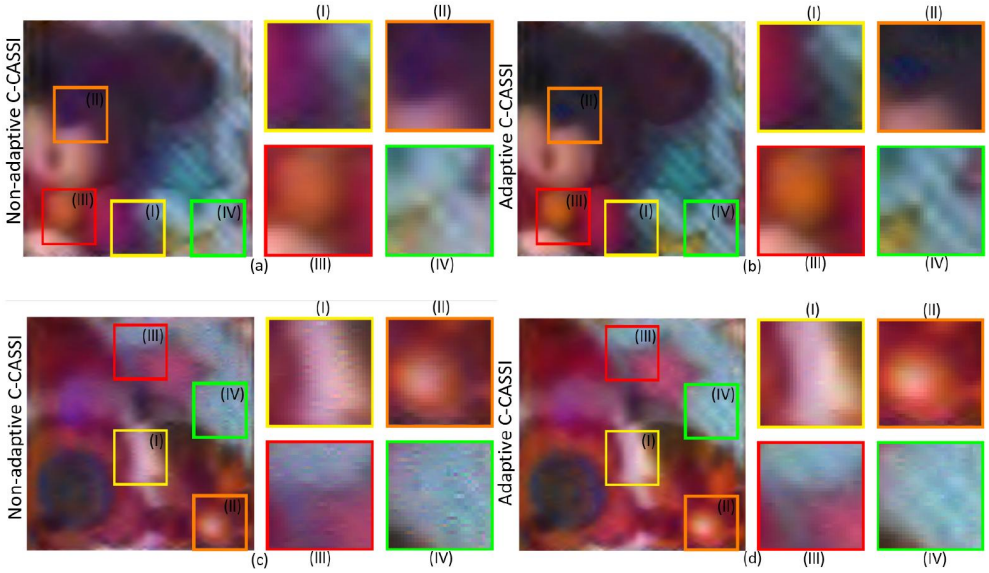


Figura 21 RGB reconstruction of the Bear-stars scene (first row), and the Flowers-start scene (second row), using $K = 4$ snapshots with the traditional C-CASSI ((a) and (c)) and the adaptive C-CASSI ((b) and (d)). Four zoomed sections (I-IV) are shown for each subfigure in order to easily appreciate the improvements.

adaptive C-CASSI in Fig. 21 (b) and (d). As can be observed, the proposed procedure improves the quality of image reconstruction. Four zoomed details of the scene depicts in the insets of each subfigure, in Fig 21 (a) and (b), inset (I) the right of the bow tie, (II)the bear's nose, (III) the center bow tie, and (IV) the yellow star. Regarding the second scene, Fig 21 (c), (d) includes (I) the white space between two flowers, (II) the center of a flower in the bottom-right corner, (III) the pink star, and (IV) the light-blue background. In these zoomed details of the scene, it is evident the improvement in the quality of image reconstruction of the proposed method. Notice that the diagonal

lines crossing the stars are more appreciable in the reconstruction with adaptive C-CASSI than in traditional C-CASSI due to the higher transmittance in the maximum local region in comparison with the minimum local region. Figure 22 represents a comparison of the reconstructed spectral bands of the bear scene using non-adaptive C-CASSI and adaptive C-CASSI with $K = 4$ snapshots for the spectral bands $\lambda_6 = 534$, $\lambda_7 = 562$, and $\lambda_8 = 597$ nm. In more detail, the first row depicts the spectral reconstruction using non-adaptive C-CASSI, and the second row shows the spectral reconstruction using adaptive C-CASSI. The same four zoomed details of the scene are depicts for each spectral band. In particular, notice for the proposed approach, the better spectral quality in diagonal lines of zoomed details (I) and (IV). To test the fidelity of the spectral reconstruction, Fig. 23 shows the spectral signatures for three spatial positions **P1**, **P2**, **P3** of the bear-stars scene for $K = 3$ snapshots, and the flower-star scene for $K = 3$ snapshots. For the scene in the top row, the point **P1** corresponds to the light-blue background, the **P2** represents the red bow tie, and **P3** denotes the yellow star. For the scene in the bottom row, the spatial points correspond to **P1** light-blue background, **P2** the dark-red petal of the flower in the top-left corner, and **P3** the orange petal of the flower in the bottom-right corner. In each spatial point, the light is measured using an Ocean Optics Flame spectrometer. The resulting spectral signature is compared with the spectral signature for the reconstructions using the non-adaptive C-CASSI and the adaptive C-CASSI. The three subplots show the improvement in the quality of image reconstruction of the proposed adaptive method in the bear-stars scene, and Flower-stars scene in comparison with the traditional random method.

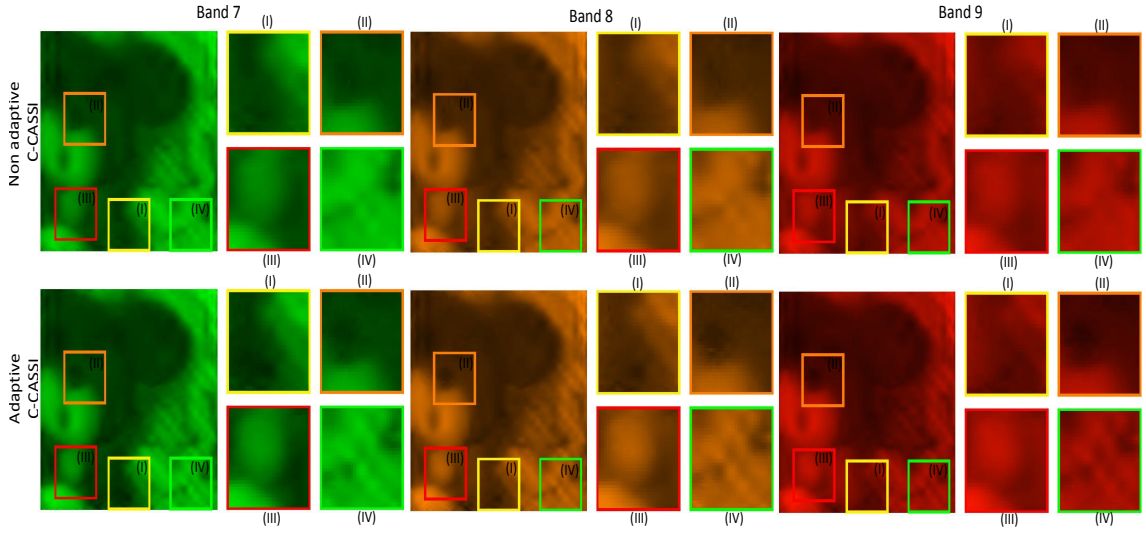


Figura 22 Reconstruction Bear-stars spectral scene at wavelength $\lambda_6 = 534$ nm, $\lambda_7 = 562$ nm, and $\lambda_8 = 597$ nm. In the right of each spectral band there are shown four zoomed sections (I-IV). The first row depicts the recovered spectral with traditional C-CASSI using $K = 4$ snapshots, whereas the second row shows the spectral reconstruction with adaptive C-CASSI.

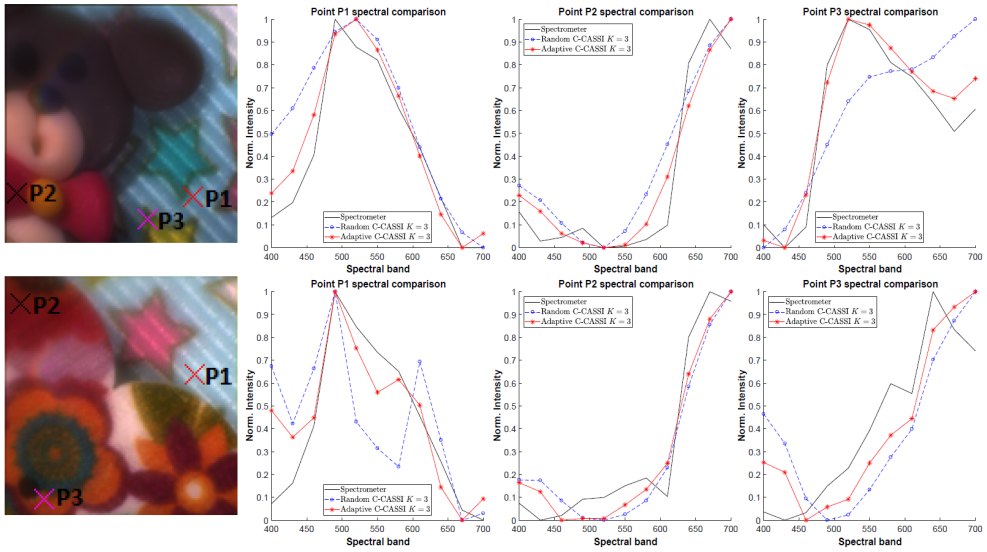


*Figura 23 Spectral signatures for three different spatial points in the target scenes, using $K = 3$. (Top row-column 1) three spatial points in the Bear-stars scene. (Top row-column 2) **P1** point in the light-blue background. (Top row-column 3) **P2** point in the red bow tie. (Top row-column 4) **P3** point in the yellow star. (Bottom row-column 1)selection fo the three spatial points in the Flower-star scene. (Bottom row-column 2) **P1** point in the light-blue background. (Bottom row-column 3) **P2** point in the dark red flower petal. (Bottom row-column 4) **P3** point in the orange flower petal.*

3.5. Conclusions

This chapter has shown the benefits of the adaptive design of colored coded apertures in C-CASSI. In particular, the code design of the CCA creates a matched filter of the scene adaptively. The resulting CCA adjusts to the scene. The simulations and experimental implementation show the feasibility of an adaptive C-CASSI system as an alternative to non-adaptive C-CASSI, with the advantage of improved quality in image reconstruction. The proposed architecture takes advantage of the previous captured compressive measurements and sensing matrix to compute the subsequently coded aperture relying on a low-resolution reconstruction. The key feature of adaptive C-CASSI is the subsequent design of coded apertures performing a matched filter and sampling better the high intensities of the spectral signatures of the low-resolution data cube. The results of this study imply that the designed coded apertures not only improve the quality of image reconstruction but better deal with Gaussian noise. The simulations and numerical analysis support the idea that the proposed approach outperforms the traditional method is up to 3 dB when the measurements contain Gaussian noise of SNR = 10 dB. Also, experimental reconstructions evidence the improvement of the proposed adaptive colored-coded aperture in comparison with the traditional non-adaptive colored random apertures.

4. Adaptive multisensor compressive acquisition for spectral image classification

4.1. Introduction

Spectral image (SI) classification is a relevant topic in remote sensing that aims at assigning predefined labels to the corresponding spectral pixels, allowing them to detect and characterize the different materials present in a scene. In particular, SI classification has many applications, for instance, forest classification (Goodenough et al., 2003), monitoring of agricultural land use (Amato et al., 2013), and precision agriculture (Yang et al., 2013). In the last two decades, outstanding pixel-wise approaches have been proposed to classify SI, including the methods based on support vector machines (SVMs) (Melgani and Bruzzone, 2004), neural networks (Ratle et al., 2010), dictionary-based sparse representation (Chen et al., 2011), and multinomial logistic regression (Li et al., 2012).

When the number of captured bands increases, the curse of dimensionality (Hughes, 1968) becomes a limitation, which degrades the classification performance, because the amount of training data is limited. The deficiency in the training samples induces low generalization in the pixel-wise classifier yielding to low classification accuracy. For that reason, various pixel-wise classifiers use feature extraction and dimensionality reduction methods, such as principal component analysis (PCA) (Ye et al., 2017), and independent component analysis (ICA) (Falco et al., 2014), to reduce the dimensionality related to the spectral signatures and thus, to improve the classification accuracy. Also, it is well-known that considering the spatial contextual information (SCI) improves the labeling accuracy, minimizing the salt and pepper noise in the classification maps (Li et al., 2012),

(Falco et al., 2015), (Fauvel et al., 2013) (Zhao et al., 2016). In spite of that, pixel-wise classifiers do not take into account the contextual information.

On the other hand, SI classifiers are applied to datasets acquired by traditional sensors which capture the entire image based on the Nyquist sampling theorem which leads to high costs in storage and transmission during the acquisition process. To avoid these, compressive spectral imaging (CSI) has emerged as an acquisition alternative that simultaneously senses and compresses the SI of interest. Specifically, this acquisition framework is based on the compressive sensing (CS) theory. Different architectures have been proposed to capture projections of an encoded-and-spectrally-shifted version of the input spectral density (Arguello and Arce, 2014), (Cao et al., 2016), (Lin et al., 2014b), and (Wagadarikar et al., 2008). Therefore, the classification might be performed in the compressive domain avoiding the disadvantages of conventional pixel-wise classifiers.

The intuitive approach to perform classification in the compressive domain would involve the reconstruction of the datacube, and use a conventional pixel-wise classifier. In 2014, Ramirez et al. proposed the first supervised approach that classifies spectral images in the compressive domain using sparse representation. The method uses compressive measurements of CASSI architecture to attain sparse representations of test pixels in a given dictionary (Ramirez et al., 2014). In this approach, the coded aperture is designed according to the restricted isometry property (RIP) (Samuel et al., 2016). Nevertheless, the method attains low accuracy classification results because the RIP-based design ignores the spatial contextual information, yielding poor features for the pixel-wise classifier. Moreover, the dictionary-based sparse representation is a computationally expensive method.

Another approach that performs classification in the compressive domain is shown in (Dunlop-Gray et al., 2016). It proposes an adaptive classification scheme from compressive measurements of the dual disperser CASSI (DD-CASSI) (Lin et al., 2014b) to classify the materials present in the scene avoiding the reconstruction of the entire datacube. This approach designs the coded apertures adaptively according to the adaptive Bayesian classifier and a feature design framework, which attains high accuracy classification results in a low SNR. However, the coded aperture design is Ad hoc. This classifications scheme requires comparison using the state-of-art dataset, and also the spatial context in the pixel-wise classifier is ignored.

Recently, a multisensor CSI architecture has been used to acquire hyperspectral and multispectral compressive measurements. This architecture requires the performing of fusion algorithms to reconstruct a high resolution multispectral and hyperspectral datacube (Vargas et al., 2019). The multisensor architecture has been used in classification by (Ramirez and Arguello, 2019a), (Ramirez and Arguello, 2019b). However, these approaches obtain both multispectral and hyperspectral compressive measurements with salt and pepper noise. Moreover, the approach is non-adaptive, avoiding the utilization of extracted spatial contextual information to prevent salt and pepper noise. It is known that adaptive approaches have been used in the past in CSI to prevent saturation (Diaz et al., 2015) (Diaz et al., 2019), and to improve the quality of reconstruction (Diaz et al., 2018). However, as far as we have known that there are not approaches performing adaptive CSI classification. The motivation of this adaptive approach is to use the spatial contextual information of a multiple sensor CSI architecture to adaptively design the coded apertures of two arms of a CSI architecture in order to improve the classification accuracy via removing the salt and pepper

noise of both types of measurements. This work presents two contributions. First, the spatial contextual information is included to design the coded aperture with the goal of removing remaining salt and pepper noise introduced by the compressive measurements in a multisensor CSI architecture, which also improves the extraction spatial classification features. The second contribution to adapt some approaches of the state-of-art in a multisensor scenario to compare with the proposed approach.

This chapter is organized as follows. Section 4.2 presents the observation model, which describes the optical model of dual-arm 3D-CASSI, one arm captures multispectral compressive measurements, and the other one samples hyperspectral compressive measurements. In this research, the CSI architectures used are the 3D coded aperture snapshot spectral imager (3D-CASSI) (Cao et al., 2016), and the colored coded aperture snapshot spectral imager (C-CASSI) (Arguello and Arce, 2014), both adapted in a multisensor architecture. Section 4.3 describes the proposed adaptive CSI classification algorithm. Section 4.4 outlines the results using the adaptive approach using two hyperspectral datasets. Conclusions are summarized in section 4.5.

4.2. Observation model

The multiple sensor architecture used in this paper is based on two CSI systems 3D-CASSI (Cao et al., 2016) and C-CASSI (Arguello and Arce, 2014). This section presents the mathematical models of both CSI architectures. These architectures have a multisensor with a high spatial resolution arm and a high spectral resolution arm.

4.2.1. Hyperspectral image acquisition model. The hyperspectral compressive measurements are given by $\mathbf{y}_h^k = \mathbf{H}_h^k \mathbf{D}_h \mathbf{f} + \boldsymbol{\omega}_h$, being $\mathbf{H}_h^k \in \mathbb{R}^{\frac{MN}{p^2} \times \frac{MNL}{p^2}}$ the sensing matrix at the k^{th}

snapshot, $\mathbf{f} \in \mathbb{R}^{MNL}$ a vectorized version of the datacube, the matrix $\mathbf{f}_h = \mathbf{D}_h\mathbf{f}$ being $\mathbf{f}_h \in \mathbb{R}^{\frac{MNL}{p^2}}$ the hyperspectral image, and $\boldsymbol{\omega}_h \in \mathbb{R}^{\frac{MN}{p^2}}$ is the additive noise. A detailed description of $\mathbf{D}_h$ is presented below:

- The matrix $\mathbf{D}_h \in \mathbb{R}^{\frac{MNL}{p^2} \times MNL}$ is the spatial decimation matrix, given by $\mathbf{D}_h = \mathbf{I}_h \otimes \mathbf{S}_h$ where $\mathbf{I}_h \in \mathbf{R}^{L \times L}$ is an identity matrix, $\otimes$ is the Kronecker product, and $\mathbf{S}_h \in \mathbb{R}^{\frac{MN}{p^2} \times MN}$, it is a matrix that degrades spatially each band, it is given by $\mathbf{S}_h = \mathbf{D}_c\mathbf{D}_r$.

- The matrix $\mathbf{D}_c \in \mathbb{R}^{\frac{MN}{p^2} \times \frac{MN}{p}}$ decimates the columns of each spectral band. In more detail, $\mathbf{D}_c = \mathbf{I}_c \otimes \mathbf{G}^T$ where $\mathbf{I}_c \in \mathbb{R}^{\frac{M}{p} \times \frac{M}{p}}$ is an identity matrix. The matrix $\mathbf{G}^T \in \mathbb{R}^{\frac{N}{p} \times N}$ is given by $G_l = \boldsymbol{\Theta}_c^{(l)}\mathbf{d}_c$ where $l = \{0, 1 \cdots, \frac{N}{p} - 1\}$ corresponds to the index of the columns of matrix $\mathbf{G}$, $\mathbf{d}_c \in \mathbb{R}^N$ is a vector that selects of columns, the position of the ones are determined by $\frac{iN}{p}$ with $i = \{0, 1 \cdots, p-1\}$ and $\boldsymbol{\Theta}_c^{(l)} \in \mathbb{R}^{N \times N}$ is a permutation matrix is given by,

$$\boldsymbol{\Theta}_c = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{19}$$

- The matrix $\mathbf{D}_r \in \mathbb{R}^{\frac{MN}{p} \times MN}$ decimates the rows of each spectral band, which is given by $\mathbf{D}_r = \mathbf{I}_r \otimes \mathbf{d}_r$ where $\mathbf{I}_r \in \mathbb{R}^{\frac{MN}{p} \times \frac{MN}{p}}$ stands for an identity matrix, p corresponds to the decimation

ratio, and $\mathbf{d}_r = \mathbf{1}^T$ is a vector of all ones such as $\mathbf{d}_r \in \mathbb{R}^p$.

4.2.2. Multispectral image acquisition model. The multispectral compressive measurements are given by $\mathbf{y}_m^w = \mathbf{H}_m^w \mathbf{D}_m \mathbf{f} + \boldsymbol{\omega}_m$, where $\mathbf{y}_m^w \in \mathbb{R}^{MN}$, and $\mathbf{H}_m^w \in \mathbb{R}^{MN \times \frac{MNL}{q}}$ is the sensing matrix at w^{th} each snapshot, q is the ratio of the spectral decimation, where $\mathbf{D}_m \in \mathbb{R}^{\frac{MNL}{q} \times MNL}$ is the decimation matrix of the spectral bands, given by $\mathbf{D}_m = \mathbf{D}_\lambda \otimes \mathbf{I}_m$, being $\mathbf{D}_\lambda \in \mathbf{R}^{\frac{L}{q} \times L}$ the decimation matrix that summarizes the bands of the original datacube, $\otimes$ denotes the Kronecker product, $\mathbf{I}_m \in \mathbb{R}^{MN \times MN}$ is an identity matrix, $\mathbf{f}$ is datacube, the matrix $\mathbf{f}_m = \mathbf{D}_m \mathbf{f}$ denotes to the multispectral image, $\boldsymbol{\omega}_m \in \mathbb{R}^{MN}$ corresponds to the additive noise.

4.3. Design of adaptive acquisition

In order to perform adaptive classification using features extracted from compressive measurements, the algorithm 2 is proposed. The CSI architecture is a multisensor architecture of two arms. The first arm measures compressive measurements $\mathbf{y}_h^k$ of low spatial and high spectral resolution (hyperspectral compressive measurements). The second arm samples compressive measurements $\mathbf{y}_m^w$ of low spectral and high spatial resolution (multispectral compressive measurements), where the index of multispectral shots is between $w = \{0, \ldots W-1\}$, and the index of hyperspectral shots is between $k = \{0, \ldots K-1\}$ notice that $L < K$ which means that the number of multispectral snapshots is smaller than the number of hyperspectral snapshots. Figure 24 depicts a scheme of the adaptive classification approach using compressive measurements of multiple sensors. The procedure is composed of nine stages. (1) Acquisition of grayscale image, (2) low-pass filter, (3) quantization, (4) design of sampling matrix, (5) Acquisition of compressive measurements, (6) denoising, (7) interpolation, (8) feature extraction, and (9) classification.

4.3.1. Acquisition of grayscale image. In the first stage, in the first snapshot, the filters of the colored coded aperture in the multispectral arm (or sensing matrix) are set to one for the 3D-CASSI, and the filters are set to one in specific contiguous bands that are complementary along with the snapshots for the C-CASSI, the purpose of this step is to attain the first multispectral compressive measurement $\mathbf{Y}^0$ or equivalently a grayscale image in the first snapshot.

4.3.2. Low-pass filter. In the second stage, a high-pass filtering process obtains the components of the high frequency of the grayscale image $\overline{\mathbf{Y}}^0$, i.e., detecting the edge of the grayscale image.

4.3.3. Quantization. In the third stage, the filtered image is quantized according to the number of classes in the scene. The number of classes can be assumed because this approach is supervised, resulting in the matrix $\mathbf{Q}$, which provides the prior spatial contextual information to the algorithm.

4.3.4. Design of sampling matrix. In the following stage, the design of the sampling matrix is computed using a matched filter. The matched filter is computed to attain the best match between the training spectral signatures and designed filters. Then, the filters are sorted in descending order to sample, first, the filters with the best match, and so on. The sensing matrices of the two arms are designed $\mathbf{H}_h^k$, and $\mathbf{H}_m^w$, or equivalently the colored coded aperture $\mathbf{T}_m^w$, and $\mathbf{T}_h^k$. The designed patterns are spectrally complementary along with the snapshots. The spatial distribution of the colored coded apertures follows the spatial distribution of the quantized and filtered grayscale image $\mathbf{Q}$. Notice that the coded apertures $\mathbf{T}_h^k$, and $\mathbf{T}_m^w$ are adaptively computed rearranging the

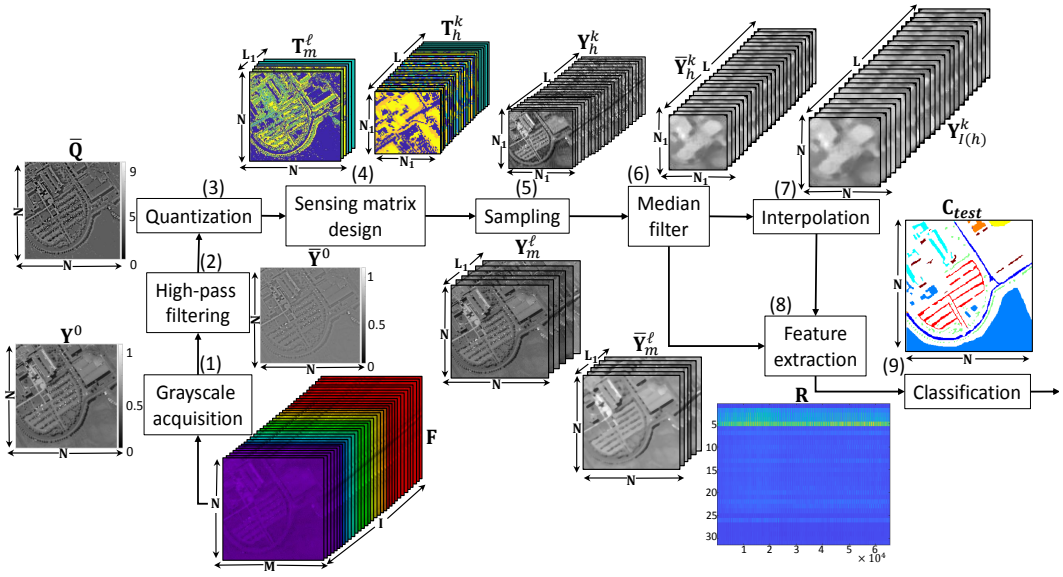


Figura 24 Adaptive classification scheme via designing of the coded apertures using spatial contextual information provided by the grayscale image of the scene. The multispectral and hyperspectral coded apertures $\mathbf{T}_m^w$ *and* $\mathbf{T}_h^k$ *or equivalently sensing matrix* $\mathbf{H}_m^w$*, and* $\mathbf{H}_h^k$ *in the multisensor 3D-CASSI are designed adaptively.*

coded apertures, which yields to the sensing matrices $\mathbf{H}_h^k$, and $\mathbf{H}_m^w$.

4.3.5. Acquisition of compressive measurements. In the fifth stage, each arm captures a different type of compressive measurement. At each snapshot, the multispectral compressive measurements $\mathbf{y}_h^k = \mathbf{H}_h^k \mathbf{D}_h \mathbf{f} + \boldsymbol{\omega}$ are captured, and the hyperspectral compressive measurements $\mathbf{y}_m^w = \mathbf{H}_m^w \mathbf{D}_m \mathbf{f} + \boldsymbol{\omega}$ are captured. The matrices $\mathbf{D}_m$, and $\mathbf{D}_h$ correspond to the spatial, and spectral decimation matrices allowing to sample measurements with different spatial, and spectral resolution. Figure 25 depicts an example of the compressive measurements of the multispectral and hyperspectral arms. Notice that when the coded apertures are spatially and spectrally complementary but randomly distributed, a salt and pepper noise is produced over the compressive measurements (see Fig. 25(a), and 25(b)). In contrast, when the coded apertures are designed the salt pepper noise is prevented (see Fig. 25(c), and 25(d)). However, in order to include the spatial contextual informa-

tion in the extracted features, a pre-processing stage is required.

4.3.6. Feature extraction. In the pre-processing stage, a median filter is applied to obtain the spatial contextual information from the compressive measurements. A median filter with window $[7 \times 7]$ is performed to promote the uniformity in both compressive measurements resulting in $\overline{\mathbf{Y}}_m^k$, and $\overline{\mathbf{Y}}_h^w$. In addition, the filtered hyperspectral measurements $\overline{\mathbf{Y}}_h^k$ are interpolated to attain the size of the multispectral measurements, resulting in $\overline{\mathbf{Y}}_{I(hs)}^k$. Using the filtered compressive multispectral measurements $\overline{\mathbf{Y}}_m^w$, and the filtered-interpolated compressive hyperspectral measurements $\overline{\mathbf{Y}}_{I(hs)}^k$, the features are extracted resulting in the matrix $\mathbf{R}$, this is depicted in Fig. 24. Figures 25(e) and 25(f) depict the spatial and spectral features after applying the median filter to the non-adaptive compressive measurements. In contrast, figure 25(g), and 25(h) shows the spatial $\overline{\mathbf{Y}}_m^w$, and spectral $\overline{\mathbf{Y}}_{I(hs)}^k$ features after applying the median filter to the adaptive compressive measurements. Notice that compressive measurements using the non-adaptive approach attain lower spatial uniformity Fig. 25(e) and Fig. 25(f) than the compressive measurements obtained using the proposed adaptive designed of the coded apertures Fig. 25(g) and Fig. 25(h).

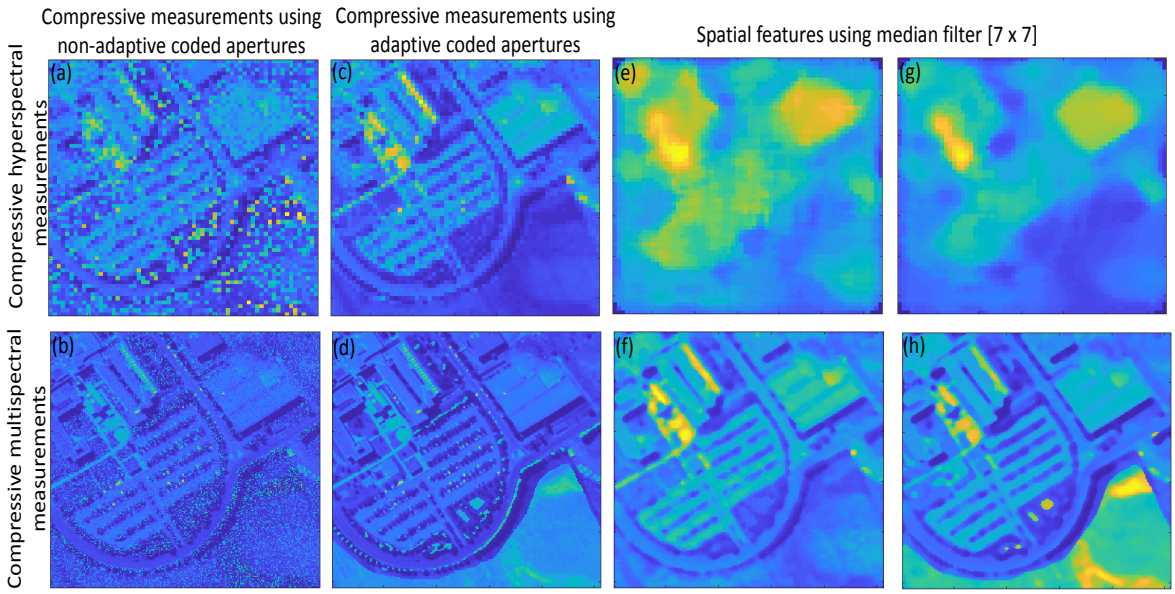


Figura 25 Denoising of compressive measurements. Multispectral compressive measurements using non-adaptive coded apertures (a), hyperspectral compressive measurements using non-adaptive coded apertures (b), Multispectral compressive measurements using adaptive coded apertures (c), hyperspectral compressive measurements using adaptive coded apertures (d). Compressive measurements after median filter with window size $[7 \times 7]$*, using non-adaptive coded apertures (e), and (f), using adaptive coded apertures (g) and (h).*

4.3.7. Classification. The classification is performed using as input the extracted features of the previous step. The support vector machine (SVM) classifier (Melgani and Bruzzone, 2004), computes the class labels for training $\mathbf{C}_{training}$, and the class labels for test $\mathbf{C}_{test}$ to each spatial position of the scene $\mathbf{f}$.

4.3.8. Algorithm. The following algorithm summarizes the proposed method. The algorithm 2 have two inputs. First, the vectorization of the datacube denoted by $\mathbf{f}$ which is used only to compute the compressive measurements of both arms in the CSI multisensor. Second, the colored coded aperture $\mathbf{T}_m^0$. The acquisition of grayscale image is computed in step 4. Then, the low-pass filter is calculated in step 7. After that, the quantization is computed in step 8. And then, it

is designed the sampling matrix, this step is performed in lines 11 and 18. Subsequently, the compressive measurements are acquired in matrix form $\mathbf{Y}_m^w$, $\mathbf{Y}_h^k$, in line 13, and line 20. The processing before attains the features is composed of denoising, and the interpolation. The denoising is conducted in lines 14, and 21. And, the interpolation is only performed to hyperspectral measurements in line 15. And the extracted features denoted as matrix $\mathbf{R}$ are computed in line 22. After, the classification is computed in lines 23, and 24. The output of the algorithm is the resulting classification labels $\mathbf{C}_{test}$.

4.4. Simulations and results

The spatial resolution of the datacube used is $M = 256$ and $N = 256$, the spatial decimation $p = 4$, and $q = 4$, the number of spectral bands $W = 24$, and $K = 96$.

Figure 26 shows a comparison of the non-adaptive 3D-CASSI and adaptive 3D-CASSI. The test varies the index of multispectral and hyperspectral snapshots $w = \{2,3,4,6\}$, and $k = \{8,12,16,24\}$, respectively. It is clear that the non-adaptive approach attains low classification values in the AO metric 93.32 %, 94.89 %, 95.05 %, and 96.26 % in comparison with the results obtain using the adaptive approach 99.25 %, 99.49 %, 99.64 %, and 99.68 %. The proposed approach attains 99 % of accuracy in the classification using 2 multispectral shots, and 8 hyperspectral shots, in contrast with the non-adaptive approach which obtains 93 % with the same number of snapshots. Notice, that there is a significant reduction in the number of shots using the proposed approach.

A test using a CSI architecture with dispersion is more challenging because the contextual information is blurred in compressive measurements. Fig. 27 depicts the results using C-CASSI. It depicts a comparison of the non-adaptive C-CASSI and adaptive C-CASSI. The test varies the in-

Algorithm 2 Adaptive classification using spatial contextual information in a multisensor CSI.

Input: $\mathbf{T}_m^0$. ▷ Inputs of the algorithm.
Output: $\mathbf{C}_{test}$. ▷ Output of the algorithm.
function AC($\mathbf{f}$, $\mathbf{T}_m^0 = \mathbf{1}$)
  $\mathbf{T}_m \leftarrow [(\mathbf{T}_m^0)^T]$ ▷ Concatenation of multispectral codes.
  $\mathbf{H}_m^0 \leftarrow$ **Rearrange**$(\mathbf{T}_m)$ ▷ Rearrange of patterns to multispectral sensing matrix.
  $\mathbf{y}_m^0 \leftarrow \mathbf{H}_m^0 \mathbf{D}_m \mathbf{f} + \boldsymbol{\omega}$ ▷ Compute a grayscale image of the scene.
  $\mathbf{y}_m \leftarrow [(\mathbf{y}_m^0)^T]$ ▷ Concatenation of grayscale.
  $\mathbf{Y}_m \leftarrow$ **Rearrange**$(\mathbf{y}_m)$ ▷ Concatenation of multispectral measurements.
  $\overline{\mathbf{Y}}_m \leftarrow$ **High-pass filter**$(\mathbf{Y}_m)$ ▷ High-pass filtering of grayscale image.
  $\mathbf{Q} \leftarrow$ **Quantization**$(\overline{\mathbf{Y}}_m)$ ▷ Quantization of grayscale image.
  for $k \leftarrow 0, K-1$ **do**
    $\mathbf{T}_h^k \leftarrow$ **Adaptive coding**$(\mathbf{Q}, \mathbf{T}_h)$ ▷ Adaptive hiperspectral coding design.
    $\mathbf{H}_h^k \leftarrow$ **Rearrange**$(\mathbf{T}_h^k)$ ▷ Concatenation of hiperspectral measurements.
    $\mathbf{y}_h^k \leftarrow \mathbf{H}_h^k \mathbf{D}_h \mathbf{f} + \boldsymbol{\omega}$ ▷ Compute CH measurements.
    $\mathbf{Y}_h^k \leftarrow$ **Rearrange**$(\mathbf{y}_h^k)$ ▷ Rearrange of hyperspectral measurements.
    $\overline{\mathbf{Y}}_h^k \leftarrow$ **Median filter**$(\mathbf{Y}_h^k)$ ▷ Compute median filtering.
    $\overline{\mathbf{Y}}_{I(h)}^k \leftarrow$ **Interpolation**$(\overline{\mathbf{Y}}_h^k)$ ▷ Interpolate hyperspectral measurements.
  for $w \leftarrow 1, W-1$ **do**
    $\mathbf{T}_m^w \leftarrow$ **Adaptive coding**$(\mathbf{Q}, \mathbf{T}_m)$ ▷ Adaptive multispectral coding design.
    $\mathbf{H}_m^w \leftarrow$ **Rearrange**$(\mathbf{T}_m^w)$ ▷ Concatenation of multispectral sensing matrix.
    $\mathbf{y}_m^w \leftarrow \mathbf{H}_m^w \mathbf{D}_m \mathbf{f} + \boldsymbol{\omega}$ ▷ Compute multispectral compressive measurements.
    $\mathbf{Y}_m^w \leftarrow$ **Rearrange**$(\mathbf{y}_m^w)$ ▷ Rearrange of multispectral measurements.
    $\overline{\mathbf{Y}}_m^w \leftarrow$ **Median filter**$(\mathbf{Y}_m^w)$ ▷ Compute median filtering.
  $\mathbf{R} \leftarrow [vec(\overline{\mathbf{Y}}_m^1), \ldots, vec(\overline{\mathbf{Y}}_m^w), vec(\overline{\mathbf{Y}}_{I(h)}^1), \ldots, vec(\overline{\mathbf{Y}}_{I(h)}^k)]^T$ ▷ Spatio-spectral feature extraction.
  $\mathbf{C}_{\mathbf{training}} \leftarrow \mathbf{SVM}_{training}(\mathbf{R})$ ▷ Classification labels using training data.
  $\mathbf{C}_{\mathbf{test}} \leftarrow \mathbf{SVM}_{test}(\mathbf{R})$ ▷ Classification labels using test data.
  return $\mathbf{C}_{test}$ ▷ Output of the algorithm.

dex of multispectral and hyperspectral in the same manner that for 3D-CASSI. Notice that the non-adaptive approach achieves low classification values in the AO metric 90.93 %, 93.52 %, 94.49 %, and 96.0 % against the results obtain utilizing the adaptive approach 95.47 %, 99.54 %, 97.07 %, and 97.66 %. There is a reduction in the number of snapshots. The non-adaptive approach attains

an accuracy of 90.93 % using 2 and 8 multispectral, and hyperspectral snapshots, respectively. In contrast, using the same number of snapshots, the adaptive approach obtains an accuracy of 95.47 %.

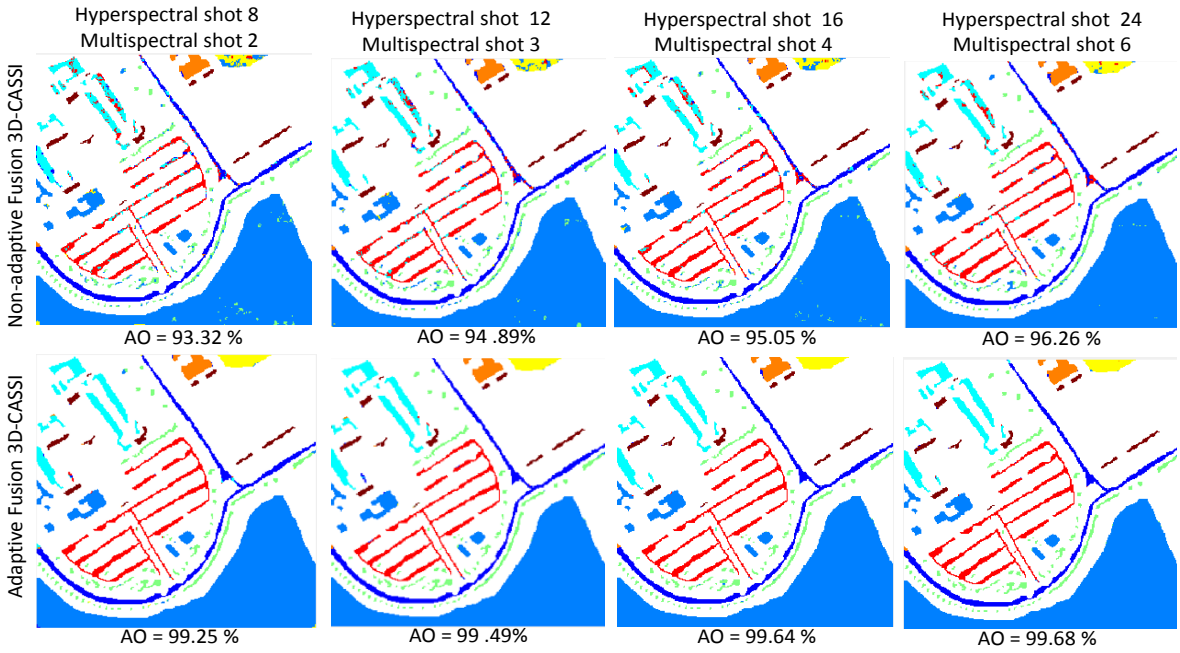


Figura 26 Classification using non-adaptive coded aperture 3D-CASSI against adaptive coded aperture 3D-CASSI.

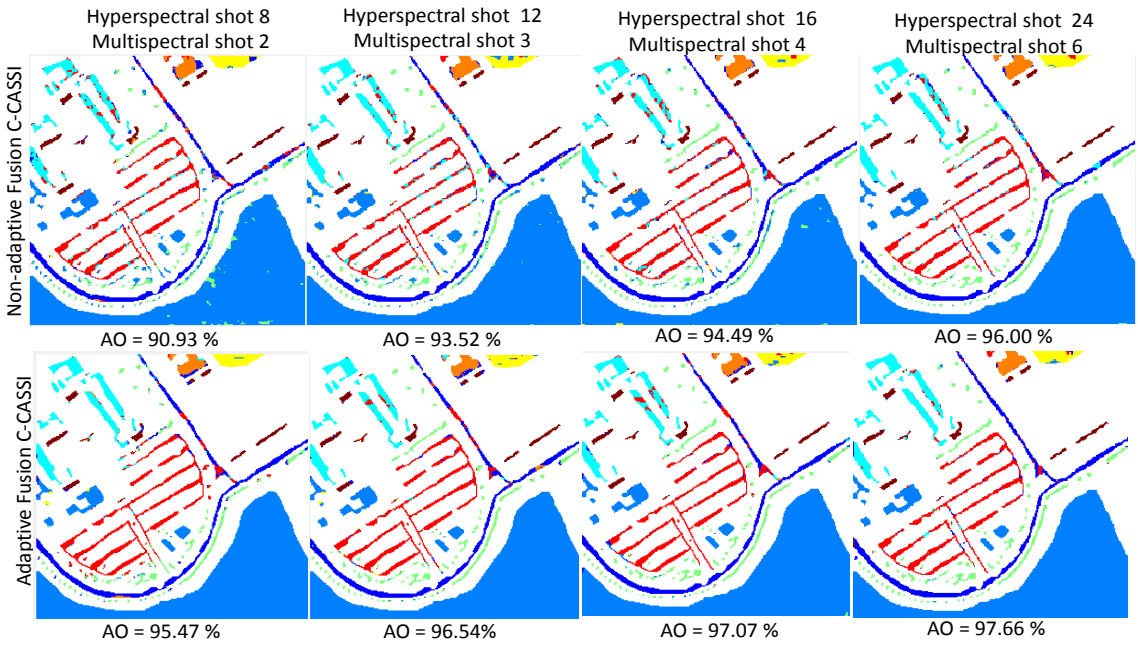


Figura 27 Classification using non-adaptive coded aperture C-CASSI against adaptive coded aperture C-CASSI.

4.5. Conclusions

The adaptive 3D-CASSI was proposed to design colored-coded apertures of the dual-arm CSI adaptively. According to the simulation results, a significant reduction in the number of snapshots in the hyperspectral, and multispectral arm is achieved using the proposed approach. Besides, the simulation test shows remarkable improvement in the accuracy classification in up to 99.68 % in the OA metric.

5. Cardiac motion estimation and convolutional sparse coding

5.1. Introduction

Chapter 5 is part of this dissertation because the motion estimation method proposed here is used in the chapter 6 to estimate the motion in spectral images, and then design the coded aperture for a video compressive spectral imaging architecture (V-CSI). Chapter 6 will show that despite the motion estimation is developed for Ultrasound Imaging (UI), it might be extended to other areas such as spectral imaging. In chapter 5 the kind of images used are ultrasound images. UI is a high temporal resolution imaging modality used in many clinical applications such as echocardiography due to its low cost, non-ionization characteristics, and comfort for the patient. An active research area in UI is tissue motion estimation. In particular, the problem of cardiac motion estimation in UI has been addressed with different approaches based on block-matching (Kaluzynski et al., 2001), the monogenic signal (Alessandrini et al., 2013), and B-splines (Myronenko et al., 2009). More recently, Ouzir et al. have introduced in (Ouzir et al., 2018) an energy minimization problem to estimate the motion of the heart. The proposed energy was a linear combination of a data fidelity term, a smoothness term and a regularization constructed from a sparse decomposition in a dictionary of motions determined using a standard patch-based dictionary learning method. The resulting motion estimation algorithm showed better results than classical methods such as block-matching, the monogenic signal, and B-splines.

This chapter investigates a new motion estimation method based on convolutional dictionary learning (CDL). The motivation of this study is to estimate the heart motion by using conse-

cutive frames of ultrasound images as inputs to build a convolutional dictionary (Wohlberg, 2016). Particularly, the k^{th} cardiac motion image $\mathbf{s}_k$ is modeled as a convolution between the coefficient maps $\mathbf{x}_{m,k}$ and a set of M filters $\mathbf{d}_m$. The coefficient maps indicate where the filters are activated, and the filters are supposed to model specific structures contained in the images of interest. A particular example is displayed in Fig. 28, where Fig. 28(a) displays one frame of the heart motion, Fig. 28(b) shows the estimated filters for the image and Fig. 28(c) shows the map of the cardiac motions associated with the red patch of the image. Note that the convolutional dictionary of Fig. 28(b) was obtained using $M = 32$ filters of size $L \times L$ with $L = 8$. In Fig. 28(c), the cardiac motions of the red patch are written as the linear combination of 10 filters convolved with a respective set of coefficient maps. Note that only 10 filters are required to represent this patch and that the 22 remaining filters are inactive, i.e., with zero coefficients. In order to improve the visualization quality, only the non-zero values of the coefficient maps have been shown. The key advantage of using a convolutional sparse model is its translation-invariant property which may offer a better representation in comparison with standard dictionary learning strategies. Indeed, each patch of the image can be sparsely represented with the proposed model by a single shift-invariant local dictionary (Papyan et al., 2017).

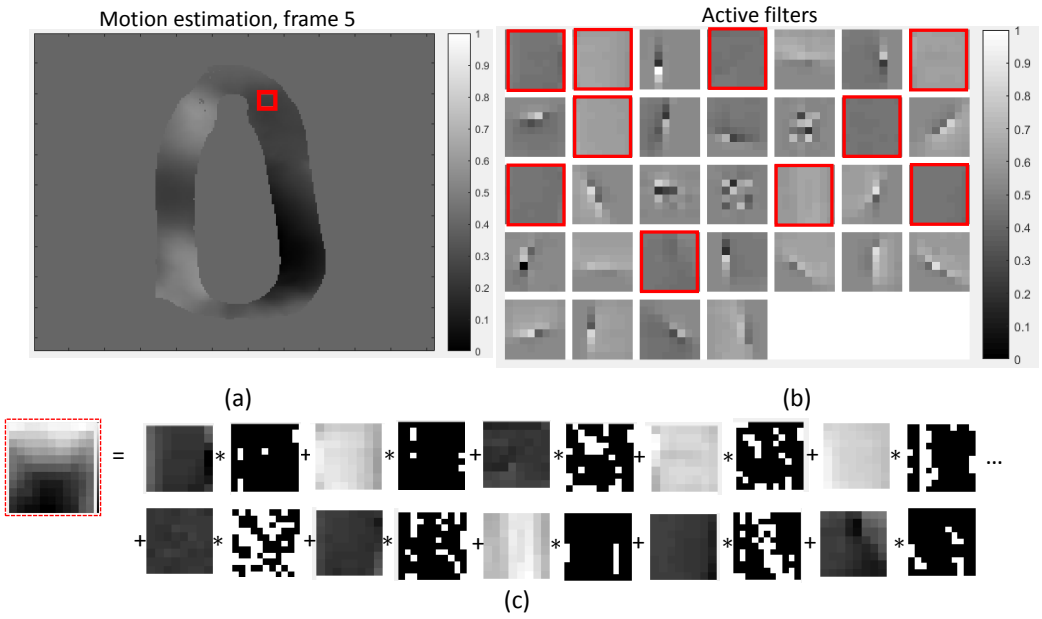


Figura 28 Example of a sparse representation of a cardiac motion patch using a convolutional dictionary. The filter size is $L \times L$ with $L = 8$ and the number of filters is $M = 32$.

This chapter is organized as follows. Section 5.2 introduces the motion estimation problem along with the proposed solution based on sparse regularization. Section 5.3 summarizes some key elements of CDL and its applications to cardiac motion estimation. Section 5.4 evaluates the performance of the proposed motion estimation method highlighting the interest of CDL for UI. Conclusions are reported in Section 5.5.

5.2. Motion estimation

The 2D motions for a pair of successive frames $(\mathbf{r}_k, \mathbf{r}_{k+1}) \in \mathbb{R}^{J \times N}$ acquired at time instants k and $k+1$ are denoted as $(\mathbf{s}_{k,x}, \mathbf{s}_{k,y}) \in \mathbb{R}^{J \times N}$ where $\mathbf{s}_{k,x}$ and $\mathbf{s}_{k,y}$ are the motions along the x and y axes. Since the motion estimation problem is considered independently, the displacement vector is equal to $\mathbf{s}_k = \mathbf{s}_{k,x}$ or $\mathbf{s}_k = \mathbf{s}_{k,y}$. The motion estimation field is formulated as the minimization of a

cost function with energy $E_{\text{data}}(\mathbf{s}_k)$ penalized by spatial and sparse regularizations, i.e.,

$$\underset{\mathbf{x},\mathbf{s}_k}{\text{argmín}}\left\{E_{\text{data}}(\mathbf{s}_k)+\lambda_d E_{\text{sparse}}(\mathbf{s}_k,\mathbf{x})+\lambda_s E_{\text{spatial}}(\mathbf{s}_k)\right\} \tag{20}$$

where $(\lambda_d,\lambda_s)\in\mathbb{R}^2$ are two parameters balancing the importance of the data fidelity and regularization terms.

5.2.1. Data fidelity term. The maximum likelihood (ML) method is a well accepted technique for motion estimation (Cohen and Dinstein, 2002). It maximizes the conditional probability density function of the measurement vector $\mathbf{r}_{k+1}$ given $\mathbf{r}_k$ and $\mathbf{s}$. The ML estimator is classically computed in the negative log-domain

$$\underset{\mathbf{s}}{\text{argmín}}-\ln\left[p(\mathbf{r}_{k+1})|\mathbf{r}_k(n),\mathbf{s}\right]. \tag{21}$$

Straightforward computations exploiting the Rayleigh statistics of ultrasound images detailed in (Ouzir et al., 2018) lead to the following data fidelity term

$$E_{\text{data}}(\mathbf{s})=-2d(\mathbf{s})+2\log[e^{2d(\mathbf{s})}+1]+C \tag{22}$$

where

$$d(\mathbf{s})=\frac{1}{b}\sum_{n=1}^{N}\left[\mathbf{r}_{k+1}(n+\mathbf{s}(n))-\mathbf{r}_k(n)\right]$$

n indicates the pixel index, $\mathbf{s} = [s(1), \ldots, s(N)]^T$ is the vectorized motion, $\mathbf{r}_k = [r_k(1), \ldots, r_k(N)]^T$ is the vectorized ultrasound image in frame k, and $C = -\log(2\sigma^4/b)$ is a known constant (depending on a scale parameter $\sigma \in \mathbb{R}^+$ and on the linear gain associated with the formation of the log-compressed B-mode image).

5.2.2. Spatial regularization. The spatial regularization term promotes the smoothness of the motion estimation field and is defined as

$$E_{\text{spatial}}(\mathbf{s}) = \|\nabla \mathbf{s}\|_2^2 \tag{23}$$

where ∇ denotes the gradient operator, and $\|.\|_2^2$ is the squared ℓ_2 norm which promotes low spatial gradients. This constraint imposes smooth fluctuations of the motion field which corresponds to a first-order spatial regularization (Horn and Schunck, 1981).

5.2.3. Sparse regularization. The proposed sparse regularization determines the motion $\mathbf{s}_k$ that is best represented as a convolution between M filters $\mathbf{d}_m$ and the sparse coefficient maps $\mathbf{x}_m$, i.e,

$$E_{\text{sparse}}(\mathbf{s}_k, \mathbf{x}_k) = \left\| \mathbf{s}_k - \sum_{m=1}^{M} \mathbf{x}_m * \mathbf{d}_m \right\|_2^2 \tag{24}$$

with a sparse constrain on $\mathbf{x}_m$ (which will appear in 27). The filters are composed of specific patterns contained in the training motions and $\mathbf{x}_m$ are the activation maps of each atom. Motivations for using this kind of regularization include the fact that convolutional sparse representations are invariant to translations contrary to standard dictionary learning techniques (Wohlberg, 2016). Linearly combining (22), (23) and (24) yields to the proposed energy which is minimized for motion

estimation in (20). This energy exploits the Rayleigh distribution of the noise and the spatial and sparse regularizations. The next section introduces the algorithm proposed to solve (20).

5.3. Optimization strategy

The regularization (24) assumes that an image frame of cardiac motion can be well represented by the sum of M convolutions between the coefficient maps and the corresponding filters. More precisely, the kth cardiac frame image $\mathbf{s}_k \in \mathbb{R}^{J\times N}$ is approximated as follows

$$\mathbf{s}_k \approx \sum_{m=1}^{M} \mathbf{d}_m * \mathbf{x}_m \tag{25}$$

where $*$ denotes the two-dimensional convolution. In order to solve (20), Algorithm 3 is proposed, which consists of three steps: 1) dictionary learning (see line 2), sparse coding (see line 3), and cardiac motion estimation (see line 6). These three steps are detailed below.

Algorithm 3 Motion estimation field for a pair of images using convolutional dictionary learning.

Input: $\mathbf{r}_{b,1}, \mathbf{r}_{b,2}, \lambda_s, \lambda_d, K, J, \lambda, \rho,\ \widetilde{\mathbf{s}} =$ campe2 motions, $\hat{\mathbf{s}} =$ campe1 motions
Output: $\mathbf{s}$

1: **function** MEFCDL($\mathbf{r}_{b,1}, \mathbf{r}_{b,2}, \lambda_s, \lambda_d, K, J, \lambda, \rho, \widetilde{\mathbf{s}}_0, \hat{\mathbf{s}}_0$)
2: $\quad\mathbf{d}_m \leftarrow$ Computes the dictionary by solving (26)
3: $\quad\mathbf{x}_m \leftarrow$ Computes the coefficient maps by solving 27
4: $\quad$**for** $k \leftarrow 1, K$ **do**
5: $\quad\quad$**for** $j \leftarrow 1, J$ **do**
6: $\quad\quad\quad$ $\text{argmín}_{\mathbf{s}}\{E_{\text{data}}(\mathbf{r}_{b,1}, \mathbf{r}_{b,2}, \mathbf{s}_{j-1}) +$
$\quad\quad\quad\quad\lambda_s\|\nabla\mathbf{s}_{j-1}\|_2^2 + \lambda_d(k)\|\mathbf{s}_{j-1} - \sum_m \mathbf{d}_m * \mathbf{x}_m\|_2^2\}$
$\quad\quad\quad\quad$ s.t. $\|\mathbf{d}_m\| = 1\ \forall m$ ▷ Motion estimation
7: $\quad$**return s** ▷ (Estimated motion field)

5.3.1. Dictionary learning. In the first step of Algorithm 3, a dictionary is estimated off-line by using a set of training cardiac motions denoted as $\widetilde{\mathbf{s}}$. The dictionary is obtained by

solving the following optimization problem

$$\begin{aligned} \underset{\mathbf{d}_m, \mathbf{x}_{k,m}}{\text{argmín}} \frac{1}{2} \sum_k \left\| \sum_m \mathbf{x}_{k,m} * \mathbf{d}_m - \widetilde{\mathbf{s}}_k \right\|_2^2 + \lambda \sum_m \sum_k \|\mathbf{x}_{k,m}\|_1 \\ \text{s.t.} \quad \|\mathbf{d}_m\| = 1 \; \forall m = 1, ..., M \end{aligned} \tag{26}$$

which was solved by using the alternating direction method of multipliers (ADMM) (Boyd et al., 2011).

5.3.2. Sparse coding. In the second step of Algorithm 3, the coefficient maps $\mathbf{x}_m$ are computed from cardiac motions $\hat{\mathbf{s}}_k$. More precisely, $\mathbf{x}_m$ is estimated by using the dictionary $\mathbf{d}_m$ obtained in (26) (see line 2) and by solving the following problem using ADMM

$$\underset{x_m}{\text{argmín}} \frac{1}{2} \left\| \sum_m \mathbf{x}_m * \mathbf{d}_m - \hat{\mathbf{s}}_k \right\|_2^2 + \lambda \sum_m \|\mathbf{x}_m\|_1. \tag{27}$$

5.3.3. Motion estimation . In the last step of Algorithm 3, the cardiac motion estimation are estimated using an algorithm similar to the one proposed in (Ouzir et al., 2018). In order to take into account the modified dictionary learning regularization, the following optimization problem is considered

$$\begin{aligned} \underset{\mathbf{s}}{\text{argmín}} \big\{ E_{\text{data}}(\mathbf{r}_{b,1}, \mathbf{r}_{b,2}, \mathbf{s}_{j-1}) + \lambda_s \|\nabla \mathbf{s}_{j-1}\|_2^2 + \lambda_d \\ \times \|\mathbf{s}_{j-1} - \sum_m \mathbf{d}_m * \mathbf{x}_m\|_2^2 \big\} \text{ s.t. } \|\mathbf{d}_m\| = 1 \; \forall m. \end{aligned} \tag{28}$$

The minimization problem (28) can be solved by setting the gradient of the cost function to zero and following the approach in (Myronenko et al., 2009). Note that the horizontal and vertical motions $\mathbf{s}_x$ and $\mathbf{s}_y$ are computed independently.

5.4. Experimental results

This section analyzes the performance of the proposed motion estimation based on CDL and compares it with standard dictionary learning (Ouzir et al., 2018) and with other state-of-the-art methods. For this comparison, this chapter considers highly realistic simulations performed using data with a controlled ground-truth generated using the method studied in (Alessandrini et al., 2015). The proposed approach is compared with the recent method of (Ouzir et al., 2018) (which showed very competitive results when compared to block-matching (Kaluzynski et al., 2001), the monogenic signal (Alessandrini et al., 2013), and B-splines (Myronenko et al., 2009)).

5.4.1. Simulation scenarios. Filters: The filters were computed with the pathological sequence LADdist by solving (26) with 500 iterations. The number of filters was set to $M = 32$, and the filter size was $L \times L$ with $L = 48$. The regularization parameter was $\lambda = 0.001$. Fig. 29 shows the corresponding set of convolutional dictionaries trained with horizontal, and vertical cardiac motions. Note that the filters trained with horizontal motions are depicted in Fig. 29 (a), (c), (e), (g), whereas the filters trained with vertical motions are shown in Fig. 29 (b), (d), (f), (h). Three scenarios were investigated: 1) one dictionary of filters for each frame of motions (top-row and Fig. 29(a), (b)), 2) one dictionary of filters for the whole cardiac cycle (second row and Fig. 29(c), (d)), 3) one dictionary for systole frames (frames 1-12) (third-row and Fig. 29(e), (f)) and one dictionary for diastole frames (frames 13-33) (bottom-row and Fig. 29(g), (h)). Finally, the

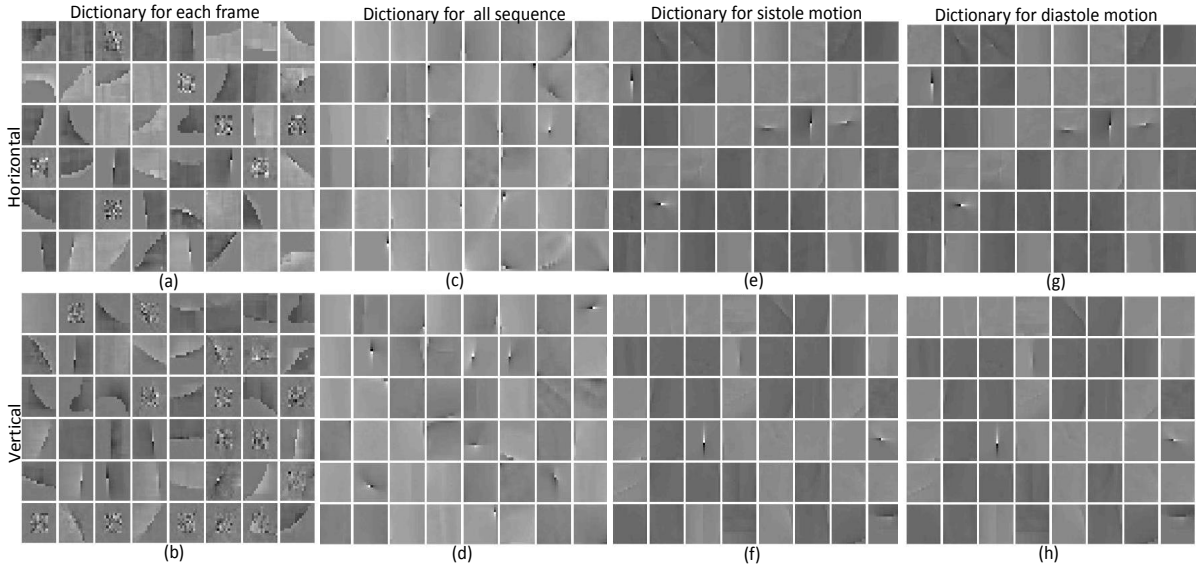


Figura 29 Dictionaries obtained using LADdist motions (with a filter size $L \times L$ with $L = 32$, and a number of filters $M = 48$): (first column) An example of filters for the 1^{st} frame. (second column) A dictionary estimated for the whole cardiac cycle. (third column) A dictionary for systole. (fourth column) A dictionary for diastole. The first rows (a),(c),(e),(g) show dictionaries trained with horizontal motions, (b),(d),(f),(h) are for dictionaries trained with vertical motions. All the dictionaries are obtained with 500 *iterations.*

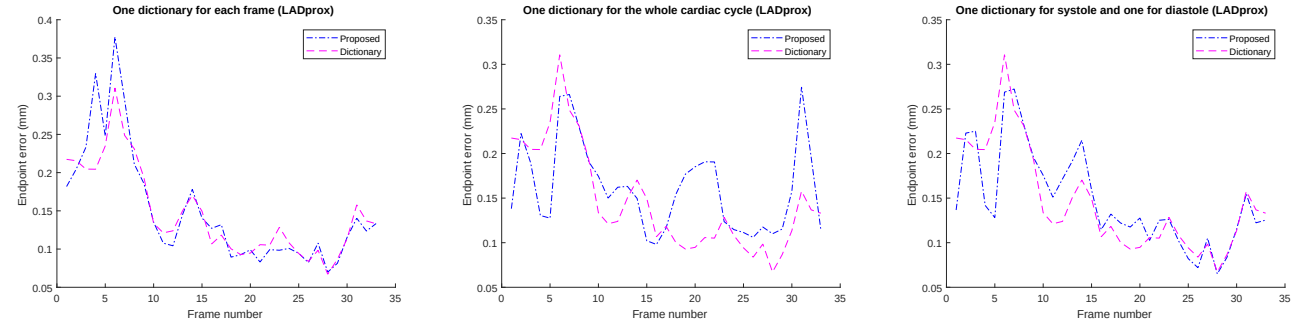


Figura 30 Mean endpoint error (in mm) for the LADprox sequence by (a) training a convolutional dictionary for each frame (error: 0.1556), (b) training a convolutional dictionary for the whole sequence (error: 0.1601) and (c) training two convolutional dictionaries (one for systole and one for diastole motions) (error: 0.147). The error for the method of (Ouzir et al., 2018) is 0.147.

coefficient maps were obtained from 27 with the pathological sequence of LADprox displacements (500 iterations) and a fixed dictionary of filters (see Section 5.4.1).

Motions: The regularization parameter was set to $\lambda_s = 0.75$ and λ_d was logarithmically

increased from 1×10^{-9} to 1×10^{-3} in 12 iterations. The parameters of the three steps of algorithm 3 are summarized in Table 2. In the table 3 is depicted a detailed cross-validation to select the filter size and filter number.

Table 2
Parameters for each step of algorithm 3, dictionary learning, sparse coding, and cardiac motion estimation.

Step	Parameters	Values
Dictionary learning	Database	LADdist
	Filter size	48×48
	Filters number	$M = 32$
	Sparsity term	$\lambda = 0.001$
	Number of iteration	500
Sparse coding	Database	LADprox
	Number of iteration	500
Cardiac motion estimation	Regularization parameter	$\lambda_s = 0.75$
	Sparsity term (Systole)	$\lambda_d = \{1 \times 10^{-6} \times 10^{-3}\}$
	Sparsity term (Diastole)	$\lambda_d = \{1 \times 10^{-9} \times 10^{-2}\}$

Table 3
Comparison of mean endpoint error varying the filters number and the filter size for a one dictionary systole and one dictionary for diastole. Notice the best result is attained with $M = 32$ and $L = 48$.

		Filter number, M				
		8	16	24	32	48
Filter size, L	8	0.1473	0.1477	0.1477	0.1475	0.1477
	16	0.1476	0.1477	0.1478	0.1478	0.1479
	24	0.1474	0.1476	0.1469	0.1470	0.1475
	32	0.1477	0.1471	0.1477	0.1469	0.1475
	40	0.1476	0.1473	0.1478	0.1482	0.1474
	48	0.1470	0.1467	0.1466	0.1465	0.1470

5.4.2. Performance measure. In order to evaluate the performance of the different methods, the endpoint error was computed as in (Alessandrini et al., 2013). This error is defined for the nth pixel as $e_n = \sqrt{[\mathbf{s}_x(n) - \hat{\mathbf{s}}_\mathbf{x}(n)]^2 + [\mathbf{s}_y(n) - \hat{\mathbf{s}}_y(n)]^2}$, where $\mathbf{s}_x(n)$, $\mathbf{s}_y(n)$, $\hat{\mathbf{s}}_x(n)$, $\hat{\mathbf{s}}_y(n)$ are the true and estimated (horizontal and vertical) motions at pixel n. Fig. 30 shows the mean endpoint error for the LADprox sequence and the three scenarios. The averaged endpoint errors for the different scenarios are (1) 0.1556 (2) 0.1601, (3) 0.147, with a preference for learning different dictionaries for the systole and diastole frames.

5.4.3. Realistic simulations. This section considers highly realistic simulations using B-mode ultrasound data published in (Alessandrini et al., 2016) including ground-truth of

horizontal and vertical motions. More precisely, the LADprox sequence was used, which corresponds to a proximal occlusion of the left anterior descending artery. Each sequence is a set of 3D images (with $224 \times 176 \times 208$ voxels, voxel size $0.7 \times 0.9 \times 0.6$ mm, frame rate $21 - 23$ Hz (Alessandrini et al., 2016)). The sequence contains 34 images of one complete cardiac cycle. For more details, the reader is invited to consult (Alessandrini et al., 2016) and (Alessandrini et al., 2015).

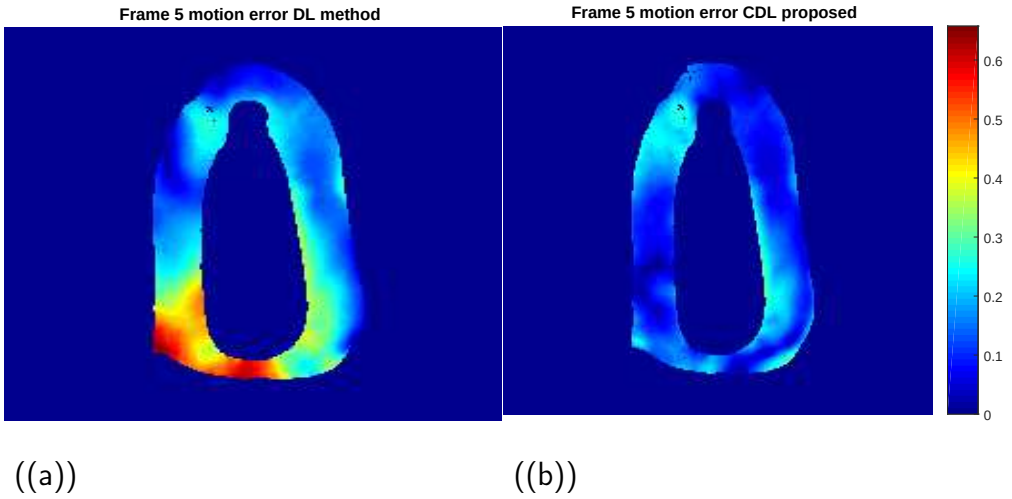


Figura 31 Error map for the 5th frame. Motion estimation using standard dictionary learning (Ouzir et al., 2018) (left) and proposed method (right).

In order to analyze the performance of the different algorithms, the error maps of the displacement estimates were computed. Fig. 31 displays the error maps for the 5th frame of the LADprox sequence. Fig. 31 (left) shows the error maps obtained for (Ouzir et al., 2018) (which uses a standard dictionary learning), and Fig. 31 (right) for the algorithm 3 (which uses CDL). Fig. 32 shows a representative example of the estimated motion field for the 5th frame which includes the ground-truth Fig. 32 (top), the approach in (Ouzir et al., 2018) with standard dictionary learning Fig. 32 (center), the proposed method in algorithm 3 by using convolutional dictionary Fig. 32 (bottom).

The zoomed version displays the motion vectors obtained using the approach in (Ouzir et al., 2018) (standard dictionary) and Algorithm 3 (convolutional dictionary). It is clear that the vector computed using a standard dictionary looks vertical. In contrast, the vector computed using CDL better matches the ground-truth. To show the interest of using a convolutional dictionary, the principal component analysis (Jackson, 2005) of the coefficient maps $\mathbf{x}$ was computed, in order to highlight some correlations between these coefficients and the cardiac motions. The coefficient maps for the 5^{th} frame were projected on their two first principal components and are shown in Fig. 33(a)(a). Clustering the projected coefficient maps using kmeans with 2 classes and comparing the resulting clusters displayed in Fig. 33(b)(b) with the corresponding motion estimation in Fig. 33(c)(c) shows a 94 % correlation between the coefficient maps and the cardiac motions, which shows the interest of using coefficient maps to represent the motions 33(d)(d).

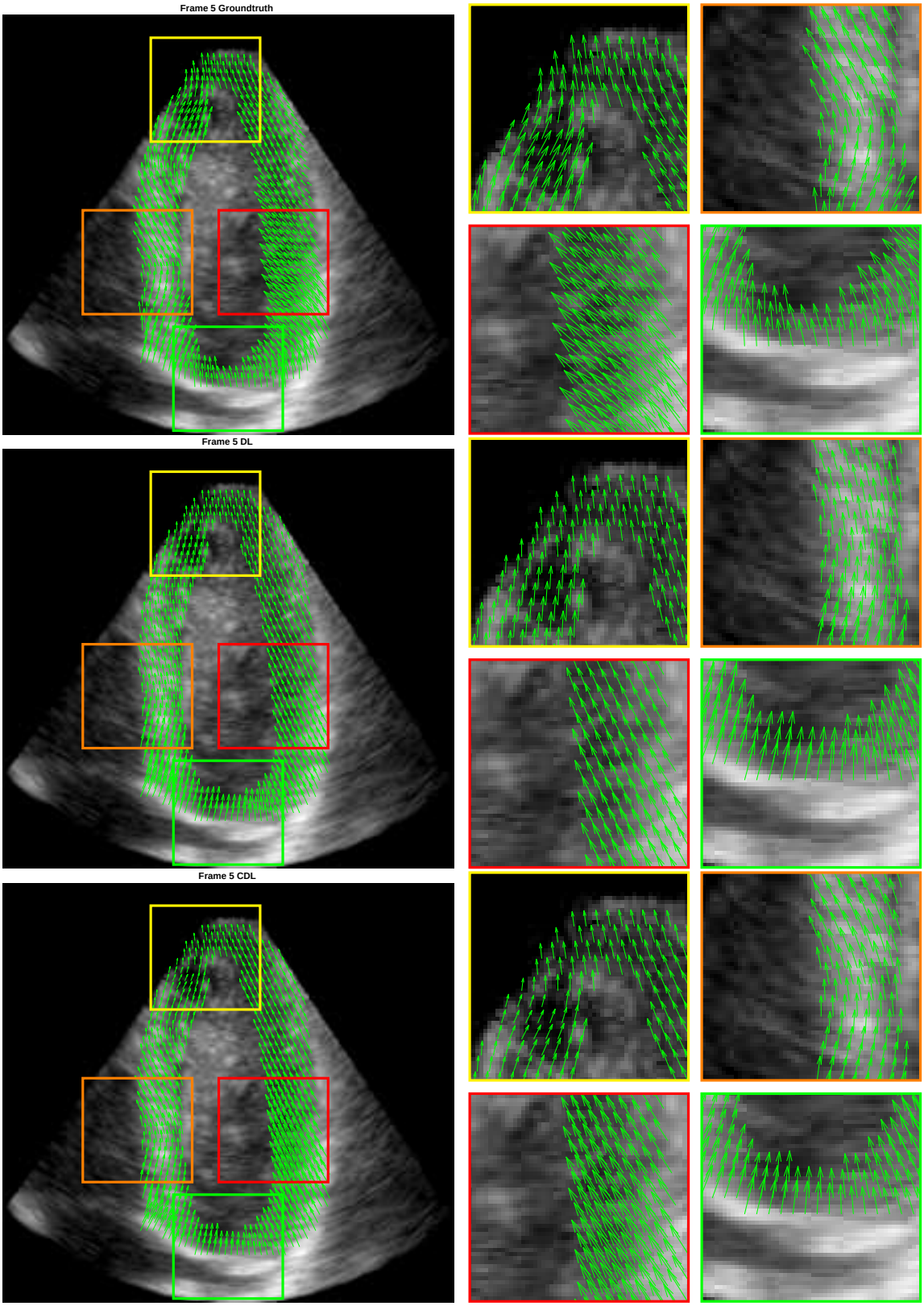


Figura 32 Ground-truth (top) and estimated meshes of the 5th frame of motion estimation using standard dictionary learning (Ouzir et al., 2018) (center), and motion estimation using convolutional dictionary (bottom).

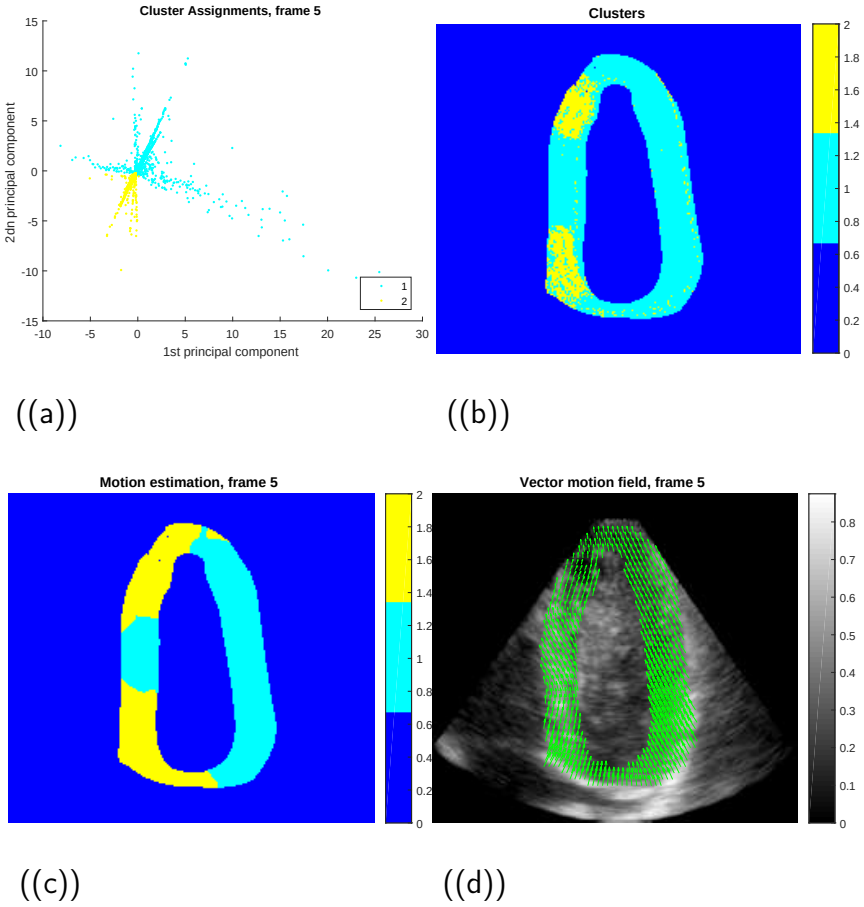


Figura 33 Projection of the coefficient maps on their two first principal components (a), and cluster assignment of the projected coefficients using kmeans with 2 classes (b). Motions associated with the 2 clusters identified by kmeans, which can be compared with the thresholding of the true motions of Fig. 33(c)(c). Thresholding of the motions for the 5^{th} frame (c) and their estimates (d).

5.5. Conclusions

This chapter introduced a new method for cardiac motion estimation in 2D ultrasound images based on a sparse decomposition of the motions on convolutional dictionaries. The method exploits the noise characteristics based on the B-mode distribution of ultrasound images and regularizes the estimation problem using a smoothing term and a sparse decomposition of the motions in a convolutional dictionary. The results obtained with the proposed method compares favourably with other state-of-the-art methods. An interesting property of the coefficients of the sparse de-

composition on a convolutional dictionary is a strong correlation with the corresponding motions. Future work will be devoted to the study of classification and anomaly detection methods based on the parameters resulting from this sparse decomposition on convolutional dictionaries.

6. Adaptive acquisition using motion estimation in compressive spectral video

6.1. Introduction

Traditional-C-CASSI has been employed in capturing static scenes, but its use might be extended to dynamic scenes. Recently, a C-CASSI-based architecture for compressive spectral video sensing (CSVS) has been called video-C-CASSI León-López et al. (2019). Previously, various CSVS architectures have been proposed to acquire spectral videos Cao et al. (2011); Llull et al. (2013); Ma et al. (2014); Tsai et al. (2015); Wang et al. (2015a). Most CSVS systems rely on coding and dispersion of the incoming light towards the camera sensor. However, video-CASSI improves flexibility in the encoding of the dynamic scenes by using an array of optical filters, which allows uniform sampling across frames León-López et al. (2019). Besides, the approach utilizes a regularization term to reduce the errors introduced by the motion of the scene; the motion is computed using optical flow. Approaches Ouzir et al. (2018); Ouzir et al. (2019); Diaz et al. (2019) for motion estimations have been proposed, these methods may be employed in the colored coded aperture design of dynamic scenes. Moreover, a few years ago, convolutional sparse coding (CSC) was introduced. CSC is an image representation translation-invariant suitable for dynamic scenes because it might exploit the temporal correlation between frames Wohlberg (2016).

Furthermore, traditional coded apertures-designs for C-CASSI are non-adaptive, which means that the coded apertures are independently designed from the scene Wagadarikar et al. (2009); Ma et al. (2014); Tsai et al. (2015); Cao et al. (2011); Wang et al. (2015a); Arguello and Arce (2014); Correa et al. (2016). Although, recently adaptive designs in C-CASSI Diaz et al.

(2018), and CASSI Diaz et al. (2019), have been proposed taking into account static scenes version. The motivation of this chapter is to utilize CSC which is translation-invariant, to represent the frames of a spectral video sequence, to compute the motion, and adaptively design the coded apertures by differentiating the sampling of the static parts of the scene to the sampling of the dynamic parts of the scene. The method estimates the motion by solving a minimization problem between pair of frames which involves a cost function with energy penalized by a spatial term to promote smoothness of the motion, and a sparsity regularization which uses convolutional sparse coding as image representation. Simulation results show that the proposed adaptive video-C-CASSI outperforms approaches of state-of-the-art such as random C-CASSI and blue-noise C-CASSI. This chapter is organized as follows: Section 6.2 presents the mathematical model via motion estimation presented in chapter 5 applied in C-CASSI. In section 6.3 is the proposed algorithm to design the colored coded apertures using motion estimation. Section 6.4 depicts the set of experiments and simulations to verify the performance of the proposed approach. Finally, conclusions are drawn in Section 6.5.

6.2. Coded aperture design in spectral videos using motion estimation.

6.2.1. Motion estimation.. Pairwise video motion estimation (VME) uses two consecutive frames $\mathbf{F}_H^{d-1}$ and $\mathbf{F}_H^d$ (of $\mathbb{R}^{M\times N\times L}$) from a spectral video acquired at time instants $d-1$ and d. Denote as $\mathbf{S}^d_{(\ell,x)}$ and $\mathbf{S}^d_{(\ell,y)} \in \mathbb{R}^{M\times N\times L}$ the video motions for the frame d along the x and y axes. The proposed VME method is inspired by the works conducted in Ouzir et al. (2018). It minimizes a function composed of a data fidelity term E_{data}, penalized by two regularization terms

E_{spatial} and E_{sparse}

$$\underset{\mathbf{X},\mathbf{S}^d}{\text{argmín}}\left\{E_{\text{data}}(\mathbf{S}^d,\mathbf{F}_H^d,\mathbf{F}_H^{d-1})+\lambda_s E_{\text{spatial}}(\mathbf{S}^d)+\lambda_p E_{\text{sparse}}(\mathbf{S}^d,\mathbf{X})\right\} \tag{29}$$

where $(\lambda_p,\lambda_s)\in\mathbb{R}^2$ are two regularization parameters (balancing the influence of the data fidelity term and the regularizations) and $\mathbf{S}^d=\mathbf{S}^d_{(\ell,x)}$ or $\mathbf{S}^d=\mathbf{S}^d_{(\ell,y)}$. Note that the displacement vectors components along x and y are estimated independently for simplicity. The first regularization term promotes smooth variations in the video motion field by using a standard total variation function, i.e., $E_{\text{spatial}}(\mathbf{S}^d)=\|\nabla\mathbf{S}^d\|_2^2$, where ∇ is the gradient operator and $\|.\|_2^2$ is the squared ℓ_2 norm. The other terms in (29) are defined below.

6.2.2. Data fidelity term. Optical flow assumes brightness constancy and temporal consistency leading to the following optical flow equation

$$\partial_t\mathbf{f}_H^d+\nabla\mathbf{f}_H^T\mathbf{s}^d=0 \tag{30}$$

where $\mathbf{s}^d\in\mathbb{R}^{NM}$ represents the flow field such that $\mathbf{s}_\ell^d$ is the vectorized video motion $\mathbf{S}_\ell$, $\partial_t\mathbf{f}_H^d$ denotes the temporal derivative and $\nabla\mathbf{f}_H^T$ is the spatial gradient of the brightness. The data fidelity term resulting from optical flow is

$$E_{\text{data}}(\mathbf{s}^d,\mathbf{f}_H^d,\mathbf{f}_H^{d-1})=\left\|\partial_t\mathbf{f}_H^d+\nabla\mathbf{f}_H^T\mathbf{s}^d\right\|_2^2 \tag{31}$$

where $\|.\|_2^2$ is the squared ℓ_2 norm.

6.2.3. Sparse regularization. The second regularization term promotes sparsity of the motion vectors in a dictionary of representative motions. It decomposes the video motion $\mathbf{S}^d$ as a convolution between V sparse coefficient maps $\mathbf{X}_v$ and a set of V filters $\mathbf{G}_v$, i.e.,

$$E_{\text{sparse}}(\mathbf{S}^d, \mathbf{X}) = \left\| \mathbf{S}^d - \sum_{v=1}^{V} \mathbf{G}_v * \mathbf{X}_v \right\|_2^2 \tag{32}$$

where $*$ denotes convolution. This regularization was used successfully for ultrasound images Diaz et al. (2019). One of the objectives of this chapter is to analyze its interest for multi-temporal hyperspectral images.

6.3. Adaptive coded aperture scheme

Fig. 34 summarizes the four steps proposed to design the video adaptive colored-coded apertures (VA-CCA). Specifically, the following iterative process is repeated for all the frames: (1) sample the datacube to capture the compressive measurements of a pair of frames, (2) reconstruct the datacube by solving an inverse problem to define some prior information Diaz et al. (2018), (3) estimate the video motions using optical flow, (4) design the colored-coded apertures by thresholding the motion estimates resulting from (3). Note that step (3) requires the computation of a convolutional dictionary (step (3a)) (which uses the training video motion $\widetilde{\mathbf{S}}$) and of coefficient maps ((step (3b)) using test video motions $\mathbf{S}_t$.

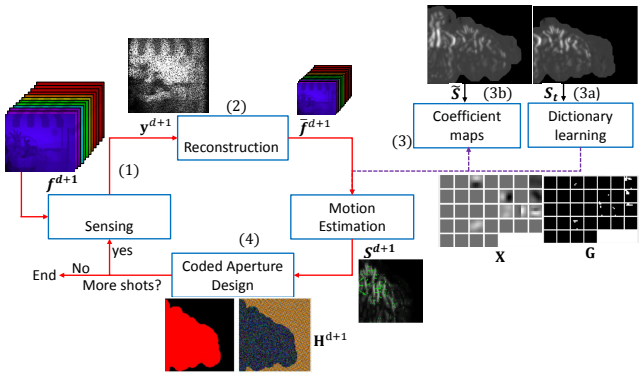


Figura 34 Flowchart of the proposed adaptive coded aperture design for VME.

6.3.1. Adaptive coded aperture design algorithm. Algorithm 4 designs the coded apertures adaptively for a compressive spectral video sequence. The algorithm uses two sets of compressive measurements and a sequence of motion fields estimated from a spectral video. It also requires to choose the hyperparameters λ_s, λ_p that balance the spatial and sparse regularization terms, respectively. Moreover, the alternating direction method of multipliers (ADMM) Boyd et al. (2011) used in this work requires to set the parameter ρ for the Lagrangian, and λ to balance the sparsity of the coefficients maps. Note that a low-resolution reconstruction of the datacube is necessary to obtain the video sequence appearing in lines 6 and 12 of the algorithm. This reconstruction is obtained by solving the following optimization problem $\hat{\mathbf{f}}_L^d = \mathbf{\Psi}_L^{-1}(\text{argmín}_{\boldsymbol{\theta}_L}\|\mathbf{y}^d - \mathbf{H}_L^d\mathbf{\Psi}_L^d\boldsymbol{\theta}_L^d\|_2^2 + \tau\|\boldsymbol{\theta}_L^d\|_1)$, where $\mathbf{y}$ contains the compressive measurements, $\mathbf{\Psi}_L^{-1}$ denotes the low-resolution representation basis, $\boldsymbol{\theta}_L^d$ represents the sparse signal, $\mathbf{H}_L^d$ corresponds to the measurement matrix, $\|.\|_1$ is the ℓ_1-norm, and τ is a regularization parameter. Line 2 computes the convolutional dictionary by assuming that each band in a specific spectral frame of the video can be represented as a set of V filters $\mathbf{G}_v$ convolved with a set of sparse coefficient maps

$\mathbf{X}_v$, i.e., $\mathbf{S}_d \approx \sum_{v=1}^{V} \mathbf{G}_v * \mathbf{X}_v$.

6.3.2. Dictionary learning. The determination of a convolutional dictionary satisfying $\mathbf{S}_d \approx \sum_{v=1}^{V} \mathbf{G}_v * \mathbf{X}_v$ can be achieved using several off-the-shelf algorithms Wohlberg (2016). In this chapter, the dictionary learning step is performed by solving the following problem (where $\widetilde{S}_d$ denotes the training video sequence which was obtained using Horn-Schunck optical flow estimation)

$$\begin{aligned} \underset{\mathbf{G}_v, \mathbf{X}_{d,v}}{\text{argmín}} \frac{1}{2} \sum_d \left\| \sum_v \mathbf{X}_{d,v} * \mathbf{G}_v - \widetilde{\mathbf{S}}_d \right\|_2^2 + \lambda \sum_{v=1}^{V} \sum_d \|\mathbf{X}_{d,v}\|_1 \\ \text{s.t.} \qquad \|\mathbf{G}_v\| = 1 \ \forall v = 1, ..., V. \end{aligned} \tag{33}$$

The minimization of (33) can be handled efficiently using the ADMM.

6.3.3. Sparse coding. Once the dictionary $\mathbf{G}_v$ has been determined, the coefficient maps of a sequence of test images denoted as $\mathbf{S}_t^d$ are obtained by solving the following optimization problem (see line 3 of Algorithm 4)

$$\underset{X_v}{\text{argmín}} \frac{1}{2} \left\| \sum_{v=1}^{V} \mathbf{X}_v * \mathbf{G}_v - \mathbf{S}_t^d \right\|_2^2 + \lambda \sum_{v=1}^{V} \|\mathbf{X}_v\|_1 \tag{34}$$

which can again be done using the ADMM algorithm.

6.3.4. Video reconstruction. In order to compute the adaptive coded apertures, some prior information is required. We propose to use a low resolution reconstruction of the image $\hat{\mathbf{f}}_L^d$ (see line 11 of Algorithm 4), where $\mathbf{H}_L^d$ is a decimated sensing matrix $\mathbf{H}_L^d = \mathbf{H}^d\mathbf{D}$, $\mathbf{D}$ is another decimation matrix such that $\hat{\mathbf{f}}_L^d = \mathbf{D}\hat{\mathbf{f}}^d$, the representation basis is denoted by $\mathbf{\Psi}_L^{-1}$ and the vectorized sparse signal is $\boldsymbol{\theta}_L^d$. An interpolation is used in Line 13 of Algorithm 4 to create a high

Algorithm 4 Adaptive coded aperture design for compressive spectral video using motion estimation.

Input: $\lambda_s, \lambda_p, K, D, \lambda, \rho,\ \widetilde{\mathbf{S}}, \mathbf{S}_t$: Training/test video motions
Output: $\mathbf{S}_\ell^d$

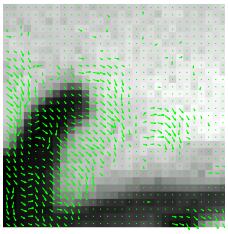
1: **function** CODED APERTURE DESIGN USING VIDEO MOTION ESTIMATION $(\mathbf{y}^0, \mathbf{y}^1, \lambda_s, \lambda_p, K, J, \lambda, \rho, \widetilde{\mathbf{S}}, \mathbf{S}_t)$
2: $\quad \mathbf{G}_v \leftarrow$ Computes the dictionary by solving (33)
3: $\quad \mathbf{X}_v \leftarrow$ Computes the coefficient maps by solving (34)
4: $\quad \mathbf{y}^0 \leftarrow \mathbf{H}^0\mathbf{f}$ ▷ First snapshot
5: $\quad \hat{\mathbf{f}}_L^0 \leftarrow \boldsymbol{\Psi}_L^{-1}(\text{argmín}_{\boldsymbol{\theta}_L}\|\mathbf{y}^0 - \mathbf{H}_L^0\boldsymbol{\Psi}_L^d\boldsymbol{\theta}_L^d\|_2^2 + \tau\|\boldsymbol{\theta}_L^d\|_1)$
6: ▷ Low-resolution
7: $\quad \hat{\mathbf{f}}_H^0 \leftarrow \mathbf{P}(\hat{\mathbf{f}}_L^0)$ ▷ Interpolation
8: $\quad \hat{\mathbf{F}}_H^0 \leftarrow rearrange(\hat{\mathbf{f}}_H^0)$ ▷ Rearrange
9: $\quad$ **for** $k \leftarrow 1, K$ **do**
10: $\quad\quad$ **for** $d \leftarrow 1, D$ **do**
11: $\quad\quad \hat{\mathbf{f}}_L^d \leftarrow \boldsymbol{\Psi}_L^{-1}(\text{argmín}_{\boldsymbol{\theta}_L}\|\mathbf{y}^d - \mathbf{H}_L^d\boldsymbol{\Psi}_L^d\boldsymbol{\theta}_L^d\|_2^2 + \tau\|\boldsymbol{\theta}_L^d\|_1)$
12: ▷ Low-resolution
13: $\quad\quad\quad \hat{\mathbf{f}}_H^d \leftarrow \mathbf{P}(\hat{\mathbf{f}}_L^d)$ ▷ Interpolation
14: $\quad\quad\quad \hat{\mathbf{F}}_H^d \leftarrow rearrange(\hat{\mathbf{f}}_H^d)$ ▷ Rearrange
15: $\quad\quad\quad$ **for** $\ell \leftarrow 1, L$ **do**
16: $\quad\quad\quad\quad \text{argmín}_{\mathbf{S}_\ell^d}\{E_{\text{data}}(\hat{\mathbf{F}}_H^{d-1}, \hat{\mathbf{F}}_H^d, \mathbf{S}_\ell^{d-1}) + \lambda_s\|\nabla\mathbf{S}_\ell^{d-1}\|_2^2 + \lambda_p(k)\|\mathbf{S}_\ell^{d-1} - \sum_v \mathbf{G}_v * \mathbf{X}_v\|_2^2\}$ s.t. $\|\mathbf{G}_v\| = 1\ \forall v$ ▷ Video motion estimation
17: $\quad\quad\quad\quad \mathbf{Q}_\ell^d \leftarrow (\mathbf{S}_\ell^{d-1}, \mathbf{S}_\ell^d)$ ▷ Thresholding motion
18: $\quad\quad\quad\quad \mathbf{q}_\ell^d \leftarrow vec(\mathbf{Q}_\ell^d)$ ▷ Vectorized motion areas
19: $\quad\quad\quad \mathbf{r}_\ell^d \leftarrow \mathbf{q}_\ell^d \odot \mathbf{b}_\ell^d + (\mathbf{1} - \mathbf{q}_\ell^d) \odot \hat{\mathbf{b}}_\ell^d$ ▷ Next code
20: $\quad\quad\quad\quad \mathbf{H}_\ell^d \leftarrow rearrange(\mathbf{r}_\ell^d)$ ▷ Rearrange
21: $\quad\quad\quad \mathbf{y}^d \leftarrow \mathbf{H}^d\mathbf{f}$ ▷ Next snapshot
22: $\quad \hat{\mathbf{f}} \leftarrow \boldsymbol{\Psi}^{-1}(\text{argmín}_{\boldsymbol{\theta}}\|\mathbf{y} - \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}\|_2^2 + \tau\|\boldsymbol{\theta}\|_1)$
23: $\quad$ **return** $\mathbf{S}_\ell^d$ ▷ (Estimated motion field)

resolution image $\hat{\mathbf{f}}_H^d$ from a low resolution reconstruction $\hat{\mathbf{f}}_L^d$, where $\mathbf{P}$ is a bilinear interpolator. The interpolated datacube $\hat{\mathbf{f}}_H^d$ is rearranged in Line 14 to obtain the datacube of the dth frame $\hat{\mathbf{F}}_H^d$.

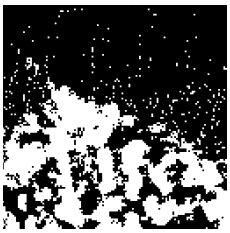
The VME is reported in Line 16. In line 17 of the algorithm, a thresholding using a binary mask $\mathbf{Q}_\ell^d$ divides each pixel in background and target areas using a pair of motions $\mathbf{S}_\ell^{d-1}$, and $\mathbf{S}_\ell^d$.

The matrix $\mathbf{Q}_\ell^d$ is vectorized as $\mathbf{q}_\ell^d$ in line 18 and the resulting coded aperture $\mathbf{r}_\ell^d$ is computed in line 19. Note that the coded aperture depends on two blue noise code apertures $\hat{\mathbf{b}}_\ell^d$ and $\mathbf{b}_\ell^d$ Correa et al. (2016). One blue noise coded aperture corresponds to the background area (complement of matrix $\mathbf{q}_\ell^d$, $\mathbf{1}-\mathbf{q}_\ell^d$) and a moving blue noise coded aperture $\mathbf{b}_\ell^d$ is attributed to the mobile target (the subset of the scene $\mathbf{q}_\ell^d$). For an example of designed codes, the reader is invited to look at Fig. 35 in Section 6.4. In line 20 the coded aperture is reorganized leading to the matrix $\mathbf{H}_\ell^d$. In the following step (line 21) the compressive measurements $\mathbf{y}_d$ are obtained by sampling with the adaptive coded aperture $\mathbf{H}_\ell^d$. The last step in line 22 reconstructs the spectral video with improved quality due to the designed patterns.

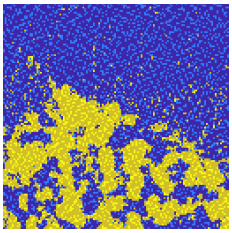
(a) Fr. 1 motion estimation.



(b) Fr. 1 zoomed version.



(c) Threshold estimation



(d) Adaptive Video CCA

Figura 35 (a) Example of motion field of the 1st frame, and the 1st spectral band. (b) Zoomed version of the motion field which corresponds to the red square in (a). (c) Threshold motion estimation which divides in two the video scene, i.e., static area and dynamic area. (d) Hybrid blue noise coded aperture, notice that a yellow area depicts the dynamic blue noise coded aperture, and the blue area represent the static blue noise coded aperture.

6.4. Simulations and results

In order to validate the performance of the proposed coded aperture design, a set of C-CASSI video measurements was simulated using the model (12). These measurements were constructed using a real test spectral source acquired in the Optics Lab of the High Dimensional Signal Processing (HDSP) research group at Universidad Industrial de Santander with a CCD camera using wavelength steps of 10 nm. The resulting discrete source $\mathbf{F}$ used in simulations has 12 frames of 128×128 pixels and $L = 10$ spectral bands ranging from 400 nm to 500 nm. Given the compressive projections, the compressive sensing algorithm GPSR (Gradient Projection for Sparse Reconstruction) was used to recover the data Figueiredo et al. (2007). The 4D sparse representation basis used in this experiment was the Kronecker product between a 2D-Wavelet Symmlet 8 basis for the spatial dimensions denoted as $\mathbf{\Psi}^{2D}$, a 1D-Discrete Cosine basis (DCT) for the spectral dimension denoted as $\mathbf{W}$ and a 1D-DCT basis for the temporal dimension denoted as $\mathbf{U}$ Correa-Pugliese et al. (2016).

Algorithm 4 was used to design the video adaptive colored-coded apertures (VA-CCA) in an adaptive manner. The designed apertures were then used to create a C-CASSI spectral source and to reconstruct the image sequence of interest. The performance of the designed apertures was compared with random colored-coded apertures (R-CCA) Arguello and Arce (2014) with the same transmittance, with blue noise apertures (BNA) Correa et al. (2016), and with 50 % transmittance blocking-unblocking coded apertures (BUA) Galvis-Carreño et al. (2014). For the experiments, the estimated low-resolution image had a spatial resolution of 32×32 pixels, which corresponds

to a spatial downsampling by a factor of 4. An example of the motion field obtained for the first frame, and the first spectral band is depicted in Fig. 35. Fig. 35(a) displays the motion field, Fig. 35(b) shows a zoomed version of the motion field, Fig. 35(c) depicts the Otzu thresholding of the motion field, which divides the scene into static and dynamic regions, Fig. 35(d) displays a hybrid blue noise coded aperture which is composed of one blue noise code aperture for the static part of the scene and a dynamic blue noise that moves to keep the complementarity between frames.

The quality of image reconstructions was evaluated in terms of peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM). The PSNR, given in decibels (dB), is related to the mean squared error (MSE) as $10\log_{10}(\text{máx}^2/\text{MSE})$, where máx represents the maximum possible value of an image pixel. SSIM measures the structure similarity between two imageswith values varying from 0 to 1, 1 being the value obtained for two identical images. Table 4 summarizes the results in terms of PSNR mean and SSIM for the different coded apertures. The PSNR and SSIM obtained with the proposed VA-CCA patterns are higher than those obtained with BUA, BNA, and R-CCA. Fig. 36 shows the RGB reconstructions of the frames #4, #8, and #12 and provides the quality of reconstructions in terms of PSNR. The VA-CCA provides the best reconstructions, with a PSNR up to 2 dB higher than the block-unblock coded apertures, 2.9 dB higher than blue noise coded apertures, and 0.7 dB higher than the random colored-coded apertures. These results are very promising.

6.5. Conclusions

This chapter studied a new design of adaptive colored-coded apertures (VA-CCA) for compressive spectral sampling in video C-CASSI. After introducing a mathematical model for dynamic

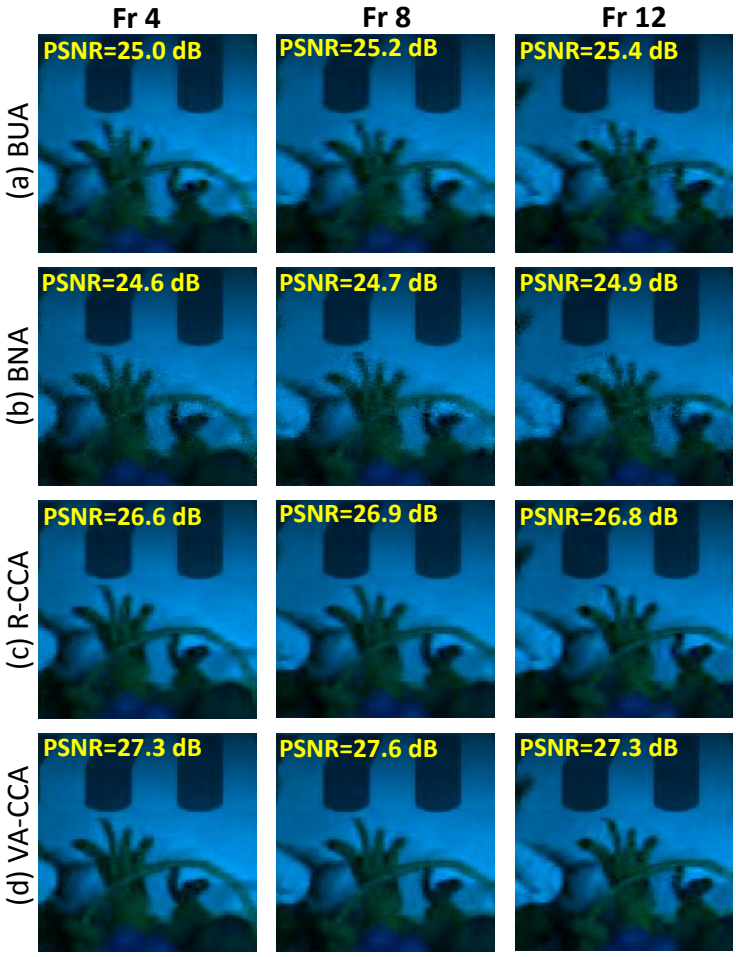


Figura 36 Comparison of video quality reconstruction in frames 4, 8, and 12 respectively per column, using non-adaptive and adaptive coded apertures. (a) block-unblock apertures, (b) blue noise apertures, (c) random CCA, (d) video adaptive CCA. The average PSNR across the spectral band is shown in each frame.

Table 4

Average PSNR and SSIM of the Reconstruction in dB Across the Spectral and Temporal dimension.

Coded Aperture Patterns	BU	BN	CCA	VA-CCA
PSNR **average**	25.04	24.56	26.75	27.42
SSIM **average**	0.844	0.843	0.895	0.906

scenes, we investigated the colored-coded apertures in C-CASSI relying on prior information on the scene, provided by motion estimation between pairs of consecutive frames. The proposed design divides the video scene into static and dynamic regions, assigning two different blue noise patterns to the static and mobile parts of the image. The improvement in PSNR obtained with the proposed approach is up to 2 dB compared to traditional blocking-unblocking apertures, 2.9 dB in comparison with non-adaptive blue noise patterns, and 0.7 dB higher than random colored-coded apertures.

Conclusions. Spectral images are a crucial technology in remote sensing, astronomy, and medical imaging. However, capturing the datacube is a difficult task due to issues of acquisition time, storage, and transmission. Nowadays, the most outstanding technology to acquire SI is compressive spectral imaging, which is snapshot technology based on compressive sensing. The critical element of those systems is the coded aperture, whose designs improve the performance. For that reason, coded aperture design still is an open problem. The objective of this Ph.D. dissertation was to propose novel approaches of coded aperture design for CSI architectures. The principle of this work was to design the coded aperture adaptively. Three applications have been identified to design coded aperture adaptively. First, the improvement of the quality of image reconstruction in static scenes, second the enhancement of classification spectral images in the compressive domain, and third the improving of the quality of image reconstruction in dynamic scenes. For each of these applications an algorithm that adaptively designs the coded aperture has been proposed.

The first adaptive algorithm introduced in chapter 2 computes the coded aperture using a low-resolution reconstruction of the spectral image. The second adaptive algorithm was presented in chapter 3, and it uses the compressive measurements of dual-arm CSI architecture to compute the coded aperture adaptively, promoting the spatial contextual information in the compressive measurements in order to improve the classification accuracy. The third adaptive method exploits a novel regularization algorithm that uses the fact that the motion can be decomposed in a convolutional dictionary of displacements. The proposed motion estimation approach extends to an

adaptive algorithm presented in chapter 5 to design the coded aperture in video CSI.

The first proposed algorithm adaptively designs the coded aperture using as prior information a low-resolution reconstruction of the scene to characterize the coding filters. The method performs thresholding to the second derivative of each spectral signature present in the scene. That means that the coded aperture samples the scene according to the matched filter. Simulation and implementation results show that the quality of image reconstruction improves in up to 2 dB using the proposed matched filter adaptive approach.

The second adaptive approach proposes a scheme that performs classification from the compressive measurements. The approach uses an architecture of two arms. One arm to capture compressive spectral measurements and one arm to acquire compressive spatial measurements. The adaptive approach samples the compressive measurement preserving the contextual information. The contextual information is crucial in order to improve classification accuracy because it improves the features which are the inputs of the classifier. The algorithm first measures the grayscale image, then performes high-pass filtering and the quantization of that filtered image. The coded aperture of both arms is designed according to the quantized image, which is referred to as contextual information. The contextual information indicates where the spectral signature might be similar. In this way, the compressive measurements preserve the spatial structure improving the classification accuracy.

The third proposed method estimates the motion of dynamic scenes. The motion estimation method is initially developed for ultrasound images. In particular, the proposed motion estimation

is composed of a minimization cost function including a data fidelity term penalized by two regularization terms, one that promotes spatial smoothness and other that encourages sparsity. The spatial regularization promotes smooth fluctuation in motion. The sparse regularization induces that an image will be represented by the summation of M filters convolved with the coefficient maps. The simulations show comparable results between the proposed method and the methods of state-of-the-art. Also, it shows the high correlation between the coefficient maps and the motion estimation.

In fourth method, the motion estimation is employed in compressive spectral imaging to track the object that moves in the scene in order to design the sampling of the compressive measurement adaptively. The coded aperture is adaptively designed in the video C-CASSI to attain compressive measurements that allow reconstructing multispectral video with high image quality. The proposed approach uses two sets of blue noise coded apertures, one that remains static and one that moves with the dynamic object. The proposed method compares against approaches of state of the art, for instance, block-unblock coded apertures, random colored-coded apertures, static blue-noise colored-coded aperture overcoming those approaches of state of the art.

Perspective. Each of the following paragraphs illustrates about the future work:

- **Chapter 2:** Previous works in the literature have reported the design of colored coded apertures using just a reduced number of optimal filters (Arguello and Arce, 2014), including high-pass, low-pass, and band-pass filters. In future works, the adaptive designed introduced

in chapter 3 and (Diaz et al., 2018), would focus in limiting the set of filters so as to reduce the implementation complexity; sideways, it would be interesting to optimize the entries of such limited set of filters to satisfy the restricted isometric property tightly.

- **Chapter 3:** To implement the adaptive classification approach from real compressive measurements obtained by a laboratory test-bed of the dual-arm architecture. Also, it is worth to compare the support vector machine classifier against novel classifier methods based on deep learning (Li et al., 2019) such as stacked autoencoders (SAEs), deep belief networks (DBNs), convolutional neural network (CNN), recurrent neural networks (RNN), and generative adversarial networks (GANs).

- **Chapter 4:** The cardiac motion estimation algorithm might be evaluated using *in vivo* data in order to test the proposed approach in a real scenario, comparing the performance with a robust approach (Ouzir et al., 2019) and Spatio-temporal approach. Moreover, the proposed cardiac motion estimation approach is pairwise. However, it does not exploits the Spatio-temporal information present in the whole image sequence. A way of computing the motion estimation is taking into account the Spatio-temporal information.

- **Chapter 5:** To make a test-bed implementation of the adaptive video compressive spectral imaging scheme. A factor that has a significant impact on the performance of the approach is the frame rate and the presence of outliers. Some test could be performed using a dataset of ultrafast video dataset. Also, robust motion estimation algorithms could be used in order to mitigate the presence of outliers which made challenging to segment the dynamic scene

from the static scene correctly.

References

Aeron, S., Saligrama, V., and Zhao, M. (2010). Information theoretic bounds for compressed sensing. *IEEE Transactions on Information Theory*, 56(10):5111–5130.

Alessandrini, M., Basarab, A., Liebgott, H., and Bernard, O. (2013). Myocardial motion estimation from medical images using the monogenic signal. *IEEE Trans. Image Process.*, 22(3):1084–1095.

Alessandrini, M., De Craene, M., Bernard, O., Giffard-Roisin, S., Allain, P., Waechter-Stehle, I., Weese, J., Saloux, E., Delingette, H., Sermesant, M., and D'hooge, J. (2015). A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: Methodology and open-access database. *IEEE Trans. Med. Imaging*, 34(7):1436–1451.

Alessandrini, M., Heyde, B., Queirós, S., Cygan, S., Zontak, M., Somphone, O., Bernard, O., Sermesant, M., Delingette, H., Barbosa, D., De Craene, M., O'Donnell, M., and D'hooge, J. (2016). Detailed evaluation of five 3D speckle tracking algorithms using synthetic echocardiographic recordings. *IEEE Trans. Med. ImaMyronenko-2009ging*, 35(8):1915–1926.

Amato, U., Antoniadis, A., Carfora, M. F., Colandrea, P., Cuomo, V., Franzese, M., Pignatti, S., and Serio, C. (2013). Statistical classification for assessing prisma hyperspectral potential for agricultural land use. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):615–625.

Arce, G. R., Brady, D. J., Carin, L., Arguello, H., and Kittle, D. S. (2014). Compressive Coded Aperture Spectral Imaging: An Introduction. *IEEE Signal Processing Magazine*, 31(1):105–115.

Arguello, H. and Arce, G. R. (2011). Code aperture optimization for spectrally agile compressive imaging. *Journal of the Optical Society of America A*, 23:2400–2413.

Arguello, H. and Arce, G. R. (2013). Rank Minimization Code Aperture Design for Spectrally Selective Compressive Imaging. *IEEE Transactions on Image Processing*, 22(3):941–954.

Arguello, H. and Arce, G. R. (2014). Colored Coded Aperture Design by Concentration of Measure in Compressive Spectral Imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908.

Arguello, H., Rueda, H., Wu, Y., Prather, D., and Arce, G. R. (2012). Higher-order computatioal model for coded aperture spectral imaging. *Applied Optics*, 52:D12–D21.

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

Averbuch, A., Dekel, S., and Deutsch, S. (2012). Adaptive compressed image sensing using dictionaries. *SIAM J. Img. Sci.*, 5(1):57–89.

Baldry, I. and Bland-Hawthorn, J. (2000). A tunable echelle imager. *Publications of the Astronomical Society of the Pacific*, 112(774):1112.

Baraniuk, R. (2007). Compressive sensing. *IEEE Signal Processing Magazine*, 24:118–121.

Bedard, N., Schwarz, R. A., Hu, A., Bhattar, V., Howe, J., Williams, M. D., Gillenwater, A. M., Richards-Kortum, R., and Tkaczyk, T. S. (2013). Multimodal snapshot spectral imaging for oral cancer diagnostics: a pilot study. *Biomed. Opt. Express*, 4(6):938–949.

Bilbro, G. (2001). Technology options for multi-spectral infrared cameras. Technical report, DTIC Document.

Bowen, I. (1938). The image-slicer a device for reducing loss of light at slit of stellar spectrograph. *apj*, 88:113.

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

Candes, E. J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509.

Cao, X., Tong, X., Dai, Q., and Lin, S. (2011). High resolution multispectral video capture with a hybrid camera system. In *CVPR 2011*, pages 297–304.

Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Carin, L., and Brady, D. J. (2016). Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108.

Castro, R. M. (2012). Adaptive sensing performance lower bounds for sparse signal detection and support estimation. *ArXiv e-prints*.

Chen, Y., Nasrabadi, N. M., and Tran, T. D. (2011). Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3973–3985.

Coffey, V. C. (2015). Hyperspectral imaging for safety and security. *Opt. Photon. News*, 26(10):26–33.

Cohen, B. and Dinstein, I. (2002). New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition*, 35(2):455 – 463.

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

Correa, C. V., Arguello, H., and Arce, G. R. (2016). Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging. *JOSA A*, 33(12):2312–2322.

Correa-Pugliese, C. V., Galvis-Carreño, D. F., and Arguello, H. (2016). Sparse representations of dynamic scenes for compressive spectral video sensing. *DYNA*, 83(195):42–51.

Courtès, G. (1960). Méthodes d'observation et étude de l'hydrogène interstellaire en émission. *Annales d'Astrophysique*, 23:115.

Diaz, N., Basarab, A., Tourneret, J., and Fuentes, H. A. (2019). Cardiac motion estimation using

convolutional sparse coding. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5.

Diaz, N., Chacon, H. R., and Fuentes, H. A. (2015). High-dynamic range compressive spectral imaging by grayscale coded aperture adaptive filtering. *Ingeniería e Investigación*, 35(3):53–60.

Diaz, N., Hinojosa, C., and Arguello, H. (2019). Adaptive grayscale compressive spectral imaging using optimal blue noise coding patterns. *Optics & Laser Technology*, 117:147–157.

Diaz, N., Rueda, H., and Arguello, H. (2018). Adaptive filter design via a gradient thresholding algorithm for compressive spectral imaging. *Appl. Opt.*, 57(17):4890–4900.

Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306.

Dunlop-Gray, M., Poon, P., Golish, D., Vera, E., and Gehm, M. (2016). Experimental demonstration of an adaptive architecture for direct spectral imaging classification. *Opt. Express*, 24(16):18307–18321.

Falco, N., Benediktsson, J. A., and Bruzzone, L. (2014). A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2183–2199.

Falco, N., Benediktsson, J. A., and Bruzzone, L. (2015). Spectral and spatial classification of hyperspectral images based on ica and reduced morphological attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11):6223–6240.

Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., and Tilton, J. C. (2013). Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675.

Figueiredo, M. A. T., Nowak, R. D., and Wright, S. J. (2007). Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597.

Foster, D. H., Amano, K., Nascimento, S. M. C., and Foster, M. J. (2006). Frequency of metamerism in natural scenes. *J. Opt. Soc. Am. A*, 23(10):2359–2372.

France, F. G. (2011). Advanced spectral imaging for noninvasive microanalysis of cultural heritagematerials: Review of application to documents in the u.s. library of congress. *Appl. Spectrosc.*, 65(6):565–574.

Galvis, L., Lau, D., Ma, X., Arguello, H., and Arce, G. R. (2017). Coded aperture design in compressive spectral imaging based on side information. *Appl. Opt.*, 56(22):6332–6340.

Galvis-Carreño, D. F., Mejía-Melgarejo, Y. H., and Arguello-Fuentes, H. (2014). Efficient reconstruction of Raman spectroscopy imaging based on compressive sensing. *DYNA*, 81(188):116–124.

Gao, L., Kester, R. T., and Tkaczyk, T. S. (2009). Compact image slicing spectrometer (iss) for hyperspectral fluorescence microscopy. *Opt. Express*, 17(15):12293–12308.

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

Golub, M. A., Averbuch, A., Nathan, M., Zheludev, V. A., Hauser, J., Gurevitch, S., Malinsky, R., and Kagan, A. (2016). Compressed sensing snapshot spectral imaging by a regular digital camera with an added optical diffuser. *Appl. Opt.*, 55(3):432–443.

Gonzalez, R. C. and Woods, R. E. (2006). *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Goodenough, D. G., Dyk, A., Niemann, K. O., Pearlman, J. S., Hao Chen, Han, T., Murdoch, M., and West, C. (2003). Processing hyperion and ali for forest classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1321–1331.

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

Harvey, A. R. and Fletcher-Holmes, D. W. (2003). High-throughput snapshot spectral imaging in two dimensions. In *Spectral Imaging: Instrumentation, Applications, and Analysis II*, volume 4959, pages 46–54. International Society for Optics and Photonics.

Hinojosa, C., Correa, C., Arguello, H., and Arce, G. R. (2016). Compressive spectral imaging using multiple snapshot colored-mosaic detector measurements. *Proc.SPIE*, 9870:9870 – 9870 – 10.

Hirai, A., Inoue, T., and Ichioka, Y. (1994). Application of multiple-image fourier transform spectral imaging to measurement of fast phenomena. *Optical Review*, 1(2):205–207.

Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1):185 – 203.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63.

Jackson, J. (2005). *A User's Guide to Principal Components*. Wiley Series in Probability and Statistics. Wiley.

Ji, S., Xue, Y., and Carin, L. (2008). Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356.

Kaluzynski, K., Emelianov, S. Y., Skovoroda, A. R., and O'Donnell, M. (2001). Strain rate imaging using two-dimensional speckle tracking. *Trans. Ultrason., Ferroelectr., Freq. Control*, 48(4):1111–1123.

Kapany, N. (1968). *Fiber optics; principles and applications*. Academic Press.

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

Kudenov, M. W. and Dereniak, E. L. (2012). Compact real-time birefringent imaging spectrometer. *Opt. Express*, 20(16):17973–17986.

Kudenov, M. W., Jungwirth, M. E. L., Dereniak, E. L., and Gerhart, G. R. (2010). White-light sagnac interferometer for snapshot multispectral imaging. *Appl. Opt.*, 49(21):4067–4076.

León-López, K. M., Carreño, L. V. G., and Fuentes, H. A. (2019). Temporal Colored Coded Aperture Design in Compressive Spectral Video Sensing. *IEEE Transactions on Image Processing*, 28(1):253–264.

Li, J., Bioucas-Dias, J. M., and Plaza, A. (2012). Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3):809–823.

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., and Benediktsson, J. A. (2019). Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709.

Lin, X., Liu, Y., Wu, J., and Dai, Q. (2014a). Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. Graph.*, 33(6):233:1–233:11.

Lin, X., Wetzstein, G., Liu, Y., and Dai, Q. (2014b). Dual-coded compressive hyperspectral imaging. *Opt. Lett.*, 39(7):2044–2047.

Llull, P., Liao, X., Yuan, X., Yang, J., Kittle, D., Carin, L., Sapiro, G., and Brady, D. J. (2013). Coded aperture compressive temporal imaging. *Optics Express*, 21(9):10526–10545.

Ma, C., Cao, X., Tong, X., Dai, Q., and Lin, S. (2014). Acquisition of High Spatial and Spectral

Resolution Video with a Hybrid Camera System. *International Journal of Computer Vision*, 110(2):141–155.

Malloy, M. L. and Nowak, R. D. (2014). Near-optimal adaptive compressed sensing. *IEEE Transactions on Information Theory*, 60(7):4001–4012.

Mathews, J. H. and Fink, K. D. (1998). *Numerical Methods Using MATLAB*. Simon & Schuster, 3rd edition.

Mejia, Y. and Arguello, H. (2016). Filtered gradient reconstruction algorithm for compressive spectral imaging. *Optical Engineering*, 56:56 – 56 – 11.

Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790.

Myronenko, A., Song, X., and Sahn, D. J. (2009). Maximum likelihood motion estimation in 3D echocardiography through non-rigid registration in spherical coordinates. In Ayache, N., Delingette, H., and Sermesant, M., editors, *Functional Imaging and Modeling of the Heart*, pages 427–436, Berlin, Heidelberg. Springer Berlin Heidelberg.

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

Ouzir, N., Basarab, A., Lairez, O., and Tourneret, J. (2019). Robust optical flow estimation in car-

diac ultrasound images using a sparse representation. *IEEE Transactions on Medical Imaging*, 38(3):741–752.

Ouzir, N., Basarab, A., Liebgott, H., Harbaoui, B., and Tourneret, J.-Y. (2018). Motion estimation in echocardiography using sparse representation and dictionary learning. *IEEE Trans. Image Process.*, 27(1):64–77.

Papyan, V., Romano, Y., and Elad, M. (2017). Convolutional neural networks analyzed via convolutional sparse coding. *J. Mach. Learn. Res.*, 18(1):2887–2938.

Ramirez, A., Arguello, H., Arce, G. R., and Sadler, B. M. (2014). Spectral image classification from optimal coded-aperture compressive measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6):3299–3309.

Ramirez, J. M. and Arguello, H. (2019a). Multiresolution compressive feature fusion for spectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–12.

Ramirez, J. M. and Arguello, H. (2019b). Spectral image classification from multi-sensor compressive measurements. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–11.

Ratle, F., Camps-Valls, G., and Weston, J. (2010). Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2271–2282.

Romberg, J. (2008). Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25:14–20.

Romberg, J. (2009). Compressive sensing by random convolution. *SIAM J. Img. Sci.*, 2(4):1098–1128.

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

Samuel, P., Vargas, H., and Arguello, H. (2016). Probability of correct reconstruction in compressive spectral imaging. *Ingeniería e Investigación*, 36:68 – 77.

Sellar, R. G. and Boreman, G. D. (2005a). Classification of imaging spectrometers for remote sensing applications. *Optical Engineering*, 44(1):013602–013602–3.

Sellar, R. G. and Boreman, G. D. (2005b). Comparison of relative signal-to-noise ratios of different classes of imaging spectrometer. *Appl. Opt.*, 44(9):1614–1624.

Starling, D. J., Storer, I., and Howland, G. A. (2016). Compressive sensing spectroscopy with a single pixel camera. *Appl. Opt.*, 55(19):5198–5202.

Stoffels, J., Bluekens, A., and Petrus, J. (1978). Color splitting prism assembly. US Patent 4,084,180.

Tsai, T.-H., Llull, P., Yuan, X., Carin, L., and Brady, D. J. (2015). Spectral-temporal compressive imaging. *Optics Letters*, 40(17):4054–4057.

Tsaig, D. D. Y., Drori, I., and Starck, J. (2012). Sparse solution of underdetermined systems of

linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 58:1094–1121.

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

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

Wagadarikar, A. A., Pitsianis, N. P., Sun, X., and Brady, D. J. (2009). Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics Express*, 17(8):6368–6388.

Wainwright, M. J. (2009a). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Transactions on Information Theory*, 55(12):5728–5741.

Wainwright, M. J. (2009b). Sharp thresholds for high-dimensional and noisy sparsity recovery using l1-constrained quadratic programming (lasso). *IEEE transactions on information theory*, 55(5):2183–2202.

Wang, L., Xiong, Z., Gao, D., Shi, G., and Wu, a. F. (2015a). High-speed hyperspectral video acquisition with a dual-camera architecture. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4942–4950.

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

Wang, Z. and Arce, G. (2010). Variable density compressed image sampling. *IEEE Transactions on Image Processing*, 19:264–270.

Wohlberg, B. (2016). Efficient algorithms for convolutional sparse representations. *IEEE Trans. Image Process.*, 25(1):301–315.

Wu, Y., Mirza, I. O., Arce, G. R., and Prather, D. W. (2011a). Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Opt. Lett.*, 36(14):2692–2694.

Wu, Y., Mirza, I. O., Arce, G. R., and Prather, D. W. (2011b). Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Optics Letters*, 36(14):2692–2694.

Yang, C., Everitt, J. H., Du, Q., Luo, B., and Chanussot, J. (2013). Using high-resolution airborne and satellite imagery to assess crop growth and yield variability for precision agriculture. *Proceedings of the IEEE*, 101(3):582–592.

Yang, M., de Hoog, F., Fan, Y., and Hu, W. (2016). Adaptive sampling by dictionary learning for hyperspectral imaging. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9):4501–4509.

Yasuma, F., Mitsunaga, T., Iso, D., and Nayar, S. K. (2010). Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253.

Ye, Z., Li, H., Song, Y., Benediktsson, J. A., and Tang, Y. Y. (2017). Hyperspectral image classification using principal components-based smooth ordering and multiple 1-d interpolation. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):1199–1209.

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

Zhao, J., Zhong, Y., Shu, H., and Zhang, L. (2016). High-resolution image classification integrating spectral-spatial-location cues by conditional random fields. *IEEE Transactions on Image Processing*, 25(9):4033–4045.

Zhu, R., Tsai, T.-H., and Brady, D. J. (2013). Coded aperture snapshot spectral imager based on liquid crystal spatial light modulator. In *Frontiers in Optics 2013*, page FW1D.4. Optical Society of America.