Mémoire

Auteur : Thomas, Clément
Promoteur(s) : Georges, Marc
Faculté : Faculté des Sciences
Diplôme : Master en sciences spatiales, à finalité spécialisée
Année académique : 2022-2023
URI/URL : http://hdl.handle.net/2268.2/17460

Exploring Compressive Sensing for Earth Observation

Author:

Clément Thomas

Supervisor:

Dr. Marc Georges

Committee member:

Prof. Serge Habraken

Dr. Murielle Kirkove

Dr. Lionel Clermont

ULiege, Faculty of Sciences

Master in Space Sciences, Professional focus

Academic year 2022-2023

Acknowledgments

First of all, I would like to express my gratitude to my supervisor, Dr. Marc Georges, for providing me with the opportunity to conduct my master's thesis at CSL on this fascinating topic, and for his invaluable guidance throughout the entire project.

I would like to thank Prof. Laurent Jacques (UCLouvain) for his helpful advice on reconstruction processes, Dr. Murielle Kirkove (CSL) for her valuable feedback, and Dr. Valentina Raimondi (Italian National Research Council) for lending the DMD used in the laboratory implementation.

Furthermore, I am grateful to my parents for giving me the opportunity to pursue my studies and for all their support during these five years.

Finally, I would like to acknowledge Julien, Colin, and Tom for their assistance in proofreading this work.

Abstract

This master thesis explores the application of compressive sensing in satellite Earth observation instruments. Firstly, a general state of the art of compressive sensing is made by introducing the mathematical concepts and describing some existing designs that implement the method. The essence of compressive sensing consists in reconstructing images with fewer measurements than in classical imaging. The method can bring drastic reduction of data quantity requirements and detector sizes as well as an increase of spatial resolution. These advantages are particularly interesting in Earth observation instruments considering the vast amount of data that they generate and the size limitations of satellites. This is even more considerable in the infrared spectrum where detectors are typically large.

A deep learning compressive sensing reconstruction algorithm dubbed ISTA-Net$^+$ is tested an proved to work on satellite multispectral data during simulations. Finally, a complete compressive sensing experimental chain has been implemented within laboratory environment. For the reconstruction, the hardware-compressed data could not be passed to the ISTA-Net$^+$ algorithm, thus a simpler algorithm applying an inpainting using iterative hard thresholding is applied. The experiment is satisfactory and the method is proven to work. Nonetheless, the optical system has to be optimized and a more efficient algorithm must be implemented. Therefore, this work opens the way to further improvements and investigations.

Contents

List of Figures

List of Tables

Acronyms

CAI Coded Aperture Imaging.

COC Circle of Confusion.

CS Compressive sensing.

CWT Continuous Wavelet Transform.

DL Deep Learning.

DMD Digital Micromirror Device.

DOF Depth Of Field.

DWT Discrete Wavelet Transform.

FOV Field Of View.

FT Fourier Transform.

GSD Ground Sampling Distance.

iFOV instantaneous Field of View.

IHT Iterative Hard Thresholding.

IR Infrared.

ISTA Iterative Shrinkage-Thresholding Algorithm.

LEO Low Earth orbit.

LWIR Long-Wavelength Infrared.

MIR Mid-Wavelength Infrared.

NDMI Normalized Difference Moisture Index.

NDVI Normalized Difference Vegetation Index.

NEO Near Earth Object.

NIR Near Infrared.

OTF Optical Transfer Function.

PB Push-Broom.

PSNR Peak Signal to Noise Ratio.

ReLU Rectified linear unit.

RIP Restricted Isometry Property.

SLM Spatial Light Modulator.

SNR Signal to Noise Ratio.

SRAM Static random access memory.

SSIM Structural Similarity Index Method.

STFT Short-Time Fourier Transform.

SWIR Short-Wavelength Infrared.

TIR Thermal Infrared.

TV Total Variation.

UV Ultraviolet.

WB Whisk-Broom.

WT Wavelet Transform.

1 Introduction

Imaging systems consist in collecting light from a scene and, facilitated by appropriately designed optics, focusing this light on a detector which will sense (ideally) all photons collected. These detectors are composed of pixels grids, each of these pixels being a semiconductor that generates an electric current upon interaction with photons of a specific range of energy. The size of these pixels along with the distance separating the scene from the detector determine the spatial resolution of the resulting image. In traditional systems, an image is acquired by sensing the scene, and integrate the total quantity of light detected per pixel. If only one wavelength range is detected, the outcome is a grayscale image where the brighter pixels are those with the most light detected. Practically, it means that one value is registered by pixel. However, if several spectral ranges are discriminated, a greater quantity of data is acquired. For instance, a traditional color image differentiates the visible spectrum into three bands: red, green, and blue. Consequently, three values are recorded per pixel and, for a detector of the same size, the quantity of data acquired is tripled with respect to a grayscale image. Thus, the quantity of data to be acquired is directly proportional to both the number of pixels of the detector and the number of spectral bands to be distinguished.

Many practical problems in scientific imaging require the acquisition of vast amounts of data. This is especially true in Earth observation, where a lot of spectral bands are discriminated, sometimes exceeding 200 for hyperspectral imagers. This leads to challenges related to data acquisition, storage, transfer, visualization, and analysis, a phenomena often referred to as "Big Data". While lots of methods are currently under investigation to handle these large data quantities, Compressive Sensing (CS) offers a different perspective by aiming to reduce the quantity of data needed to acquire and display images. The concept relies on the sparsity of natural signals, which allows to reconstruct the complete signal by detecting only its significant components. This sparse characteristic is well-known and used to compress acquired data. CS innovates by directly sensing the compressed data. Essentially, it means that an image may be acquired by measuring less data than in classical imaging systems. A simplified view of the technique is that some of the pixels of the image can be hidden by a mask and the full scene can be reconstructed with the measured ones. Nevertheless, in most cases, several acquisitions are made with different masks, with the sole requirement that the total quantity of data acquired is lower than in traditional imaging. In the most extreme case, a single-pixel is sufficient to reconstruct the images. Consequently, a noteworthy advantage of CS is the reduction in detector sizes. Each step of the CS acquisition method is extensively developed in this work.

In the modern context of satellite-based Earth observation, the demand for related data

keeps on growing due to its significance in many sectors such as meteorology, climatology, agriculture, natural disaster management, or ecosystem studies, among others. This great need of data leads to problems mostly concerning data storage and transmission from the space instrument to the ground segment. Furthermore, the development of Earth observation instruments is limited by the size and mass of detectors. Indeed, larger detectors are more difficult and expensive to manufacture. Therefore, the CS methodology can be implemented to reduce both the data quantity and the size of detectors. Moreover, the fact that the data is directly compressed results in removing the on-board compressors present in many instruments, thus further reducing the weight and size of the payload. These problems are particularly present in the domain of InfraRed (IR) imaging where the detectors tend to be significantly large.

The CS technique is proven both theoretically and in laboratory setup. Nowadays, many practical applications of the method are undergoing development including space applications such as rover cameras, stellar spectroscopy, and Earth observation instruments. This work concentrates on the application of CS in satellite Earth observation, with a particular emphasis on the advantages it offers in the IR spectral range. This thesis is structured into four main sections. First, a general state of the art of the CS method is described with a complete mathematical description of the problem and of some reconstruction algorithms. Additionally, standard CS architectures are also explained. Second, the specific application of CS for Earth observation is covered with a strong attention on the IR spectrum. Then, simulations are made with an existing CS algorithm to prove its functionality with Earth observation data. Finally, a complete chain of CS measurement and reconstruction is developed from scratch in a laboratory environment.

Overall, this thesis aims to contribute to the understanding and practical implementation of CS in the field of satellite Earth observation, particularly focusing on its advantages in the IR domain.

2 State of the art

This first section extensively develops the mathematical origins of the CS methodology in order to understand the roots of the concept. Afterwards, standard CS architectures are described to illustrate the technique. Then, the reconstruction algorithms that are used in this work are described in details. Finally, the concept of super resolution, derived from the CS method is explained. Overall, the objective of this section is to describe the different steps of CS while providing examples of its several advantages.

2.1 Mathematical description

Complete mathematical descriptions of the CS method are shown in Ref. [1] and [2], from which most information of this section come from.
In image and signal processing, the objective is to reconstruct a signal from the measured data. In the case the information acquisition is a linear process, the problem can be seen as solving a system of linear equations. In mathematical terms, the relation between $\boldsymbol{y} \in \mathbf{C}^m$ the measured data and $\boldsymbol{x} \in \mathbf{C}^N$ the original signal can be written

$$\boldsymbol{\Phi x} = \boldsymbol{y}, \tag{1}$$

with $\boldsymbol{\Phi} \in \mathbf{C}^{m \times N}$ the matrix representing the linear measurement process. Classical algebra states that m must be at least as large as N, otherwise the system is undetermined and there is an infinite number of solutions (provided that at least one exists). In other words, the amount of measured data must be at least as large as the signal length for the signal to be reconstructed. This can also be related to the Shannon-Nyquist sampling theorem which states that a signal must be sampled at a frequency at least twice as high as the highest frequency present in the signal to accurately acquire the entire information.
However, the problem is slightly different when sparsity is considered. A signal is called sparse if most of its components are zero. In fact, it appears that most of real-world signals are sparse, or at least can be well approximated by sparse signals. Actually, this sparsity often appears only after an appropriate change of basis. This property is well known and is used for file compression, notably for the JPEG, MPEG and MP3 formats. CS is an innovative technique which tackles the problem the other way around and tries to acquire directly the compressed data and to reconstruct the full signal from this reduced information. Mathematically, it means that the signal of size N can be reconstructed from m measurements with $m < N$, given that this signal is sparse in some domain. This corresponds to an ill-posed inverse problem. The CS ratio is then defined by $\frac{m}{N}$, the ratio between the number of acquired data and the signal length. The CS method aims to find balance between a CS ratio as low as possible and a maximum reconstruction accuracy. Hence, the standard problem of CS consists in the reconstruction of a sparse vector $\boldsymbol{x}$

from undetermined measurements $\boldsymbol{y} = \boldsymbol{\Phi x}$. This leads to two main questions:

1. What matrices $\boldsymbol{\Phi}$ are suitable? In other words, how should be designed the linear measurement process?

2. How can $\boldsymbol{x}$ be reconstructed from $\boldsymbol{y}$? What are efficient algorithms?

As mentioned previously, the sparse nature of the signal in some basis is crucial to the reconstruction process. In this context, it is important to introduce the notion of k-sparse vector. A vector $\boldsymbol{x}$ is called k-sparse if it has at most k nonzero coefficients, or, using the l_0-norm, if $||\boldsymbol{x}||_0 \leq k$. To represent this sparsification, the signal vector $\boldsymbol{x}$ can be written $\boldsymbol{\Psi s}$, with $\boldsymbol{\Psi}$ the $N \times N$ basis matrix, and $\boldsymbol{s}$ the K-sparse coefficient vector (with at most $K \ll N$ nonzero coefficients). Therefore, the initial problem may be rewritten

$$\boldsymbol{y} = \boldsymbol{\Phi \Psi s} + \boldsymbol{e}, \tag{2}$$

where the noise $\boldsymbol{e}$ is introduced, including the sensor noise and the model loss; or, considering a noiseless model and defining $\boldsymbol{\Theta} = \boldsymbol{\Phi \Psi}$,

$$\boldsymbol{y} = \boldsymbol{\Theta s}. \tag{3}$$

This measurement process is represented in Fig. 1.

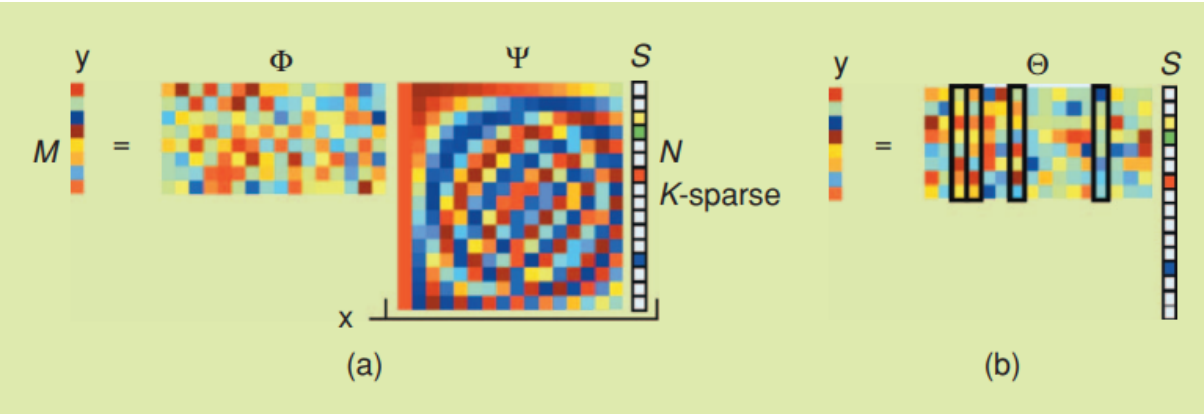


Figure 1: (a) CS measurement process with a random Gaussian measurement matrix $\boldsymbol{\Phi}$ and discrete cosine transform (DCT) matrix $\boldsymbol{\Psi}$. The vector of coefficients $\boldsymbol{s}$ is sparse with $K = 4$. (b) Measurement process with $\boldsymbol{\Theta} = \boldsymbol{\Phi \Psi}$. There are four columns that correspond to nonzero s_i coefficients; the measurement vector $\boldsymbol{y}$ is a linear combination of these columns. From [3].

The main challenge of CS relies in finding the locations of the nonzero entries of $\boldsymbol{x}$. If they were known, reducing the matrix $\boldsymbol{\Phi}$ to the columns indexed by these locations would be sufficient. However, not knowing them introduces some non-linearity in the reconstruction method. A first method to solve this problem would be to search for the

sparsest vector $\boldsymbol{x}$ that respects Eq. (1). If the noise is not taken into account, it is equal to solving the l_0 minimization problem

$$\min ||\boldsymbol{z}||_0 \qquad \text{subject to } \boldsymbol{\Phi z} = \boldsymbol{y}, \tag{4}$$

with $\boldsymbol{z}$ the best k-sparse approximation of $\boldsymbol{x}$. Nevertheless, in practice there exists no efficient algorithm that can solve this general problem for any $\boldsymbol{\Phi}$ and any $\boldsymbol{y}$. Consequently, one turns to a l_1 minimization approach which considers the solution of

$$\min ||\boldsymbol{z}||_1 \qquad \text{subject to } \boldsymbol{\Phi z} = \boldsymbol{y}. \tag{5}$$

Now, the question of which matrices $\boldsymbol{\Phi}$ allow an efficient reconstruction using an appropriate algorithm must still be answered. One widely used property to determine these matrices is the Restricted Isometry Property (RIP) which is defined as follows:

The restricted isometry constant δ_k of a matrix $\boldsymbol{\Phi} \in \mathbf{C}^{m \times N}$ is the smallest number such that

$$(1 - \delta_k)||\boldsymbol{z}||_2^2 \leq ||\boldsymbol{\Phi z}||_2^2 \leq (1 + \delta_k)||\boldsymbol{z}||_2^2 \tag{6}$$

for all sparse vectors $\boldsymbol{z}$ having no more than k nonzero entries. A matrix $\boldsymbol{\Phi}$ is said to satisfy the restricted isometry property of order k with constant δ_k if $\delta_k \in [0, 1]$.
A related condition called incoherence states that the rows of $\boldsymbol{\Phi}$ cannot sparsely represent the columns of $\boldsymbol{\Psi}$. Hence, the matrices $\boldsymbol{\Phi}$ must satisfy both of these conditions. Verifying these conditions in practice seems complicated. However, it appears that both can be achieved by simply selecting $\boldsymbol{\Phi}$ as a random matrix [3].

2.2 Compressive sensing imaging devices

Several CS imaging devices have been built and tested in the laboratory. The development of such systems that exploit CS theory is a considerable challenge since there are many considerations such as size, cost, reconstruction accuracy, and reconstruction speed. The imaging system characteristics also depend on the nature of the target signal, one can easily understand that the measurement method for IR or ultraviolet (UV) light will not be exactly identical, or at least will not have the same accuracy [4]. Besides the reduction of the required data measurement quantity, another possible outcome of CS imaging devices is to use the CS properties to minimize the number of optical pieces in the system. In some cases, no lenses are used at all. This allows to build cameras of smaller size and lower mass. In this section are presented the most popular proposed architectures for CS imagers. But first, Spatial Light Modulators (SLM), used for masking the images in a CS system, are introduced.

2.2.1 Spatial light modulators

In any case, CS requires masking the image. In practice, the mask is implemented with a SLM. By definition, it is an instrument that modulates the intensity of a light beam according to a control signal. SLMs exist under two main forms: transmissive or reflective. Transmissive SLMs are designed such that they either pass, block, or attenuate parts of the incoming beam. The most used are based on Liquid Crystal Display (LCD) in which the light modulation occurs through the different absorption and refraction properties of liquid crystals arranged in matrix form. However, due to manufacturing constraints and physical properties of the liquid crystals, the use of LCD is limited in the visible and Near IR (NIR) spectral range [5]. On the other hand, reflective SLMs reflect part of the light, instead of transmitting it. The most popular type of reflective SLM, and that will also be used in the practical part of this work, is a Digital Micromirror Device (DMD), sometimes also called digital micromirror arrays. The DMD is composed of an array of micromirrors. Each micromirror is connected to an individual Static Random Access Memory (SRAM) , is able to rotate, and can be set in one of two states depending on the bit value of the SRAM. The two states are similar to simple 'on' and 'off' states. In the 'on' state, the mirror is positioned so that it reflects the light towards the detector, and in the 'off' mode it reflects the light away from it, ideally towards an absorbing surface to avoid stray light from back scattering. This way, different masks can be applied at a relatively high rate with an appropriate coding, allowing a high-speed acquisition of the required measurements.

2.2.2 Single-pixel camera

The single-pixel camera is probably the most well-known CS imaging system. It was developed by Duarte et al.[6] at Rice University. The method consists in applying some mask that will hide part of the signal produced by the target and then measure the resulting signal in one single pixel. This means that after passing by the mask, the resulting light will be entirely integrated and the measurement will represent the total intensity received on the detector. In other words, random parts (pixels) of the scene being imaged are hidden and the detector measures the total intensity of the non-masked pixels. By applying random masks and taking m measurements (with $m < N$ the signal length), it is possible to use CS image reconstruction methods from these measurements. In practice, the single-pixel camera is designed to apply masks with a DMD. A schematic view of a single-pixel camera is shown in Fig. 2.

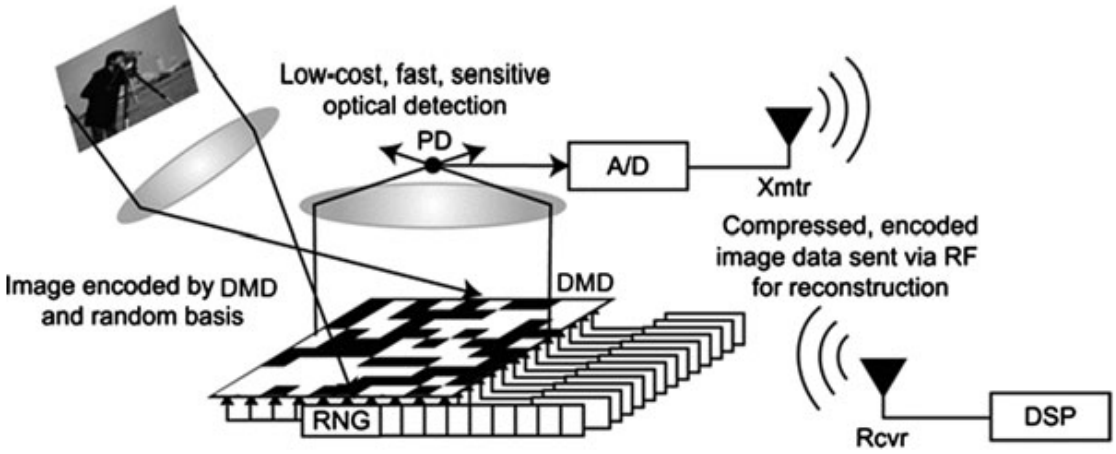


Figure 2: Schematic representation of a single-pixel camera. From [1]

2.2.3 Coded aperture imagers

In Coded Aperture Imaging (CAI), wavefronts from a scene are encoded by a mask and then propagate to a 2-D detector array. It has been shown that if the coded masks are designed using a pseudorandom construction, then the resulting sensing matrix satisfies the Restricted Isometry Property (RIP). Afterwards, the knowledge of the coded mask allows to decode the detected 2-D intensity pattern using deconvolution to retrieve an image. There also exists adaptive CAI methods which incorporate a reconfigurable mask (a SLM), providing higher Signal to Noise Ratio (SNR) and spatial resolution, as well as agile imaging modes. An example of applied CAI is the FlatCam: by placing the mask extremely close to the detector, it is possible to design a thin, lensless imaging system [7]. Nowadays, CAI is mostly used at X-ray and γ-ray wavelengths where classical optics such as lenses and mirrors cannot be used. Indeed, these wavelength are so short and the radiation so energetic, that the interaction occurs at the molecular level, preventing an efficient use of lenses and mirrors. Working with CAI at longer wavelengths, such as thermal IR is more challenging due to several reasons: the existence of a viable SLM technology at these wavelengths, a poor SNR, and the lack of optical flux concentration. [4] [8]

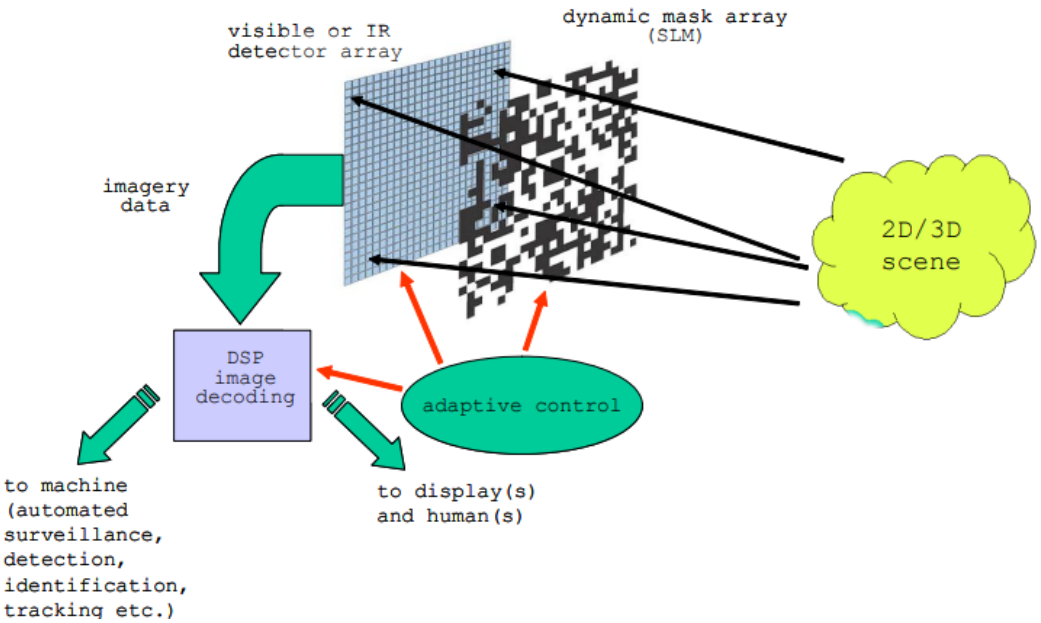


Figure 3: Illustration of the basic adaptive CAI approach. From [8].

2.2.4 Spectral imagers

A classical spectrometer uses diffractive optics such as prisms and gratings to measure the global spectrum of one target, generally point-like (e.g. a star), in some given spectral range. Detectors of such instruments are generally set so that each column of pixel represents one wavelength and only the rows give some spatial indication. However, in imaging spectroscopy each pixel contains information in the different spectral bands and represents one corresponding scene element. As a result, data is stored and represented in the form of a 3-D datacube. This involves a large amount of data to be sensed, especially for hyperspectral imagers which can sense in sometimes more than 200 spectral bands. Hence, the use of a CS method could provide a more efficient way to sense these images by measuring less data. One major challenge in image spectroscopy is to perform a trade-off between spatial resolution, spectral resolution, light collection, and acquisition time. Gehm et al.[9] and Arce et al.[10] describe the principles and design for a CS spectral imager. This system attempts to fully decouple the four operational quantities and is represented on Fig.4.

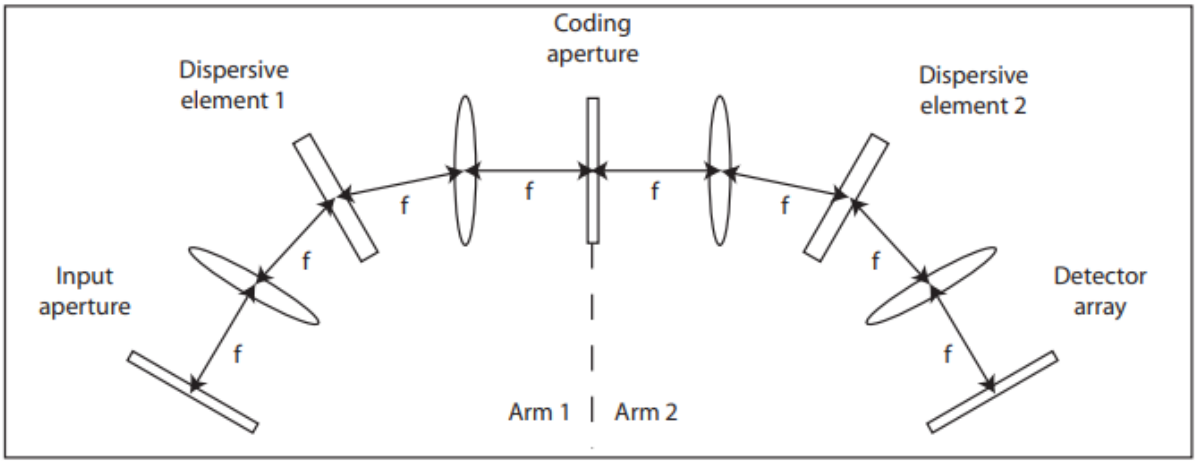


Figure 4: Schematic of the spectral imager. From [9].

The signal enters through the input aperture in the first arm. This first arm contains a dispersive element and ends with the coded aperture which separates the two arms. Due to the dispersive element, multiple images are formed in wavelength-dependent locations. When arriving at the coded aperture, the spatial structure in the plane contains a mixture of spatial and spectral information about the source. The coded aperture modulates this information with the applied pattern. Then, in the second arm, the spatial-spectral mixing of arm 1 is undone and the spatial modulation introduced by the coded aperture is transformed into a spatial and spectral modulation. Finally, an image of the source is formed on the detector.

2.3 Reconstruction algorithms

Obviously, having an efficient reconstruction method is extremely important when imaging via CS. The investigation of different algorithms was the scope of a previous master thesis done by S. Gramegna at CSL [11]. The different algorithms tested were based either on total variation (TV) methods or on deep learning (DL) methods. The outcome was that the DL methods were better in terms of quality and computation time. Therefore, this work primary focus on a DL algorithm dubbed ISTA-Net$^+$. Nevertheless, for reasons that will be detailed later, this algorithm could not be used during the laboratory experiment. Thus, an inpainting algorithm using Iterative Hard Thresholding (IHT) has been used. Both methods are detailed in the following sections. As mentioned earlier, the sparsity of the signal is a key element in CS. Hence, the images measured first have to be transformed into a sparse domain to be reconstructed. This sparsification can be done by TV, wavelet transform, or Fourier transform (FT). These methods are thus introduced before the algorithms are detailed.

2.3.1 Total Variation

If a fixed tolerance ε for taking the noise of the forward model into account is introduced in the original l_1 minimization problem defined in Eq. (5), the expression becomes

$$\min ||\boldsymbol{z}||_1 \qquad \text{subject to } ||\boldsymbol{\Phi z} - \boldsymbol{y}||_2 \leq \varepsilon. \tag{7}$$

Whereas Eq. (4), (5), and (7) consist in reversing the problem with a regularization which is the minimization of the l_0-norm of $\boldsymbol{z}$ for Eq. (4) and of the l_1-norm of $\boldsymbol{z}$ for Eq. (5) and (7), the TV method uses the l_1-norm of the discrete gradient of the images instead of the signal itself. The reason being that it turns out that natural images often admit very sparse approximations in the gradient domain [12]. Indeed, homogeneous zones in images, which are often large, have a zero value in the gradient domain. A visual example is given in Fig. 5.

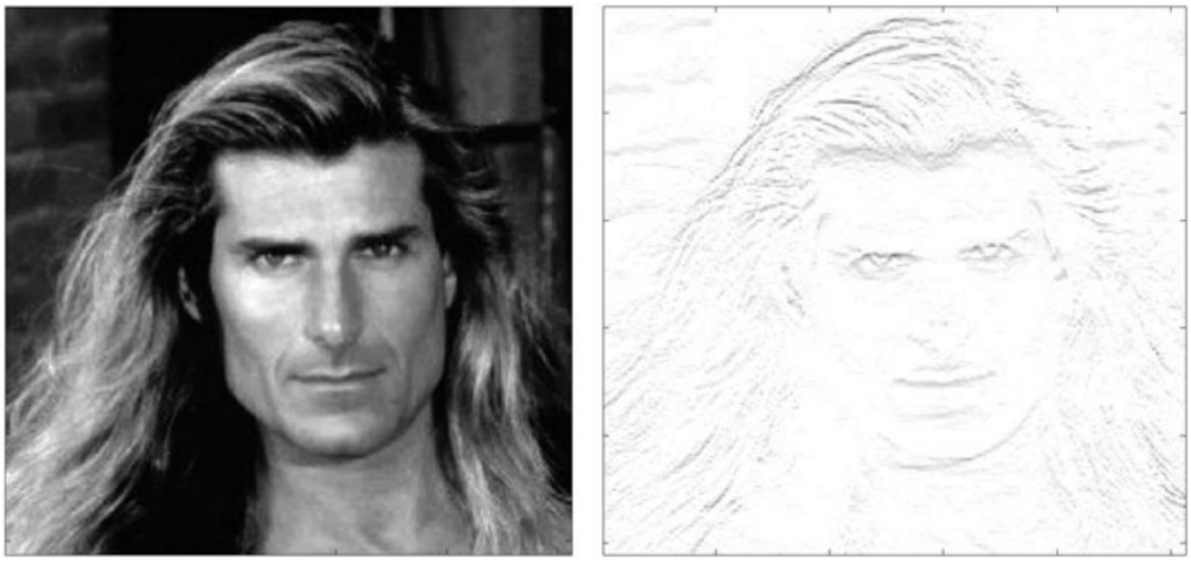


Figure 5: The original Fabio image (left) and the absolute values after application of a discrete gradient operator (right). From [12]

Stating that $z_{m,n}$ is the value of $\boldsymbol{z}$ in the pixel at position (m, n), the discrete gradients $D_1 z_{m,n}$ and $D_2 z_{m,n}$ are defined as

$$D_1 z_{m,n} = z_{m+1,n} - z_{m,n}, \tag{8}$$

$$D_2 z_{m,n} = z_{m,n+1} - z_{m,n}. \tag{9}$$

Then, the TV is defined by

$$||\boldsymbol{z}||_{TV} = \sum \sqrt{(D_1 z)^2 + (D_2 z)^2} = \sum |(\nabla z)_{m,n}| \tag{10}$$

and the problem can be rewritten

$$\min ||\boldsymbol{z}||_{\mathrm{TV}} \qquad \text{such that } ||\boldsymbol{\Phi z} - \boldsymbol{y}||_2 \leq \varepsilon. \tag{11}$$

There are two types of total variation algorithms to reconstruct compressive sensed images: by nonlocal regularization and based on the augmented lagrangian. As mentioned earlier, these algorithms are already tested and proven less efficient than the DL methods. Therefore, they will not be developed in this work.

2.3.2 Wavelet transform

Wavelet transforms can be seen as be seen as a complement to classical Fourier decomposition method. As a reminder, a FT is defined as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt, \tag{12}$$

where ω is the frequency, t the time, and f the signal. The FT is used to transform a signal from the time domain to the frequency domain. This means the resulting function will present information about which frequencies are present in the signal but no information about the time at which these frequencies are present. In order to acquire both frequency and time information, the short-time FT (STFT) is introduced as

$$STFT(\tau, \omega) = \int_{-\infty}^{\infty} s(t) w(t - \tau) e^{-i\omega t} dt, \tag{13}$$

which is the FT of the signal $s(t)$ previously windowed by the function $w(t)$ around time τ. This may be viewed as applying FT on subsamples (or windows) of the signal, providing frequency information for different time ranges in the signal. Thus, information is acquired about both frequency and time. However, the resulting frequency resolution will be reduced with respect to the classical FT. Indeed, taking a shorter time window means a lesser number of samples used in the FT. With fewer discrete frequency intervals, it is significantly more difficult for the transform to discriminate the different frequencies. In other words, a shorter time window means higher time resolution but lower frequency resolution, and inversely. The relationship between the time and frequency resolution is described, as in many physical problems, by the uncertainty principle

$$\Delta t \Delta \omega \geq \frac{1}{4\pi}, \tag{14}$$

with Δt the time resolution and $\Delta \omega$ the frequency resolution. This means that both resolutions cannot be made arbitrarily small.
The STFT describes the decomposition of $s(t)$ in the windowed basis function $w(t-\tau)e^{-i\omega t}$.

However, these basis functions can vary and a general form of the STFT equation can be written in terms of basis function $k_{\tau,\omega}$ and signal $s(t)$ as an inner product

$$STFT(\tau,\omega) = \int_{-\infty}^{\infty} s(t) k_{\tau,\omega} dt. \tag{15}$$

The Wavelet Transform (WT) is then defined in the same manner. Generally, the frequency ω is replaced by the scale variable $a = 1/\omega$, the time-shift τ is represented by b, and the basis function for WTs is

$$k_{b,a}(t) = \frac{1}{\sqrt{a}} \gamma \left(\frac{t-b}{a} \right) \tag{16}$$

and the continuous wavelet transform (CWT) is defined by

$$CWT(b,a) = \frac{1}{\sqrt{a}} \int \gamma \left(\frac{t-b}{a} \right) s(t) dt, \tag{17}$$

where $\gamma(t)$ is the wavelet. By definition a wavelet is a small wave, or more precisely an oscillation that decays quickly. Typical transform basis functions are shown in Fig. 6. [13]

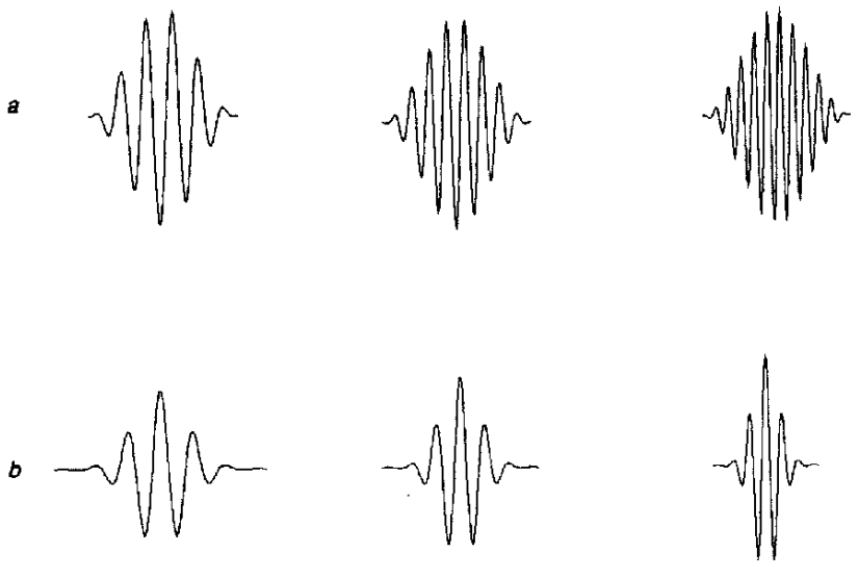


Figure 6: Real part of typical transform basis functions: (a) STFT, (b) WT. From [13].

Discrete Wavelet Transforms (DWT) can be computed by sampling the WT at discrete values of a and b, it thus have the form

$$DWT(b,a) = \frac{1}{a} \sum_{m=0}^{p-1} s(t_m) \gamma \left(\frac{t_m - b}{a} \right). \tag{18}$$

On one hand, high frequency values (low a) will have significant corresponding wavelet coefficients, called detail coefficients. On the other hand, low frequency values (high a) will

have smaller corresponding coefficients, called approximation coefficients. In the end, this enables to discriminate the different frequencies present at different times in a signal. In image processing, the time dimension is replaced by spatial dimensions and it is possible to perform a 2-dimension WT. In this case, the WT will yield few significant coefficients around discontinuities, whereas the FT would yield many significant coefficients, thus achieving a better sparsity [14].

2.3.3 Inpainting using IHT

Inpaiting consists in filling holes in images, thus corresponding to a ill-posed inverse problem. Considering the measurements $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$ with $\boldsymbol{\Phi}$ the masking operator, $\boldsymbol{x}$ the complete image, and $\boldsymbol{e}$ the noise. This algorithm uses sparsity in the form of a sparse basis $\boldsymbol{\Psi}$ to reconstruct the images. In fact, here $\boldsymbol{\Psi}$ is the matrix of wavelet transform defined in the previous section. Hence, the regularization computes a sparse set of coefficients $(s_m^*)_m$ in a frame $\boldsymbol{\Psi} = (\psi_m)_m$ that solves

$$\boldsymbol{s}^* \in \arg\min_{\boldsymbol{s}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{\Phi\Psi s}\|^2 + \lambda J(\boldsymbol{s}), \tag{19}$$

where λ is the parameter that will be used as a the hard threshold and that should be adapted to the noise level. Its value is decayed during the iteration process. During this work, damaged observations will be considered but the original signal is not noisy. Hence, in this case, the signal is considered noiseless, i.e. $\boldsymbol{e} = 0$. $J(\boldsymbol{s})$ is the l_1 sparsity prior

$$J(\boldsymbol{s}) = \|s_m\|_1. \tag{20}$$

Therefore, $\boldsymbol{s}^*$ is the wavelet coefficient vector of the sparse signal to reconstruct ($\boldsymbol{\Psi s}$). Thus, when this regularization is solved, the corresponding inverse DWT is applied on s^* to retrieve the reconstructed image.
In practice, this algorithm combines a gradient descent and a hard thresholding using a decaying threshold. The hard thresholding h_T with threshold T is defined as

$$h_T(x) = \begin{cases} x & \text{if } x > T \\ 0 & \text{if } x < T \end{cases} \tag{21}$$

thus defining a thresholding operator $H_T(\boldsymbol{f}) = h_T(\boldsymbol{s_f})$, where $\boldsymbol{s_f}$ is the wavelet coefficient vector of the DWT of the image$\boldsymbol{f}$. The algorithm computes a series of images $\boldsymbol{f}^{(l)}$ as

$$\boldsymbol{f}^{(l+1)} = H_{\lambda_l}(\boldsymbol{f}^{(l)} - (\boldsymbol{\Phi f}^{(l)} - \boldsymbol{y})), \tag{22}$$

where $\boldsymbol{f}^{(l)} - (\boldsymbol{\Phi f}^{(l)} - \boldsymbol{y})$ is the gradient descent step. The initial image $\boldsymbol{f}^{(0)}$ can be set as the inverse DWT of the wavelet coefficients of $\boldsymbol{y}$ thresholded by λ_0, i.e. the highest

threshold value that will be used during the iterations. By using decaying threshold values λ_l for a number iteration defined a priori, the algorithm can converge to a solution [15].

2.3.4 Iterative Shrinkage-Thresholding Algorithm-Net

A complete algorithm using a DL image reconstruction method, called Iterative Shrinkage-Thresholding Algorithm-Net (ISTA-Net), has been developed and implemented in Python by Zhang and Ghanem [16]. The architecture is based on the traditional ISTA which optimizes a general l_1 norm CS reconstruction model and introduces it into a deep neural network. Particularly, the ISTA-Net algorithm is composed of a fixed number of ISTA-like iterations.

The traditional ISTA algorithm aims to solve the following reconstruction problem

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{\Phi x} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\Psi x}\|_1 \tag{23}$$

by implementing the following iterative steps. First the gradient-descent method

$$\boldsymbol{r}^{(k)} = \boldsymbol{x}^{(k-1)} - \rho\boldsymbol{\Phi}^T(\boldsymbol{\Phi x}^{(k-1)} - \boldsymbol{y}) \tag{24}$$

and second, the regularization by l_1-norm minimization, which is equal to a soft thresholding with threshold λ

$$\boldsymbol{x}^{(k)} = \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{r}^{(k)}||_2^2 + \lambda||\boldsymbol{\Psi x}||_1, \tag{25}$$

where k is the ISTA iteration index, ρ the step size, $\boldsymbol{\Phi}$ the measurement matrix, $\boldsymbol{\Psi}$ the matrix that transforms $\boldsymbol{x}$ to a sparse domain, and λ the regularization parameter (generally pre-defined). While it is relatively easy to solve Eq. (25) when $\boldsymbol{\Psi}$ is orthogonal, it becomes more complex for a non-orthogonal, or even nonlinear transform. Another challenge is that the optimal transform $\boldsymbol{\Psi}$, the step size ρ and the parameter λ are defined a priori and are difficult to tune. Moreover, ISTA generally requires a lot of iterations to yield a satisfying result, giving rise to a higher computation time. Nevertheless, ISTA is very well suited for solving large-scale linear inverse problems such as the CS reconstruction.

ISTA-Net takes advantage of all ISTA merits while solving the problem of parameters optimization. The idea is to build a deep network architecture composed of a fixed number of phases, each phase corresponding to one ISTA iteration. A representation of the algorithm is shown in Fig. 7.

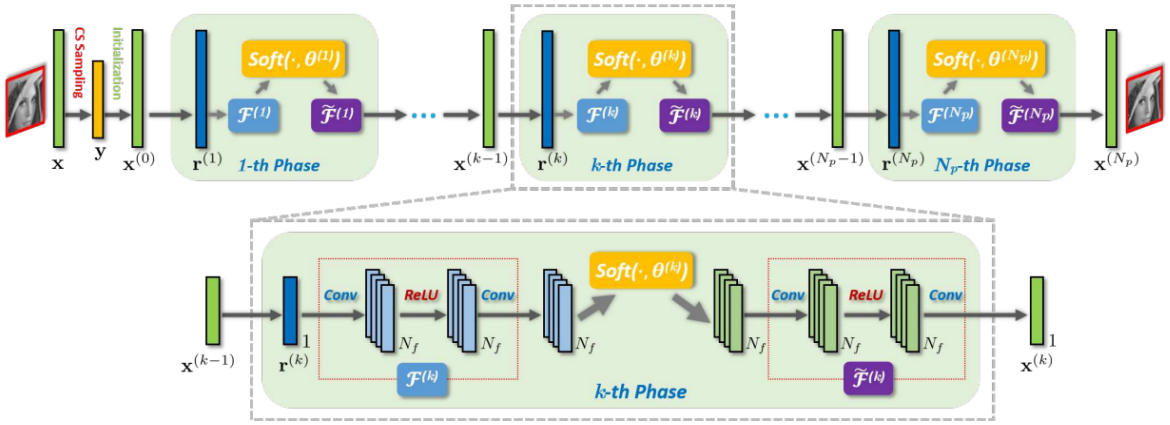


Figure 7: Illustration of the ISTA-Net framework. From [16].

Whereas ISTA uses a hand-crafted transform $\boldsymbol{\Psi}$, ISTA-Net employs a general nonlinear transform function $\mathcal{F}(\cdot)$, whose parameters are learnable, to sparsify natural images. It is composed of two linear convolutional operators separated by one rectified linear unit (ReLU). Each convolutional operator corresponds to N_f filters. Thus, in matrix form: $\mathcal{F}(\boldsymbol{x}) = \boldsymbol{B}ReLU(\boldsymbol{A}\boldsymbol{x})$ with $\boldsymbol{A}$ and $\boldsymbol{B}$ the two convolutional operators mentioned above. Replacing $\boldsymbol{\Psi}$ in Eq.(25) with $\mathcal{F}(\cdot)$, it becomes

$$\boldsymbol{x}^{(k)} = \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{r}^{(k)}||_2^2 + \lambda||\mathcal{F}(\boldsymbol{x})||_1. \tag{26}$$

Now, admitting that in the k-th ISTA iteration, Eq.(24) and (26) both have efficient solutions, one can cast them into two different modules.

1. The $\boldsymbol{r}^{(k)}$ module used to generate the immediate reconstruction result $\boldsymbol{r}^{(k)}$. A change with respect to traditional ISTA is that the step size is variable to increase the network flexibility. Therefore, the final output of this module is

$$\boldsymbol{r}^{(k)} = \boldsymbol{x}^{(k-1)} - \rho^{(k)}\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{x}^{(k-1)} - \boldsymbol{y}). \tag{27}$$

2. The $\boldsymbol{x}^{(k)}$ module which uses the first module as input in Eq. (26) to compute $\boldsymbol{x}^{(k)}$. Supposing that $\boldsymbol{r}^{(k)}$ and $\mathcal{F}(\boldsymbol{r}^{(k)})$ are the mean values of $\boldsymbol{x}$ and $\mathcal{F}(\boldsymbol{x})$ respectively, one can write[1]

$$||\mathcal{F}(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{r}^{(k)})||_2^2 \approx \alpha||\boldsymbol{x} - \boldsymbol{r}^{(k)}||_2^2, \tag{28}$$

with α a scalar related to the parameters of $\mathcal{F}(\cdot)$. Injecting this expression into Eq. (26) yields

$$\boldsymbol{x}^{(k)} = \arg\min_{\boldsymbol{x}} \frac{1}{2}||\mathcal{F}(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{r}^{(k)})||_2^2 + \theta||\mathcal{F}(\boldsymbol{x})||_1, \tag{29}$$

[1]Please refer to [16] for a more complete mathematical description of this step which goes beyond the scope of this work.

where $\theta = \lambda\alpha$.

In addition, the left inverse of $\mathcal{F}(\cdot)$ is introduced with the notation $\tilde{\mathcal{F}}(\cdot)$ such that $\tilde{\mathcal{F}} \circ \mathcal{F} = \mathcal{I}$, the identity operator. Since both functions are learnable, incorporating this relation in the loss function during network training imposes a symmetry constraint allowing $\boldsymbol{x}^{(k)}$ to be computed more efficiently. More precisely, solvin Eq. (29) allows to find $\mathcal{F}(\boldsymbol{x}^{(k)})$ which corresponds to a soft thresholding of $\mathcal{F}(\boldsymbol{r}^{(k)})$ with threshold θ

$$\mathcal{F}(\boldsymbol{x}^{(k)}) = soft(\mathcal{F}(\boldsymbol{r}^{(k)}), \theta), \tag{30}$$

where $soft(\cdot, \cdot)$ is the soft thresholding operator. Then, $\boldsymbol{x}^{(k)}$ can be efficiently computed by applying the inverse operator

$$\boldsymbol{x}^{(k)} = \tilde{\mathcal{F}}(soft(\mathcal{F}(\boldsymbol{r}^{(k)}), \theta)). \tag{31}$$

Finally, it must be noted that $\mathcal{F}(\cdot)$, $\tilde{\mathcal{F}}(\cdot)$, and θ are not constraint to be the same at each phase. Therefore the output of this module should be adapted as

$$\boldsymbol{x}^{(k)} = \tilde{\mathcal{F}}^{(k)}(soft(\mathcal{F}^{(k)}(\boldsymbol{r}^{(k)}), \theta^{(k)})). \tag{32}$$

To summarize, the learnable parameter set is composed of the step size $\rho^{(k)}$, the parameters of the forward and backward transforms $\mathcal{F}^{(k)}(\cdot)$, $\tilde{\mathcal{F}}^{(k)}(\cdot)$, and the shrinkage threshold $\theta^{(k)}$. They are learned as neural network parameters.

An enhanced version of ISTA-Net, dubbed ISTA-Net$^+$ has been created to further improve the reconstruction accuracy. Without going into details, it considers some missing high-frequency component in $\boldsymbol{r}^{(k)}$ and some noise in the expression of $\boldsymbol{x}^{(k)}$. This leads to the introduction of more learnable linear operators, increasing accuracy but also computation time. One phase of ISTA-Net$^+$ is represented in Fig.8.

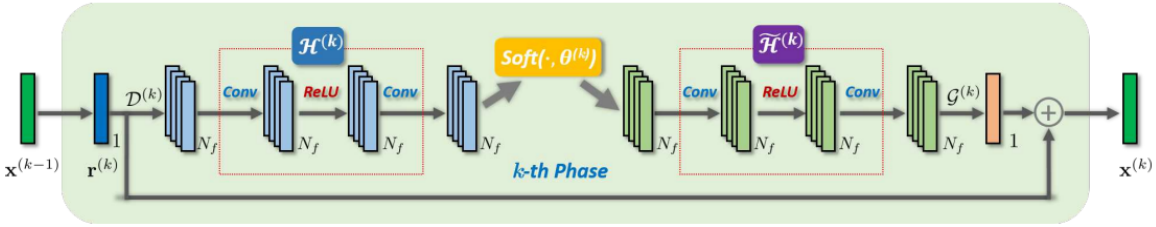


Figure 8: Illustration of the k-th phase of the proposed ISTA-Net$^+$. $\mathcal{D}^{(k)}, \mathcal{G}^{(k)}, \mathcal{H}^{(k)}, \tilde{\mathcal{H}}^{(k)}$ are learnable linear convolutional operators. From [16].

ISTA-Net$^+$ is freely available on *Github* and is the algorithm that will be used in the practical parts of this work to reconstruct the images. It is important to mention that

the code is built to preprocess the data by simulating the CS acquisition technique. This allows to assess the efficiency of the algorithm on different images without having to do the CS hardware acquisition. It does that by applying sets of masks on the input images. Each mask can modulate the signal by applying a matrix product between the image and the mask. The resulting signal then goes through the pre-trained model to be reconstructed.

More recently, You, Xie, and Zhang [17] have developed and implemented an even more improved version of the algorithm named ISTA-Net^{++}, with superior performance and flexibility. Another advantage is that it can handle CS problems with different ratios (multi-ratio tasks) through a single model. Unfortunately, this version of the algorithm is not completely available yet.

2.4 Super resolution

Most imaging applications aim to have the highest image resolution possible since a higher resolution means more details and thus more information. The most evident manner to increase resolution is to reduce the size of pixels, consequently increasing their number per unit of length. This has limitations since the smaller the pixel, the less light it receives, and the SNR decreases, which at some point may severely degrade the image quality. One promising approach to resolve this is the so-called super resolution image reconstruction, sometimes also referred to as resolution enhancement in the literature. This method provides a cheap and efficient solution that can be applied to most existing imaging systems. The basic idea of super resolution is to obtain a final image with a higher resolution than the sensor resolution. This is done thanks to a SLM with a number of elements $N \times N$ times greater than the number of elements of the detector. That way, each pixel of the detector is associated with a different group of $N \times N$ SLM elements. Then, by taking several measurements where a different binary coding mask is applied on the SLM and using appropriate reconstruction methods, it is possible to obtain a final image of the same resolution as the SLM. Therefore, the amount of data acquired is inferior to classical imaging methods (compressed) if the number of carried out modulations is lower than $N \times N$. In this case, one can talk about CS imaging and the super resolution factor is defined as the ratio between the number of SLM micropixels and the number of detectors

$$SR_{\text{factor}} = \frac{N \times N}{\#\text{ of detectors}}. \tag{33}$$

This can be seen as an extension of the single-pixel camera concept to a multi-pixel camera. [18] [19]

3 Compressive sensing in the context of Earth observation

3.1 Earth Observation

As a major part of this thesis investigates the application of the CS method to Earth observation satellites, it is necessary to understand the working principles of such instruments.

Earth observation from space began in 1959 when NASA's probe Explorer 6 took the first image of the Earth from space. Back then, a black and white imager was used since it was the only available resource to make space observations. But since then, technologies have evolved a lot with the development of color photography, multispectral imagers, and more recently hyperspectral imagers. Sensors designed for Earth observation now exist in spectral ranges from the blue to the microwave (radar). Over the years, Earth remote sensing has proven to be an extremely valuable source of information in various domains such as meteorology, climatology, agriculture, geology, geography, hydrology, study of the atmosphere and of ecosystems, to cite a few.

3.1.1 Nature of the signal

To observe the Earth from space, it is crucial to have knowledge about the nature of the observed signal if one wants to understand the data. There exists two ways of sensing the Earth remotely: active and passive systems. Active systems send their own signal and measure what is reflected by the Earth. The geometry and proportion of reflected light at different wavelengths can be used to retrieve information about the target such as its size, shape, and composition. These systems are mostly used in radar technology to observe in the microwave domain. Passive systems, on the other hand, do not send any signal but measure natural signals that can be either the reflection of the Sun emitted light or the radiation directly produced by the Earth (in the thermal IR domain). Most instruments working in the visible and IR domains are passive systems. To properly analyze their data and compute the reflectance of the target, it is thus essential to know the spectrum of the emitted light. Indeed, the quantity of radiation emitted by the Sun and by the Earth are not equivalent at every wavelength. It is described by Planck's law of black body radiation

$$L(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1}, \tag{34}$$

with L the spectral radiance, h Planck's constant, c the speed of light, λ the wavelength, T the temperature, and k Boltzmann's constant. Therefore, the quantity of emitted light is directly wavelength and temperature dependent. The radiation spectrum of the Sun

and the Earth are represented in Fig.9.

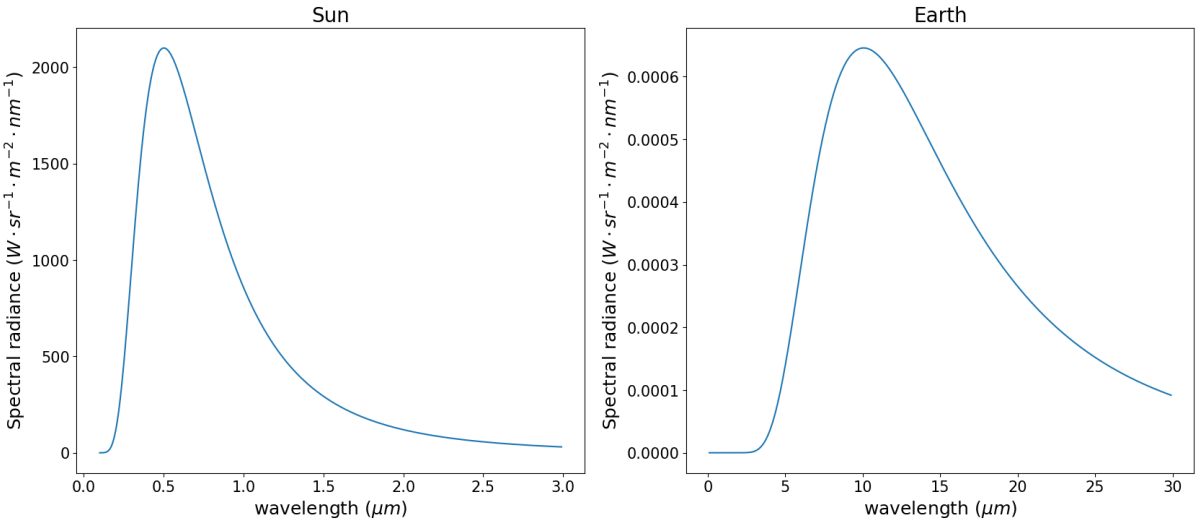


Figure 9: Radiation spectrum of the Sun and the Earth following Planck's law. The mean surface temperatures used are 5778K and 288K respectively.

By looking at the differences of the two axes between the Sun's and the Earth's spectrum, it is obvious that the peak occurs at a longer wavelength and a lower intensity for the Earth, due to the temperature difference. This peak actually occurs in the region of the thermal IR which allows to passively sense the Earth in this spectrum. These spectra can be used to normalize the flux detected by the sensor and retrieve the true reflectance of the target. However, it must still be corrected for atmospheric effects beforehand. Indeed, the different components of the atmosphere will absorb or scatter part of the light received from the Sun, thus degrading the original signal. For example, the stratospheric ozone absorbs a large part of the UV light. While this is of primary importance for life development and survival on the surface, it makes it impossible to observe the Earth surface in the UV domain. This absorption occurs at various wavelengths with different molecules such has water, carbon dioxide, or methane. This leads to atmospheric windows in the optical spectrum where some wavelengths go through, some are partially absorbed, and others totally absorbed. This is shown in Fig.10.

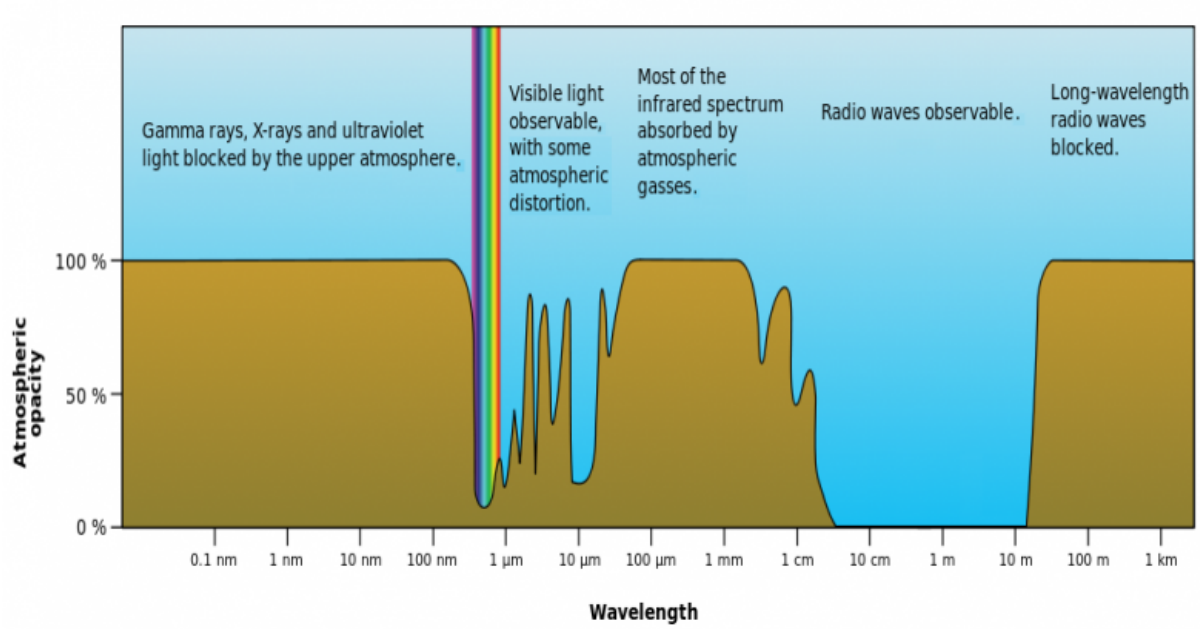


Figure 10: Atmospheric absorption as a function of wavelength. From [20].

These atmospheric effects can be used to directly sense the atmosphere and its composition. However, if the target is on the ground, they must be corrected in the measured signal. In the end, after correcting for these atmospheric effects and taking into account the non-uniform spectrum of light arriving from the Sun, the actual reflectance at different wavelengths can be computed.

3.1.2 Sensor requirements

A sensor can be characterized by three main parameters: the spatial resolution, the spectral resolution, and the radiometric resolution. In Earth observation, the objective is to have the best spatial resolution combined with a good spectral and radiometric resolution, depending on the mission of the sensor.

First, the radiometric resolution relates to how much information is perceived by a detector. More precisely, it is the minimal photon counting difference that an optical system can measure. Thus, it is directly related to the total number of photons that can be detected per pixel. For instance, in a grayscale image, the higher the radiometric resolution, the more shades of gray will be displayed. Digitally, it is related to the number of bits on which the information is recorded, i.e. a 16-bit image has a higher radiometric resolution than a 4-bit image. [21]

The spatial resolution of a sensor is characterized by the Point Spread Function (PSF) which represents how the light is spatially perceived from a point source by the detector.

More specifically, it is the spatial domain version (i.e. the inverse FT of the Optical Transfer Function (OTF) of an imaging system. The OTF is the function that specifies how different spatial frequencies are captured or transmitted through an optical system. A generic representation of a PSF is shown in Fig. 11

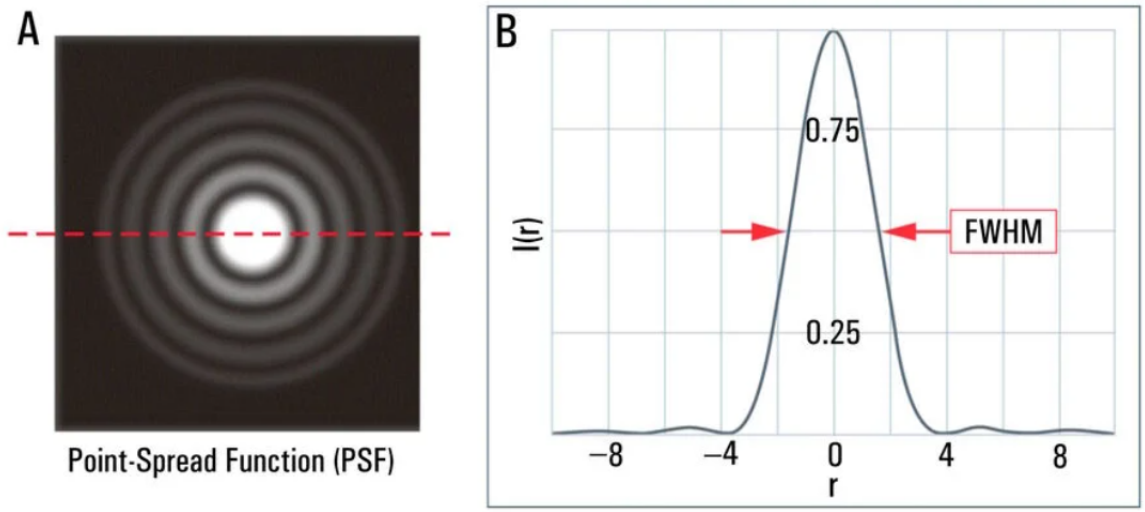


Figure 11: General representation of a PSF. From [22]

The quality of the PSF is characterized by the Full Width at Half Maximum (FWHM) of the central peak. A smaller FWHM indicates a better PSF. The PSF depends on the wavelength of the observed light and on the optical parameters of the instrument. The PSF may also be used to define the spatial or angular resolution of an instrument. The Rayleigh criterion states that two points are resolved if the peak of the first point's PSF is no closer than the first minimum of the other point's PSF. This is shown visually in Fig. 12.

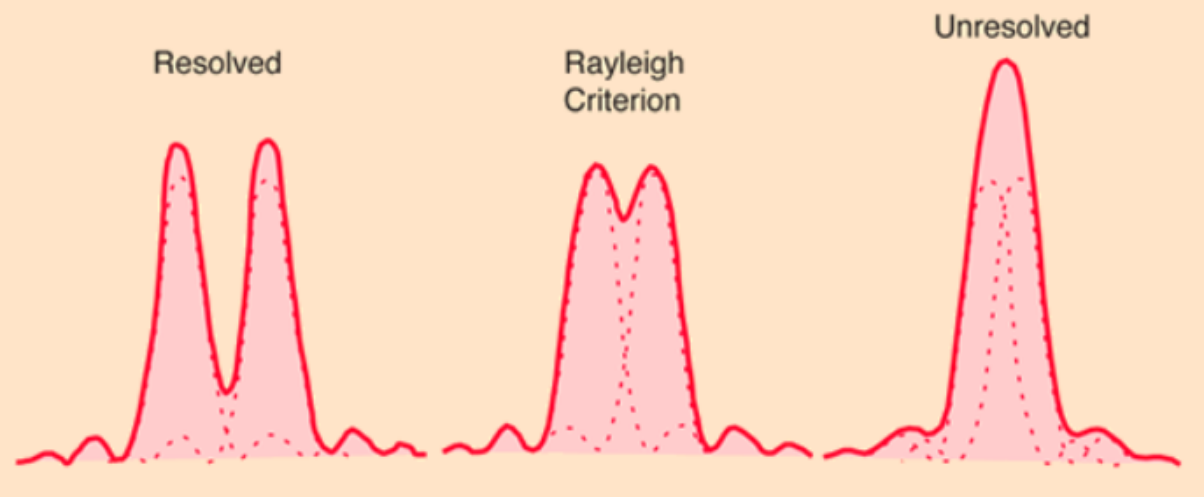


Figure 12: Rayleigh criterion for the resolution. From [23].

Mathematically, this angular resolution for a circular aperture is expressed

$$\sin\theta \approx \theta = 1.22\frac{\lambda}{D} \tag{35}$$

using the small-angle approximation and where λ is the wavelength, and D the aperture. [23]

Concerning the spectral resolution, one can distinguish different types of imaging modalities. The most simple imagers are panchromatic which means that they collect light over some spectral range, then integrate over this range to yield the total light collected per pixel. This allows to display images in grayscale, the pixels with less light collected being darker and inversely. The problem of these simple panchromatic systems is that their spectral information content is very low. Therefore, multispectral sensors are very often preferred. They operate in the same way as panchromatic sensors except that instead of integrating all collected light over one large spectral range, the sensor collects and integrates the light in several narrower spectral bands centered on different wavelengths. The separation of the different wavelengths is done either with filters disposed on different areas of the detector or with dispersive elements as in classical spectrometers. This is the concept used for color photography: by collecting in one band centered one the red wavelength, one on the green, and one on the blue, it is possible to create a RGB, also called true color, image. This concept is not limited to visible light and modern multispectral imagers often have around 10 spectral bands in wavelength ranges going from the blue to the middle or far IR. Indeed, collecting light in the IR spectrum also brings relevant scientific information. The most common example being the Normalized Difference Vegetation Index (NDVI) which is an indicator of vegetation health [24]

$$NDVI = \frac{NIR - RED}{NIR + RED}. \tag{36}$$

Here, NIR represents the measured reflectance in the Near IR (NIR) and RED the measured reflectance in the red. A healthy vegetation produces more chlorophyll which has a very high reflectance in the NIR. Hence, if the vegetation is sick or dying it will produce less chlorophyll, the reflectance in the NIR will decrease and so will the NDVI index.

To increase even further the spectral information content of images, hyperspectral imagers were developed. This is still today a booming field of research and development with more and more implementations on space probes. Hyperspectral instruments work with the same principle that multispectral images except that the spectral bands are much more numerous, narrower, and contiguous. Most of these sensors have more than 200 bands that allow to reconstitute a near continuous spectrum of the target. The difference

between the panchromatic, multispectral, and hyperspectral sensors is illustrated and summarized in Fig. 13.

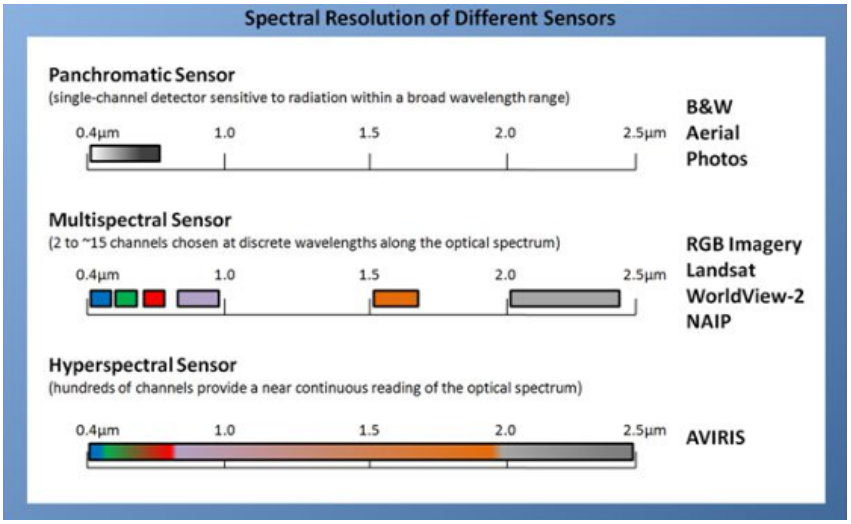


Figure 13: Spectral resolution of different sensors. From [24].

It is important to note that as the number of spectral bands increases, each band becomes narrower, resulting in a decrease in the number of photons collected by each band. To maintain a sufficiently high SNR, the size of the pixels on the detector must be increased to collect more light. Therefore, an increase in spectral resolution typically results in a decrease in spatial resolution. [20]

3.1.3 Acquisition modes

When acquiring spectral images, both multi- and hyperspectral, one cannot measure the signal in the whole scene and in all spectral bands at the same time since it implies to measure three-dimensional data (two spatial and one spectral dimension) with a two-dimensional detector. Two methods exists to acquire these spectral images. The first one is the Push-Broom (PB) mode. In this mode, the sensor scan the scene along-track line by line, i.e. a line of pixel is scanned and the spectral differentiation is done by dispersive optics or filters. The satellite motion is then used to scan the whole scene. On the other hand, the Whisk-Broom (WB) mode scans the scene pixel by pixel across-track using a rotating scan mirror. Then, it also uses the motion of satellite to scan the entire scene. While the mechanical systems of PB sensors are more simple than WB sensors, they have more complex optical systems since more detectors are needed. This may also lead to problems due to the varying sensitivity of the individual detectors. However, an advantage of the PB mode is that it allows to gather more light because the sensor spends more time on the same area.

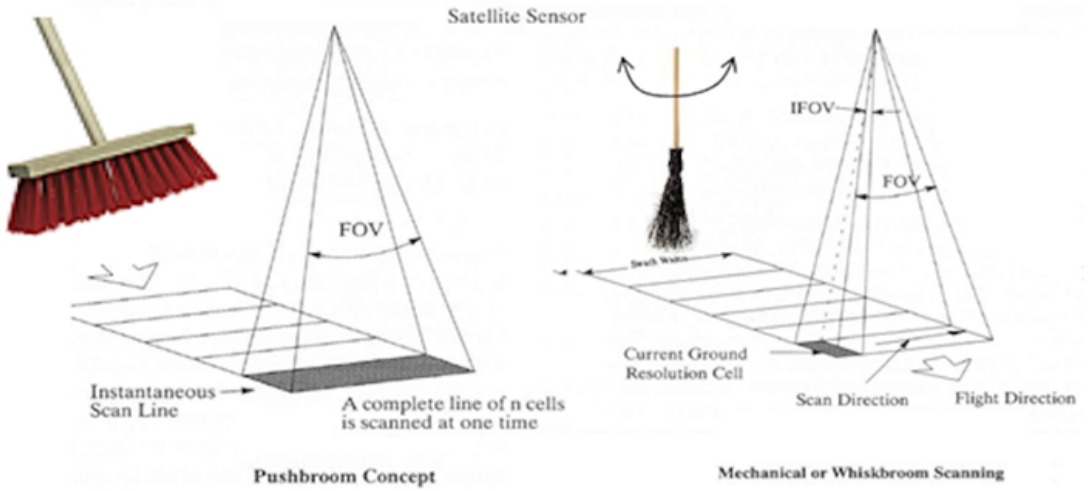


Figure 14: PB and whisk-broom acquisition modes. From [25].

3.2 Some considerations about compressive sensing in the infrared

The CS technique presents considerable advantages in IR imaging. Indeed, when observing at long wavelengths, notably in the middle or thermal IR, the optical flux is reduced. Hence, the number of photons collected per unit of area decreases with respect to shorter wavelengths. Having a satisfying SNR at long wavelengths thus requires larger pixels to collect more light. This results in the need for large focal plane arrays to reach adequate imaging performances. As a consequence, the required detectors are expensive and spatially extended, which is prohibitive in terms of manufacture and application, especially for space implementation. Hence, CS could help to overcome this problem since identical or better resolutions can be achieved with detectors containing less pixels. However, some challenges arise when applying CS in the IR domain. The most important is the development of a viable SLM technology for the Thermal IR (TIR). Another one is the lack of optical flux concentration which results in a low SNR and unsatisfying performances. A possibility to overcome this issue is to add light concentration optics in the optical train, between the SLM and the detector. Even though this is relatively easy to implement, it prevents the design of lensless cameras which is a possible outcome of CS technology. [8] [26]
The rest of this section addresses the specific case of each IR spectrum sub-region, i.e. their different advantages and challenges in the context of CS and Earth observation.

3.2.1 NIR-SWIR

While no absolute definition exists, one may consider that the near IR (NIR) extends from around 700 nm to 1.3 μm and the short-wavelength IR (SWIR) extends from around 1.3 μm to 3 μm. These wavelengths are still relatively close to the visible spectrum so

there is usually no big differences in the signal's treatment, except the need for slightly larger detectors and a slightly lower spatial resolution. This sub-region of the IR spectrum is often called the reflective IR since the Sun still emits photons at these wavelengths and reflection from the Sun's emission can thus be measured. Therefore, observing in this region allows to acquire useful information in many contexts, including Earth observation. As mentioned previously, acquiring a target's reflectance in the IR domain with multispectral or hyperspectral sensors provides data that can be used to compute indexes such as the NDVI, but also the Normalized Difference Moisture Index (NDMI). The latter compares the NIR and SWIR values to detect moisture contents in leaves

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR}. \tag{37}$$

Water has strong absorption lines in the SWIR region and a dry leaf has a lower reflectivity in the NIR region. Therefore, vegetation with high water content will have a higher NDMI. The NDVI and NDMI are just simple examples of the numerous applications resulting from Earth observation in the NIR-SWIR. Applying the CS methodology in this spectrum could drastically reduce the cost and size of the detectors.

3.2.2 TIR

TIR refers to the thermal IR range which is composed of two main windows: the Mid-Wavelength IR (MIR) which typically refers to the region of the electromagnetic spectrum between 3 and 5 μm, and the Long Wavelength IR (LWIR): which ranges from 8 to approximately 14 μm. Between these two sub-regions is an atmospheric absorption window, which is thus often not considered in Earth observation. Remote sensing of the Earth in the TIR spectral range is useful to measure self-emission coming from sources on the surface. These sources can be hotspots such as active fires or volcanic activities in the MIR, or cooler bodies such as vegetation self-emission in the LWIR[2]. A few examples of variables that can be measured in this spectral region are land surface temperature, atmospheric composition, or ocean temperature and composition. It can also be used to detect and monitor wildfires, in precision agriculture, and ecosystems monitoring. The main challenge of TIR imaging is the difficulty to achieve high resolution, typical to these long wavelengths. The lack of optical flux concentration results in a decrease in SNR of $\sim \frac{1}{\sqrt{M}}$,with M the number of pixels of the detector, leading to inadequate performances. Also, TIR detectors must often be cooled down to avoid detection of their own emission. TIR imagers are an active field of research and development, mainly to achieve better spatial resolutions. The CS methodology has great potential in this field considering its upsides of smaller detector arrays and super resolution. However, developing an efficient

[2]NB: Sometimes the TIR denomination is used to characterized only the spectral region around the Earth thermal emission, i.e. the LWIR region.

SLM technology is complex at these wavelengths due to diffraction effects.

3.3 State of the art of compressive sensing in the infrared

In this section, several architecture developments aimed at applying CS in various bands of the IR are presented. Except for the first one, all those presented are in view of space applications, mostly for Earth Observation. The objective is to show the present capabilities of CS imagers in the different IR regions.

3.3.1 InView single-pixel camera

McMackin et al. [27] managed to develop a high-resolution CS camera working in the range from 900 nm to 1.7 μm. They used the InView single-pixel camera architecture which is composed of an imaging system, a DMD to modulate the signal, and then optics to condensate the resulting light on the single-pixel detector. This architecture is represented in Fig. 15.

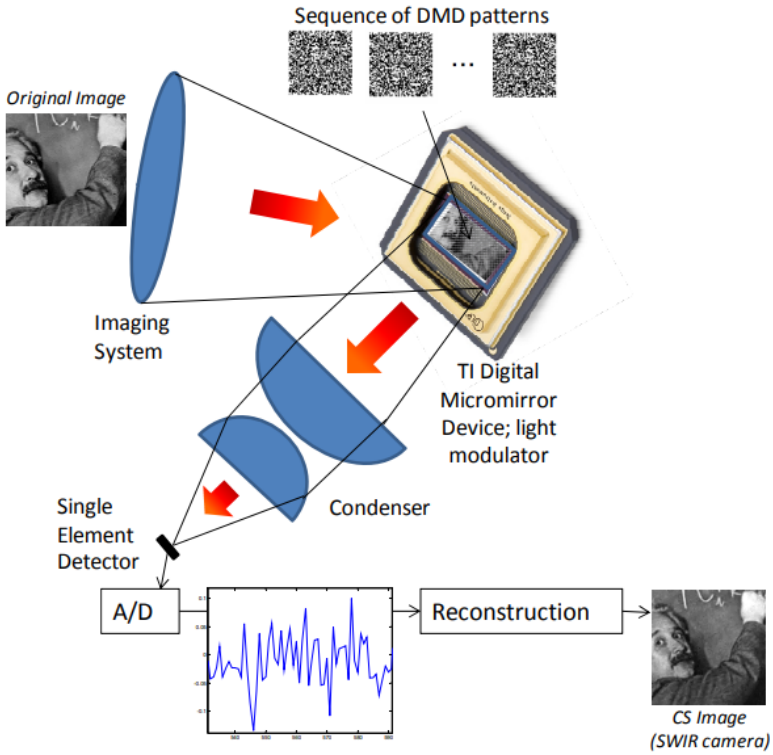


Figure 15: Schematic of the InView single-pixel camera architecture. From [27].

The DMD is composed of 1024 × 768 pixels, each of them being a micromirror that can be switched to 'on' and 'off' modes individually. This is done by the DMD Pattern Generation Board which carries a field-programmable gate array to generate patterns according to the required CS ratio and deliver them to the DMD. They managed to use the super resolution method with a SWIR single element detector and the DMD to

reconstruct images of 1024 × 768 pixels. The experiment showed that SWIR images could be reconstructed with good fidelity with a compression ratio up to 90% using a TV reconstruction algorithm.

3.3.2 The SISSI instrument

The SISSI instrument is an on-going study of a super-resolved compressive multispectral imager for Earth observation in the medium IR spectral region. The concept is described by Raimondi et al. [19] and funded by the Italian Space Agency. The principle is inspired by the single-pixel camera and uses the super resolution method. It is composed of collection optics (the telescope), a DMD, focusing optics and a detector. The telescope collects the light from the scene and send it on the DMD which acts as a SLM with suitably coded binary masks, different for each frame. The light is focused on the detector by the focusing optics. The detector is composed of $M \times M$ macro-pixels. Each of those correspond to a group of $N \times N$ micro-mirrors on the DMD. Each micro-pixel represents the super-resolved ground sampling distance (GSD). The acquisition of the same scene in different spectral bands is obtained with spectral filters on the detector, disposed along the across-track dimension, combined with the along-track apparent movement of the scene (PB acquisition mode). The working principle of SISSI is shown in Fig.16.

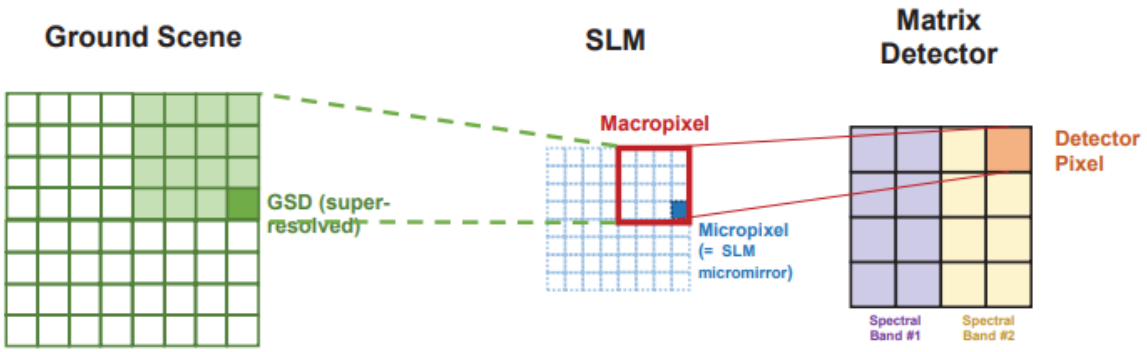


Figure 16: Working principle of the super-resolution of SISSI optical instrument. From [19].

The technical specifications of the SISSI payload, based on the analysis of end user requirements and a market survey to identify elements available, is shown in Table 1.

PARAMETER	VALUE
Acquisition mode	PB (slitless)
Operational working spectral range (μm)	3-4 (MIR)
Nr. of spectral bands	4
Spectral band central wavelengths (μm)	3.3, 3.5, 3.7, 3.9
FWHM (nm)	100 at 3.5, 3.7, and 3.9 μm, 200 at 3.3 μm
Nominal GSD (m)	15.0
Nominal altitude (km)	700
Nr. of micropixels (across track)	1024
Swath across track (km)	15.36
Super-resolution factor	4×4
SNR	100 at 1000K
Integration time (ms)	2-4
Detector	MARS by Sofradir/Lynred
SLM	DMD by Texas Instruments Inc.

Table 1: Main technical specifications of the SISSI payload. From [19].

3.3.3 The SURPRISE demonstrator

The SURPRISE (SUper-Resolve comPRessive InStrument in the visible and medium IR for Earth observation application) demonstrator aims to describe how to use SLM technology and CS to improve the performance of Earth observation super-spectral payloads in the visible, NIR and MIR. It is described by Raimondi et al. [28]. SURPRISE is based on essentially the same working principle as SISSI with two major differences. First, the detector is a single-pixel detector. Second, the instrument is conceived as a WB spectral imager with ten channels in the VIS-NIR and two channels in the MIR. The instrument relies on the super resolution concept and is designed to ensure a super resolution factor of at least 4×4 and that can rise up to 32×32. At the beginning of the measurement sequence, a portion of the scene (target) is seen by the instrument's instantaneous Field Of View (iFOV). Then, an appropriate modulation mask is applied on the SLM and a new acquisition by the detectors is triggered. Once, the integration time of all detectors has run, a new acquisition is made with a different modulation mask. This operation is repeated until the required number of CS measurements is reached. After that, a new portion of the scene is seen by the iFOV and the measurement process is repeated until the full scene is scanned. The simplified optical layout of the SURPRISE demonstrator is shown in Fig. 17.

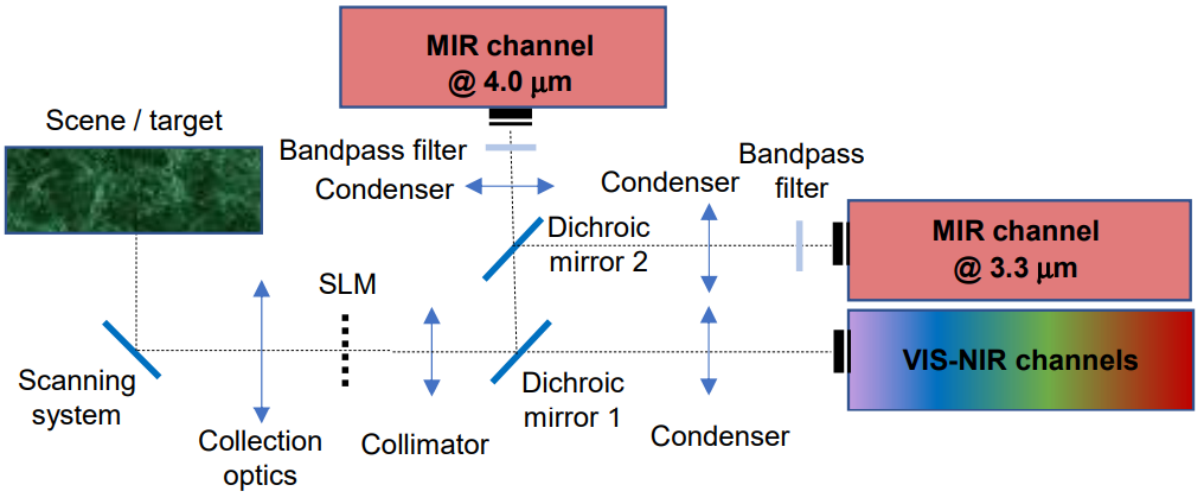


Figure 17: Optical layout of the SURPRISE demonstrator. From [29].

3.3.4 MIR camera for sky observation and real-time detection of NEOs

The proposed instrument is inspired by the NEOWISE project funded by NASA's Planetary Science Division that detects the presence of asteroid and comets from images acquired by the WISE spacecraft. The instrument concepts aims at using CS properties to combine the sky survey image acquisition and on-board detection of Near Earth Objects (NEOs), instead of doing the processing and data-mining at ground. The idea behind it is that this type of images is expected to be very sparse in the pixel domain, making the use of CS relevant. The optical concept is rather simple and shown in Fig. 18.

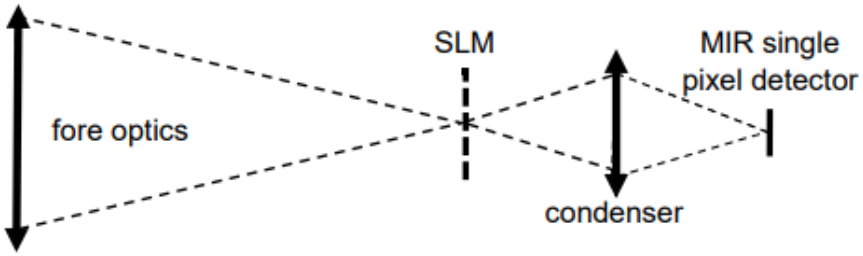


Figure 18: Diagram of the camera operating in the TIR for sky observation and real-time detection of NEOs. From [30].

The proposed payload has four principal elements: a telescope, a SLM, a condenser lens and a single-element detector. The detector is a single-pixel CS-based panchromatic camera working at 3-5μm. This working wavelength range makes it necessary to have a payload cooled down to about 150°K. Concerning the SLM, a DMD was initially investigated. The main drawback was the minimum working temperature of 233°K, incompatible with the temperature requirement of the payload. For this reason, a transmissive coding mask is used. The expected mass and volume reduction of the payload relative to traditional instruments is rather limited (about 1-kg mass reduction). The major advantage of the

CS approach is expected in terms of data transmission requirements.[30] [31]

3.4 The problem of satellite motion

One drawback of the CS method is the increase of acquisition time. Indeed, since several images must be acquired with different masks to reconstruct the complete scene, it takes more time than in classical imaging. This time is hardware dependent since it results from the integration time of the detector and the SLM framerate. This limitation can easily be overcome in laboratory systems, where the exposure time can be adapted to the available hardware. However, in satellite Low Earth Orbit (LEO) conditions, the rapid motion of the instrument restricts the exposure time. In fact, the dwell time, i.e. the time a sensor element covers one sampling distance, is set by the platform characteristics, which depend on ground velocity and GSD. In CS the spatial resolution increases with the number of mask patterns applied. But, when the dwell time is limited, increasing the number of masks results in decreasing the exposure time per pattern, with degradation of image quality or even infeasibility. In this context, the effect of satellite motion on the reconstruction quality of CS images has been studied in more details by Oggioni et al.[32]. The research compared the reconstruction quality of static images and (simulated) moving images. The authors used Landsat-8 data to compare spectral bands in the visible, NIR and TIR. The simulation mimics a PB imager. In addition to comparing static and moving scenes, the classic PB imaging has been compared with the PB CS technique. The figure of merit used is the Reconstruction Error (RE)

$$RE(\%) = \frac{\|\tilde{u} - u_0\|_F}{\|u_0\|_F} \times 100, \tag{38}$$

where $\tilde{u}$ is the reconstructed image, u_0 the original image, and $\|u\|_F$ denotes the Frobenius norm defined as

$$\|u\|_F = \sqrt{\sum_{i=1}^{q}\sum_{j=1}^{p} |u_{ij}|^2}. \tag{39}$$

To simulate the moving scene, a downsampling approach is applied, working as follows. A macropixel composed by P pixels is considered as the sensing pixel (P is in the order of a few tens). Hence, the field of view (FOV) is represented by $N \times 1$ macropixels. The movement of the scene is simulated by shifting the image one pixel along track at any instant. It takes a time equal to the dwell time for the FOV to move to the following macropixel. This approach is represented in Fig. 19.

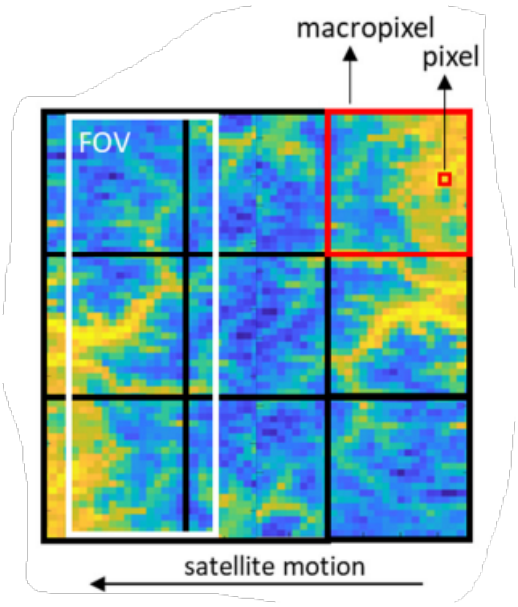


Figure 19: Schematic of the simulation in case of a moving scene. From [32].

After one dwell time, a set of frames is obtained. Two acquisition methods were compared. First, the classical PB acquisition mode where all the frames are summed up and the intensity inside the micropixel is integrated. Second, the PB method is combined with the CS approach. A 1-D mask is applied to each macropixel column and then the line of macropixels is reconstructed using a TV algorithm. The reconstruction errors were evaluated for both methods and as a function of the parameter τ defined as the ratio between the acquisition time and the dwell time. The results are presented in Fig. 20.

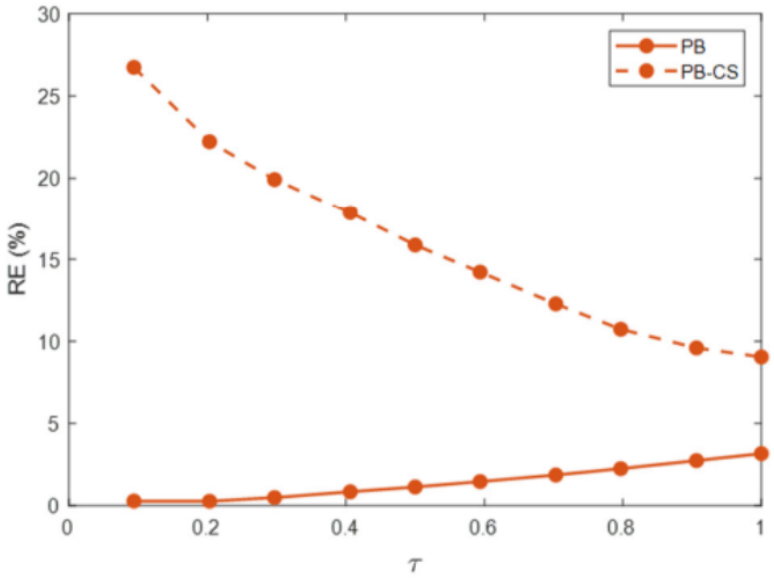


Figure 20: Reconstruction error in the green channel as a function of τ. From [32].

In case of the PB CS ,τ is equivalent to the CS ratio m/N, i.e. it takes the dwell time to acquire N patterns. In any case, the classical PB yields better results than the CS

one. The RE increases with τ in the PB mode because the scene is more stable when the acquisition time is shorter. Generally, in classical PB acquisitions, the acquisition time is shorter than the dwell time to avoid image degradation due to the satellite motion. Nevertheless, in the CS mode the RE decreases when τ increases. Indeed, an increase of τ means that more data is acquired, and when $\tau = 1$ the measured signal as the same length as the signal and the method is thus equivalent to classical imaging in terms of quantity of data acquired. Hence, the reconstruction quality increases since the reconstruction basis is more complete. In the end, if the RE is worse with the CS method it is because since the scene is moving, the last masks are applied to a very different field of view than the first ones, thus the reconstruction accuracy is worse. Therefore, care must be taken when applying CS in Earth observation. The quantity of patterns to apply and/or the framerate of the SLM used must be chosen adequately so that the acquisition time is low enough and the satellite motion does not degrade the image. However, as explained earlier, the lower the quantity of data acquired, the less precise is the reconstruction, thus a trade-off must be performed. It must also be noted that this problem of satellite motion disappears if it is placed in geostationary orbit. Nevertheless, in this orbit the satellite is further away from the Earth than in LEO, resulting in a decrease of spatial resolution. Moreover, the satellite can only cover one part of the globe and a constellation must be formed if the objective is to cover the entire planet.

3.5 Compressive hyperspectral remote sensing

CS shows great potential in hyperspectral Earth observation. Hyperspectral imagers collect information in three dimensions: two spatials and one spectral. They often have more than 200 spectral bands, resulting in large datacubes containing the information. Hence, the hyperspectral data files can reach sizes up to a few gigabytes for a single image. A successful implementation of the CS method in hyperspectral imagers could therefore drastically reduce the quantity of data to acquire. A few designs for such sensors have already been proposed. Depending on the architecture, the compressive measurement can be applied only in the spectral dimension, in the spectral and one spatial dimension, or in the spectral and both spatial dimensions. Fowler [33] proposes architectures for PB and WB CS hyperspectral imagers. In these designs, the measurements are compressive in the spectral dimension only. However, the PB or WB scanning has the advantage to not need access to the entire image at once. The PB design is shown in Fig. 21.

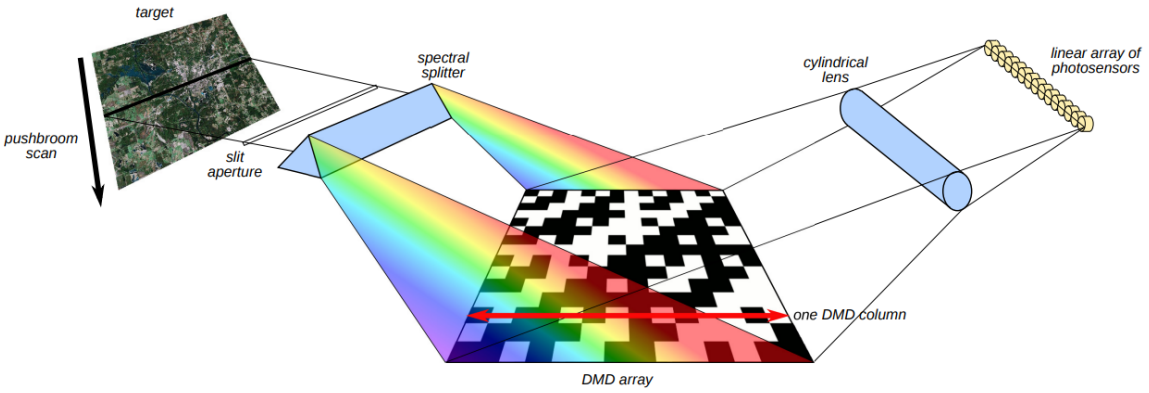


Figure 21: A compressive PB hyperspectral sensor. From [33].

As it can be seen, this design allow the acquisition of the spatial and the spectral dimension of the line being scanned with a single linear array of photosensors. The spectral differentiation is done with a dispersive element which projects each spectral band in a different line of the DMD. Each column of the DMD thus corresponds to the different spectral bands of one pixel of the line being scanned. The reflection of the DMD is then focused on the linear detector by a cylindrical lens. Mathematically, it means that the line being imaged contains M pixels $\boldsymbol{x}_m$, and each pixel is made of N spectral bands such that $\boldsymbol{x}_m \in \mathbf{C}^N$. The compressive measurement of $\boldsymbol{x}_m$ yields the measurement $\boldsymbol{y}_m = \boldsymbol{\Phi}_m^T \boldsymbol{x}_m$. With $\boldsymbol{\Phi}_m$ of size $N \times K$, $K \ll N$, and thus $\boldsymbol{y}_m \in \mathbf{C}^K$. Hence, each column of the DMD performs the inner product of one hyperspectral pixel $\boldsymbol{x}_m$ with a particular column of $\boldsymbol{\Phi}_m$. Note that $\boldsymbol{\Phi}_m$ could be identical for all pixels. In the end, by sequentially applying the K columns of $\boldsymbol{\Phi}_m$ to the corresponding column of the DMD, the measurement vector $\boldsymbol{y}_m$ of pixel $\boldsymbol{x}_m$ can be obtained. The main drawback of the PB architecture is the need to perform K measurements successively on the same pixel line. Indeed this could lead to a decrease of quality in the reconstructed image due to the problem of satellite motion detailed in the previous section.
The WB design is shown in the next figure.

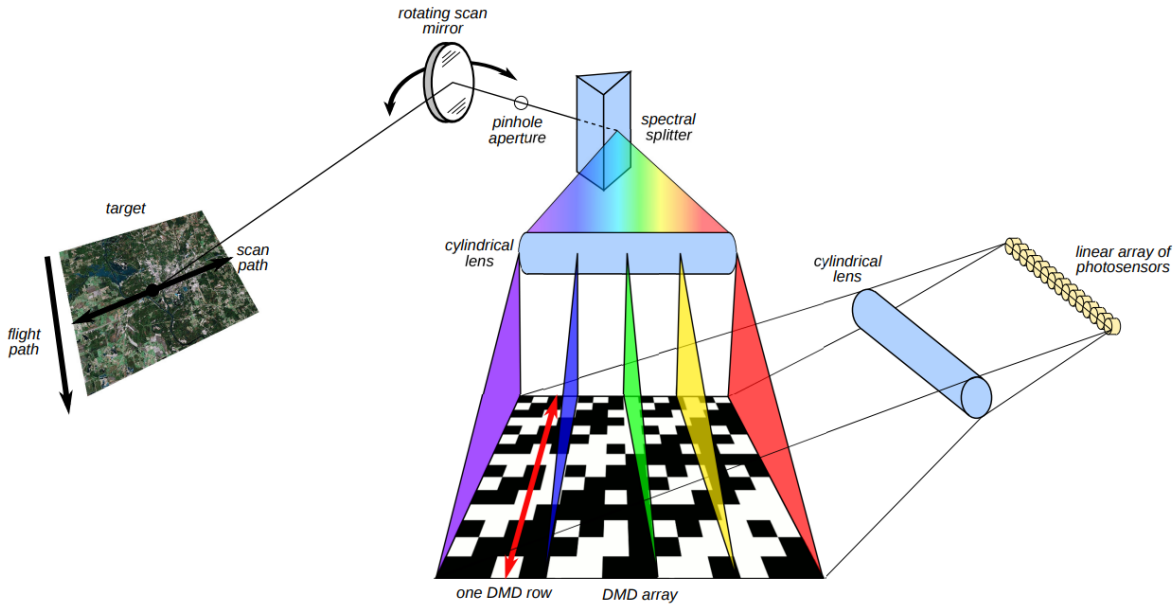


Figure 22: A compressive WB hyperspectral sensor. From [33].

Since it is a WB instrument, one pixel is scanned at a time instead of a line of pixels as in a PB sensor. The light passes through a pinhole aperture, then a dispersive element to discriminate the spectral bands, and a lens focuses each band on one specific row of the DMD. Here, since only one pixel is imaged at a time, the same information is projected on each column of the DMD. Therefore, by applying different patterns on each column, all the measurements can be done at once. Note that, if the same pattern is used to sense every pixel in the image, a DMD is not even required and a fixed coded aperture is sufficient. The advantage of the WB design is that all measurements are taken at once and thus the dwell time is increased with respect to the PB design which needs to take K measurements successively. Nevertheless, the dwell time of any WB sensors stays way below the dwell time of a classical PB sensor.

4 Simulation on multispectral data

This section has for objective to test the CS method for Earth observation applications. For this purpose, the ISTA-Net$^+$ algorithm is used to perform CS simulations on real satellite data. It must be noted that the ISTA-Net$^+$ source code has been slightly adapted to handle multispectral data. The source code from [16] takes RGB images as entries, converts them to grayscale, applies appropriate masking depending on the CS ratio requested, execute the ISTA-Net$^+$ algorithm, and returns the reconstructed images. Here, it has been modified to execute a CS reconstruction on each spectral band as a grayscale image, and then combine the reconstructed images in one single HDF5 file. The Hierarchical Data Format (HDF) is a data format specifically designed by NASA to handle Earth observation data. It is especially adapted to handle large datacubes typical of multispectral and hyperspectral images. To test the efficiency of the algorithm on real data, multispectral images from SENTINEL-2 have been retrieved from the Copernicus Earth observation data browser. This section first presents the SENTINEL-2 mission and its data acquisition strategy. After that, the selected database is presented, then estimators for the quality of the reconstruction are described, and finally the results of the simulation are presented and discussed. It is important to note that the problem of satellite motion described earlier is not considered in these simulations.

4.1 SENTINEL-2

SENTINEL-2 is a space mission from ESA composed of two satellites: SENTINEL-2A and SENTINEL-2B. The former was launched on 23 June 2015 followed by the latter on 7 March 2017. Both satellites are identical multispectral Earth observation instruments on the same sun-synchronous orbit at a mean altitude of 786 km but phased at 180° to divide the revisit time by two. The mission's orbit has an inclination of 98.62° which makes observations possible in latitude bands extending from 56° South (South America) to 82.8° North (above Greenland), with a revisit time of 5 days at the equator. With its systematic global acquisition of high-resolution multispectral data over land and coastal areas, SENTINEL-2 takes part to the european Copernicus program with contributions in various themes such as climate change, land monitoring, emergency management, and security. The instrument is a passive detector working in 13 spectral bands from the visible to the SWIR. The light is collected by a three-mirror telescope and focused onto two focal plane assemblies, one for the visible and NIR bands (VNIR) and one for the SWIR bands, via a beam-splitter. It has a field of view of 290km and works in a PB mode. The different resolutions for each band as well as the band center wavelength and bandwidth are specified in Table 2.

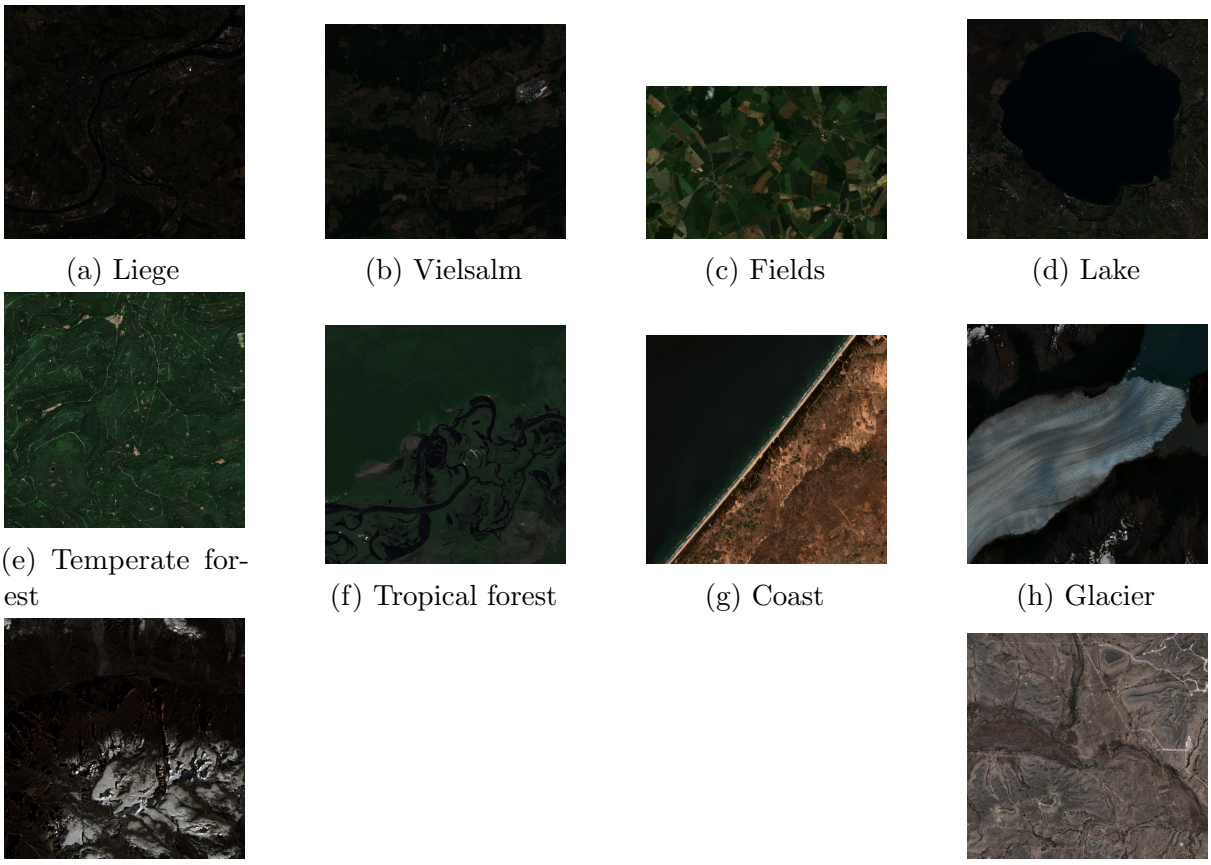
Spatial Resolution (m)	Band Number	S2A		S2B	
		Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)
10	2	492.4	66	492.1	66
	3	559.8	36	559.0	36
	4	664.6	31	664.9	31
	8	832.8	106	832.9	106
20	5	704.1	15	703.8	16
	6	740.5	15	739.1	15
	7	782.8	20	779.7	20
	8a	864.7	21	864.0	22
	11	1613.7	91	1610.4	94
	12	2202.4	175	2185.7	185
60	1	442.7	21	442.2	21
	9	945.1	20	943.2	21
	10	1373.5	31	1376.9	30

Table 2: Wavelengths and bandwidths of the three spatial resolutions of the MSI instrument. From [34].

4.2 Database

The data is obtained in the form of 12 TIFF files, one for each band[3], pre-processed so that all bands are combined in one single HDF5 file, and then injected into the algorithm. Ten images arbitrarily selected to represent various landscapes are used as a test set to evaluate the accuracy of the reconstruction. They are shown in Fig. 23 and detailed in Table 3.

[3]Band number 10 is not retrievable on the browser.



(a) Liege (b) Vielsalm (c) Fields (d) Lake

(e) Temperate forest (f) Tropical forest (g) Coast (h) Glacier

(i) Mountains (j) Desert

Figure 23: SENTINEL-2 image database.

	Landscape	**Location**	**Date**	**Resolution (pixels)**
(a)	Urban area	Liège, Belgium	13/08/22	604 × 585
(b)	Rural town	Vielsalm, Luxembourg, Belgium	10/08/22	614 × 585
(c)	Agricultural fields	Haute-Marne, Grand Est, France	21/05/22	519 × 328
(d)	Lake	Bracciano Lake, Italy	14/02/23	808 × 734
(e)	Temperate forest	Dark Forest, Germany	12/08/22	504 × 491
(f)	Tropical forest	Amazon Forest, Brazil	24/12/22	887 × 879
(g)	Coastal area	Senegal	26/02/23	534 × 507
(h)	Glacier	Perito Moreno Glacier, Argentina	10/01/23	718 × 714
(i)	Mountains	Himalaya, Nepal	11/02/23	641 × 639

Table 3: SENTINEL-2 image database details.

They were taken in different sizes to analyse a potential impact of the image size on the reconstruction quality. These images are injected into the pre-trained ISTA-Net$^+$ algorithm to apply masks and mimic the CS acquisition process, and then reconstruct the images with various CS ratios. It is important to note that the algorithm used was pre-trained by the authors of the original code, but only on classical images, not Earth observation data. The accuracy of the reconstruction would most probably increase if the algorithm is retrained with Earth observation images. However, training a DL algorithm

is a computationally intensive and resource-consuming task. The purpose of this research is to implement a complete CS experimental chain from measurement to reconstruction and to explore possibilities for space applications, not to obtain the maximum possible accuracy. Therefore, the results shown later in this section are considered satisfactory for using the algorithm without retraining it.

4.3 Reconstruction quality estimators

4.3.1 Peak signal to noise ratio

The Peak Signal to Noise Ratio (PSNR) is the ratio between the maximum value of a signal and the noise distorting this signal. In this case, the noise is the one between the original and reconstructed signals and is represented by the Mean Squared Error (MSE) over all the pixels defined as

$$MSE = \frac{1}{N_x}\frac{1}{N_y}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}(\hat{s}_{i,j} - s_{i,j})^2, \tag{40}$$

with N_x and N_y the number of pixels on the x and y axis, $\hat{s}_{i,j}$ the reconstructed signal in pixel (i, j) and $s_{i,j}$ the real value of the signal in pixel (i, j). PSNR is usually expressed using the decibel scale. Hence, the PSNR is expressed as

$$PSNR = 10\log_{10}\left(\frac{(\text{peak val})^2}{MSE}\right), \tag{41}$$

where peak val is the maximum signal value. If the signal is normalized, its values are floats between 0 and 1, thus the maximum value is 1. A higher PSNR generally indicates a lower MSE and thus a better reconstruction.

4.3.2 Structural Similarity Index Method

The PSNR is an indicator of reconstruction quality which works well but is a bit rough in the sense that it does not take human visualization into account. Indeed, the PSNR measures the difference between the reconstructed and the original signal pixel by pixel. While this method evaluates quantitatively the quality of the reconstructed signal with precision, it does not consider the representation of structures and contrasts that are more easily perceived by the human eye. In this context, Wang et al.[35] developed a new method called the Structural Similarity Index Method (SSIM). The SSIM compares the original and the reconstructed signals using three independent parameters: the luminance, the contrast and the structure.

First, the luminance of each signal is estimated as the mean intensity

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{42}$$

Stating that x and y are the two signals to compare, the luminance comparison function is given by

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{43}$$

with C_1 a small constant. An estimation of the constrast is then introduced using the standard deviation

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{\frac{1}{2}} \tag{44}$$

and the contrast comparison function takes the form

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{45}$$

with C_2 a small constant. Finally, the structure comparison function is defined using the correlation coefficient (covariance) between the two signals

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y). \tag{46}$$

Hence, the function is written

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x + \sigma_y + C_3} \tag{47}$$

with C_3 a small constant. In the end, the three components are combined to form the final similarity measure

$$SSIM(x,y) = [l(x,y)^\alpha \cdot c(x,y)^\beta \cdot s(x,y)^\gamma], \tag{48}$$

with $\alpha > 0, \beta > 0, \gamma > 0$ parameters that can be adjusted to the relative importance of each parameter. Setting all of them to 1, $C_3 = C_2/2$, and injecting in the above expressions:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2\mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{49}$$

This produces a value between -1 and 1, where a value of 1 indicates that the images are similar, and a value of -1 indicates that the images are completely dissimilar. Generally, it is more convenient to use the SSIM locally rather than globally to assess an image quality. In other words, the SSIM is evaluated in sub-blocks of the image and a mean SSIM is

computed to estimate the global SSIM.

The PSNR and the SSIM are both relevant quantities to estimate the reconstruction quality. While the PSNR is based on the absolute difference between the two signals, the SSIM focuses more on structure representation and distorsion. They can be seen as complementary to each other and will both be used in this research.

4.4 Results and discussion

Results of the reconstruction of the images specified in the database with ISTA-Net$^+$ depending on the CS ratio are shown in the following tables. Each result presented is the mean over the 12 bands for one image and one CS ratio. In addition of the two parameters explained in the previous section, the computation time is represented. As mentioned earlier in this work, the algorithm used is pre-trained by its authors on traditional images and better reconstructions could be achieved by training it with Earth observation data. However, the training process of such algorithm is extremely time and resource consuming. Hence, the simulation is executed with the pre-trained algorithm to estimate if it is accurate enough to be used in the context of testing a first full chain implementation of the CS technique, without aiming for maximum accuracy.

CS Ratio (%)	1	4	10	25	40	50
Liège	78.99	80.91	84.96	89.38	92.69	93.72
Vielsalm	75.78	78.72	83.1	88.53	92.34	94.98
Fields	72.11	76.28	81.03	87.33	91.62	94.28
Lake	75.16	78	81.59	86.55	90.26	92.77
Temperate forest	72.08	75.14	78.71	83.9	87.57	89.84
Tropical forest	75.44	79.88	84.94	90.49	94.12	96.79
Coast	69.59	73.97	77.34	81.91	85.25	87.32
Glacier	74.91	78.98	83.29	88.73	92.48	95.03
Mountains	70.41	73.67	77.55	83.18	87.24	89.87
Desert	71.05	74.43	78.1	83.11	86.79	88.99

Table 4: PSNR (dB).

CS Ratio (%)	1	4	10	25	40	50
Liège	0.8697	0.8957	0.9469	0.9756	0.9877	0.9924
Vielsalm	0.8138	0.8675	0.9338	0.9757	0.9886	0.9931
Fields	0.7502	0.8484	0.927	0.9767	0.9897	0.9938
Lake	0.814	0.8667	0.9261	0.9714	0.9862	0.9913
Temperate forest	0.7015	0.7764	0.8578	0.9372	0.9671	0.9779
Tropical forest	0.832	0.9017	0.9511	0.9799	0.9899	0.9937
Coast	0.7427	0.8091	0.8767	0.9413	0.967	0.9768
Glacier	0.8533	0.9036	0.9465	0.9773	0.9879	0.992
Mountains	0.7095	0.795	0.8868	0.9567	0.98	0.9875
Desert	0.6781	0.7632	0.8555	0.9354	0.9652	0.9764

Table 5: SSIM.

At first sight of Table 4 and 5, one can observe that for each image, both the PSNR and the SSIM increase with the CS ratio. It is what is instinctively expected since a higher CS ratio means that more data is acquired, thus the reconstruction basis is more complete and the reconstruction should be more accurate. This relationship can be seen more properly in Fig. 24 and Fig. 25. There, it can be seen that the relationships with the CS ratio are more logarithmic than linear, especially for the SSIM. Keeping in mind that the objective is to have a minimum CS ratio with a maximum reconstruction quality, fixing the CS ratio to 25% appears to achieve the best balance between these two considerations. Indeed, these results show that even without further training, the algorithm achieves a good reconstruction of these images, e.g. reaching a SSIM of at least 0.93 for each image with a CS ratio of 25%.

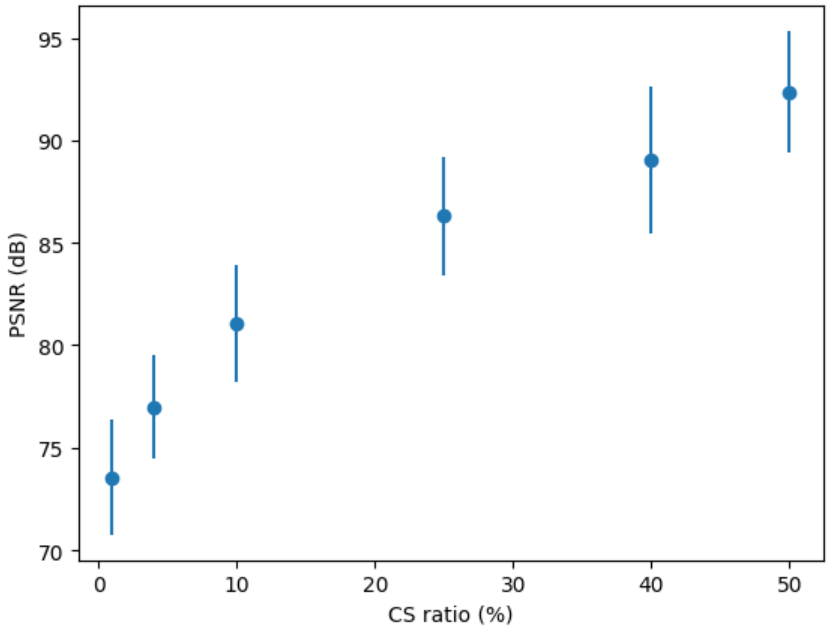


Figure 24: Mean PSNR along with the standard deviation of the 10 images as a function of the CS ratio.

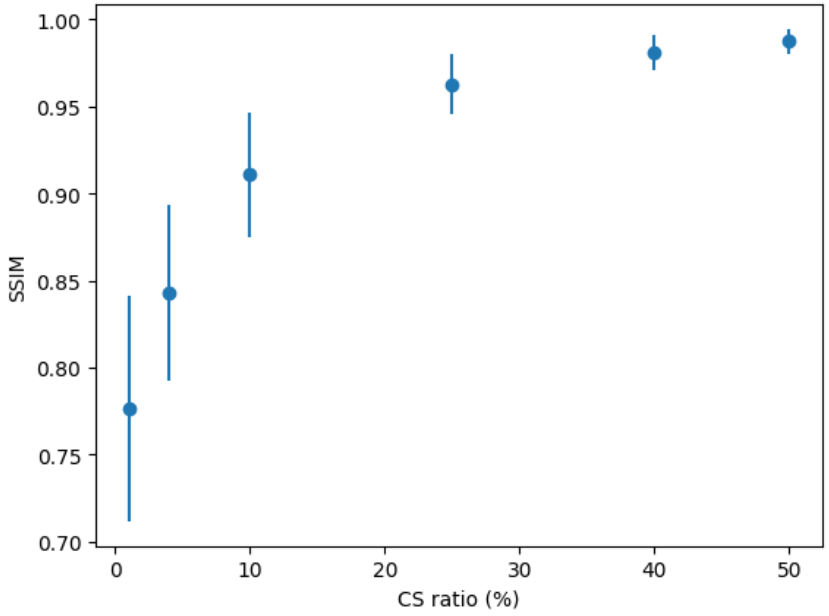


Figure 25: Mean SSIM along with the standard deviation of the 10 images as a function of the CS ratio.

CS Ratio (%)	1	4	10	25	40	50
Liège	127.88	129.91	131.34	129.31	134.07	138.63
Vielsalm	181.01	150.74	174.26	122.76	127.93	153.65
Fields	65.9	63.32	61.29	66.39	67.42	70.82
Lake	244.72	202.98	268.29	221.2	240.18	267.09
Temperate forest	127.98	133.39	131.6	112.48	98	141.29
Tropical forest	390.27	275.3	264.85	271.48	275.31	294.35
Coast	109.94	107.57	119.52	106.93	108.97	121.85
Glacier	175.2	193.91	186.05	186.68	209.31	222.23
Mountains	159.12	190.51	164.72	172.85	165.57	181.2
Desert	125.71	166.97	167.58	148.8	160.06	146.23

Table 6: Computation time (seconds).

Regarding the computation time, no direct relationship with the CS ratio is observed in Table 6 and it seems somewhat random. It is important to note that this computation time might have varied depending on other activities on the computer which could have slow it down. Thus, these results should be treated with caution. Nonetheless, significant differences seems to appear between the different images. This is probably connected to the various sizes of the images. To illustrate that, Fig. 26 shows the mean computation time for each image displayed as a function of the image size. It is clear on this graph that the computation time increases linearly with the image size. Applying a linear regression yields an increase rate of 3.624×10^{-4} seconds per pixel. While it may appear insignificant in this testing context, this increase of computation time with the file size could be important for some implementations. Indeed, scientific or industrial needs often imply larger observation fields and/or higher resolutions which could significantly increase the number of pixels in the images. Thus, the computation time would consequently largely increase which might become compelling.

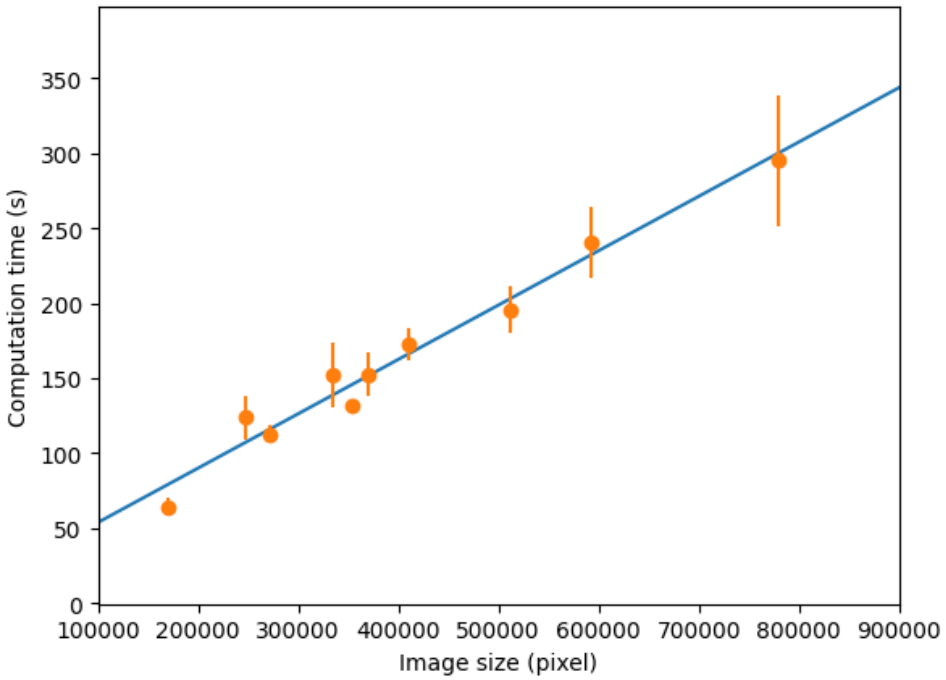


Figure 26: Mean computation time as a function of image size. The linear regression of the data points is also represented.

Now that a global assessment of the method efficiency on SENTINEL-2 multispectral images has been made, it is interesting to focus on the differences between the reconstruction accuracy in the different spectral bands. It is known that the wavelength of the incident light has an impact during the measurement on several parameters such as the spatial resolution and the bandwidth of the detector as specified in Table 2. Now, the impact of these parameters during the reconstruction process may be evaluated. In Fig. 27 and Fig. 28 are represented the mean PSNR and the mean SSIM, respectively, of all images in the database, and for each spectral channel of SENTINEL-2. To reduce the reconstruction estimators dependencies to the spectral bands exclusively, only the reconstructions with a CS ratio of 25% were used. In any case, the other CS ratios showed a similar tendency.

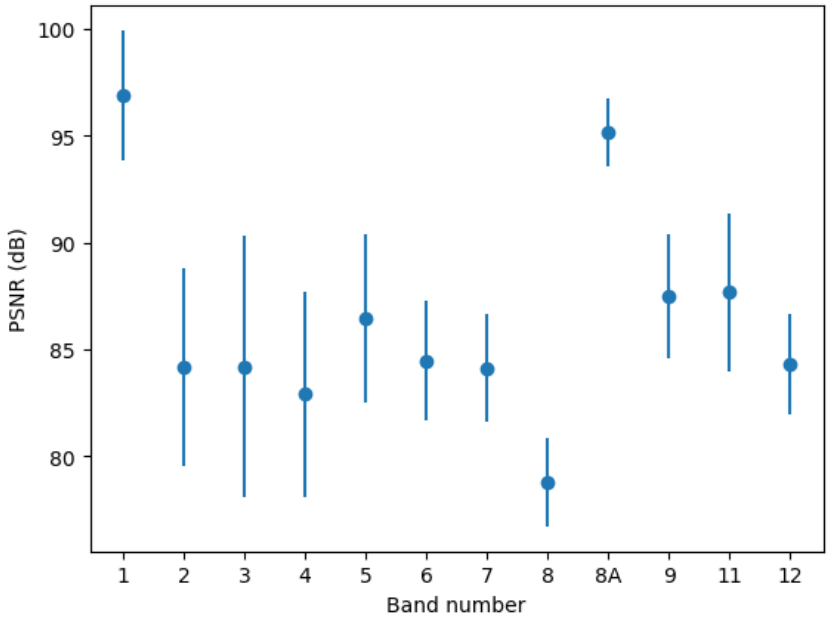


Figure 27: Mean PSNR along with the standard deviation in each spectral band.

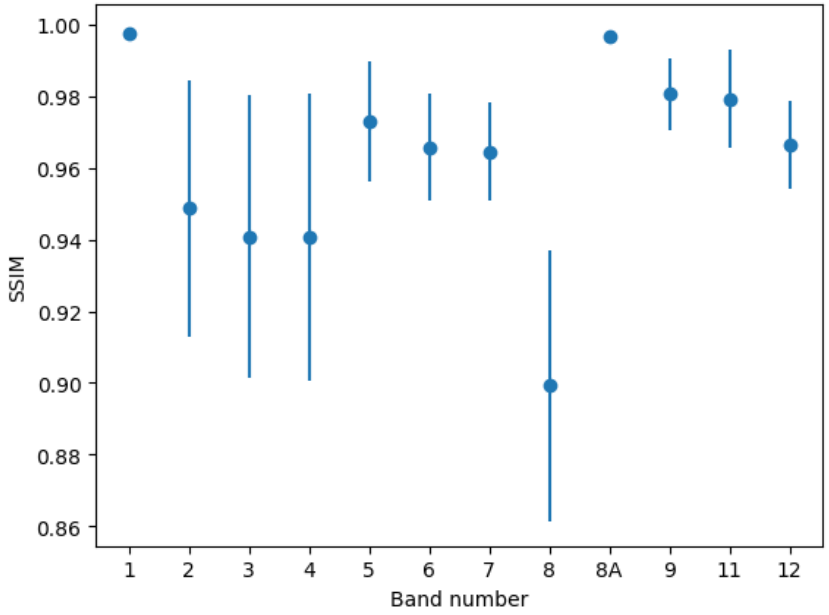


Figure 28: Mean SSIM along with the standard deviation in each spectral band.

Both estimators follow similar trends and will be treated as one in the rest of this discussion. As a reminder from Table 2, the bands are numbered in ascending wavelength order from the blue (band 1) to the MWIR (band 12). The spatial resolution is 10m for bands 2, 3, 4, and 8, 20m for bands 5, 6, 7, 8A, 11, and 12, and 60m for bands 1 and 9. Three bands stand out from the others: band 8 has a lower reconstruction accuracy while bands 1 and 8A have a higher reconstruction accuracy. The reasons are uneasy to identify. A link to the spatial resolution is most probably present. Indeed, a lower spatial

resolution means that less details are present in the image and it should then be easier for the algorithm to reconstruct an image with fewer details. It can be justified here by the fact that bands 1 and 8A have spatial resolutions of 60 and 20m, respectively, while band 8 has a spatial resolution of 10m. To properly understand this argument and the method applied, it is important to know that the images used here have the same number of pixels and cover the exact same area in all channels. The differentiation between the different spatial resolutions appears by giving blocks of pixels the same value for lower resolution bands. Nonetheless, this factor alone does not explain the results obtained here since some bands with 20m resolution show better performances than some bands with 60m resolution.

In conclusion, the pre-trained ISTA-Net$^+$ algorithm returns satisfying results and can be used without further training in the context of a first approach of an CS experimental chain. The reason for this relatively high accuracy of the algorithm is most probably that, even though it was not specifically trained on Earth observation images, the masks used are the same that those on which it was trained. Indeed, during the ISTA-Net$^+$ training, one of the learnable parameter is the initialization matrix. It is applied at the beginning of the reconstruction to avoid the system to converge towards a local minimum and it depends directly on the masks used during signal acquisition. Hence, since the masks used in this application are the same that were used during pre-training, the value of the initialization matrix is still accurate. Nevertheless, if one wants to achieve the maximum possible accuracy with this algorithm, retraining it on Earth observation images is strongly recommended. The simulation indicates that the best balance between a low CS ratio and a high reconstruction accuracy occurs at a CS ratio around 25%. Further compressing the data without retraining the algorithm would significantly degrade the reconstructed images.

Finally, to give a visual representation of the reconstruction quality, the reconstructed images of the agricultural fields image depending on the CS ratio are shown in Fig. 29. The other reconstructed images can be found in Appendix A. It is important to mention that these are RGB representations, thus only 3 out of the 12 bands are used (band 2,3, and 4) and they only partially represent the results presented above.

(a) Original

(b) Ratio 1

(c) Ratio 4

(d) Ratio 10

(e) Ratio 25

(f) Ratio 40

(g) Ratio 50

Figure 29: Results of the reconstructions of the agricultural fields image.

On the lower resolution images, some blocks of pixels appear. This is a sign of how the algorithm is working. Actually, the algorithm divides the image in blocks of 33×33 pixels, then masks and reconstructs them individually, giving rise to the observed shapes.

5 Implementation with a DMD

A CS experimental setup has been developed in a laboratory environment and is described in this section. The objective of the experiment is to create from scratch a simple but complete chain of the CS method. A complete chain means collecting light from a scene, performing the spatial modulation of the signal with a SLM, measuring the resulting signal and reconstructing the image with an algorithm. Thus, the objective is not to optimize the system but to prove the feasibility of the method with relatively simple means and create foundations for future works to improve the system. The section begins with the description of the DMD used as SLM in the setup. Then, a small digression is made about the diffraction effects that occur on a DMD and that limit the spectral range of the signal, but could also be used to turn the DMD into an efficient dispersive element. Finally, the experiment itself is extensively described.

5.1 DMD Description

A DMD is a reflective type SLM. It is a product of micromechanics and is composed of millions of micromirrors with a typical size of 10 microns. Each of them is connected to a Static Random Access Memory (SRAM) which can set the mirror in the 'on' or 'off' state, depending on the bit value. In the 'on' state the mirror is tilted at an angle of $+12°$ with respect to the flat state, and in the 'off' state it is tilted at $-12°$. Hence, by placing properly the source, detector, and optical components, this can be used to apply masks hiding precise parts of an image. In the past few decades, DMDs have been widely used in digital light processing technologies such as projectors. It has been demonstrated that DMDs have better performances than more traditional liquid crystal transimissive SLM in most applications. [36]

The SLM used in this laboratory implementation of the CS method is a DMD, and more specifically the DLP6500FLQ by Texas Instruments. It is a high resolution (1080p) DMD featuring more than two million micromirrors designed for use in the broadband visible light (300 to 700nm). It comes in two distinct parts: the DMD itself and the electronic board controller linked to it (DLPC900). The DMD is also equipped with a passive heat dissipator to cool the mirrors down. The full characteristics of the DLP6500 are provided in Table 7. [37]

Micromirror array size	1920 × 1080
Micromirror pitch (μm)	7.6
Maximum pattern rate, binary (Hz)	11574
Operating Temperature range (°C)	0 to 65
Thermal dissipation ($°C/W$)	0.7
Micromirror tilt angle	±12°

Table 7: DLP6500FLQ main parameters.

Reliable operation of the instrument requires the use of the DLP Advanced Light Control (ALC) Software Development Kit (SDK) software available on Texas Instruments' website. When connected to the DMD, the software allows to send one or several patterns to the DMD. These patterns can be set to stay a certain period of time on the DMD and they can be put in chain. Some pre-coded patterns are available with the software. They are mostly point-like or line masked, two examples are shown in Fig. 30.

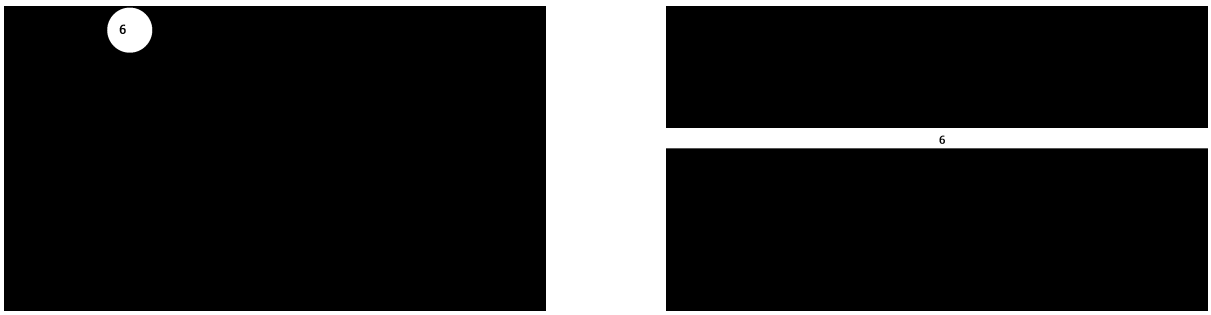


Figure 30: Examples of pre-coded masks. White pixels correspond to on state mirror and black pixels to off state mirrors.

Nevertheless, it is possible to code new masks and inject them in the software. The only requirements are that they must be binary and saved in the `bitmap` format. Other than that, the masks may be completely random or structured, depending on the needs. As an example, a mask coded such that blocks of 10 by 10 pixels are randomly set to on and off states with 50% distribution is shown in Fig. 31.

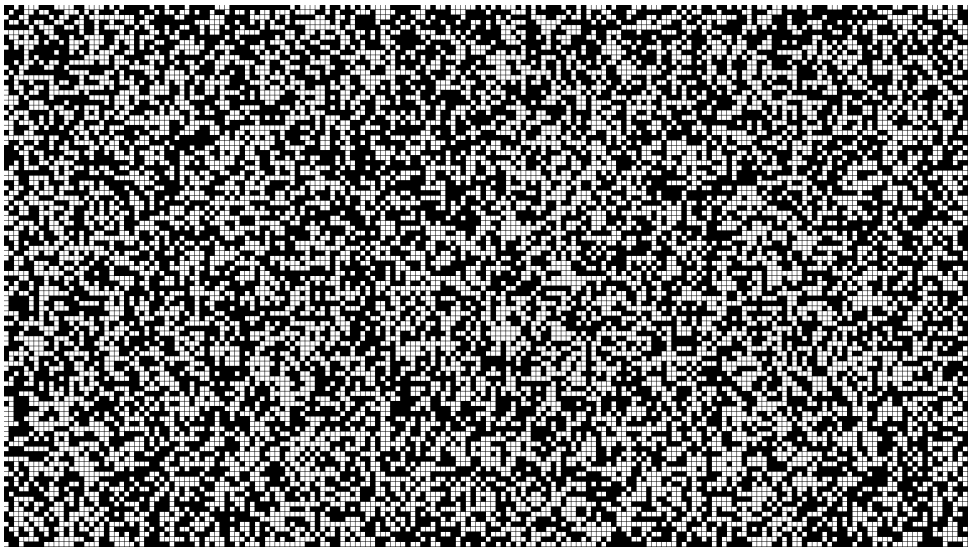


Figure 31: Random mask with 50% of on state mirrors and a binning of 10 by 10.

5.1.1 Diffraction effects

In a DMD the micromirrors size is generally of the order of a few micrometers. If the light projected onto the DMD has a wavelength of the same order, i.e. in the IR domain, diffraction will start to occur and can degrade the resulting signal. Generally the resolving power of an optical system is defined by Abbe's diffraction limit which states that the minimum distance at which two objects can be resolved is $d = \frac{\lambda}{2NA}$ with NA the numerical aperture. The minimum distance to be resolved is the micromirror pitch so that each micromirror can be distinguished. As a general rule, the Rayleigh criterion states that one should work at a wavelength not longer than around one tenth of the minimum distance to be resolved. Hence, considering the $7.6\mu m$ micromirror size of the DLP6500FLQ, the incident wavelength should not be larger than 700 nm to avoid image degradation by diffraction effects. Thus, the light source is roughly limited to the visible spectrum.

Another diffraction effect to consider is due the global shape of the DMD. As a matter of fact, a DMD can be seen as a blazed grating if all the micromirrors are in the same state [38]. As a reminder, a blazed grating is a grating designed such that its grooves have a specific angular orientation. This way, the diffracted light intensity is optimized in a specific direction. When a light ray is reflected in this specific direction, the blaze condition is said to be respected. A model of DMD seen as a grating is represented in Fig. 32. Following this model, the lines of the grating are parallel to the micromirror hinges, which are at 45° to the DMD edges. The micromirrors act as blazed facets of the grating, at angle ϕ to the DMD normal, i.e. the tilt angle of the mirror about the hinge axis.

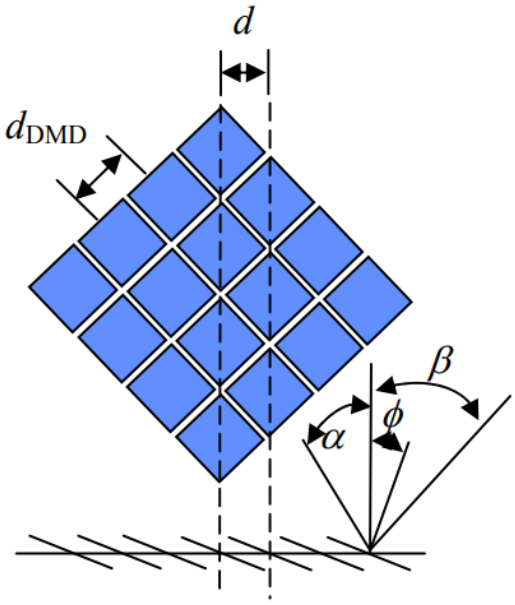


Figure 32: Description of the DMD modeled as a blazed diffraction grating. The top part of the figure depicts a section of the DMD as viewed normal to the DMD with the micromirrors in the flat state. The bottom part depicts an edge view of the same DMD section along the micromirror hinge direction (dashed lines) with all micromirrors in the on state. From [38].

As a reminder, the equation for a one-dimensional grating is

$$m\lambda = d(\sin\alpha + \sin\beta), \tag{50}$$

with m the diffraction order, λ the wavelength, d the grating pitch, α the incidence angle to the DMD normal, and β the diffraction angle to the DMD normal. Considering an angle of incidence $\alpha = 0$ the equation reduces to

$$m\lambda = d\sin\beta. \tag{51}$$

Moreover, the blaze condition occurs when the facet normal bissects the angle between the incident and reflected rays, i.e. when the incident and diffracted rays follow the law of reflection from the facet point of view. The blaze condition defines the angle at which the diffraction is the most efficient in a blazed grating [39]. Hence, for a grating under normal incidence illumination, the blaze condition is satisfied when $\phi = \beta/2$. Therefore, the blaze wavelength λ_B which defines the wavelength at which the diffraction grating efficiency is maximum (under the blaze condition) is given by

$$\lambda_B = \frac{d}{m}\sin 2\phi. \tag{52}$$

One can see from Fig. 32 that $d = d_{DMD}/\sqrt{2}$, where d_{DMD} is the DMD micromirror pitch.

Adapting the procedure described in [38] with the characteristics of the DMD used in this work and specified in Table 7, it is possible to estimate the spectral region of efficient diffraction for each diffraction order. The region of efficient diffraction for a blazed grating typically extends from about 2/3 to 3/2 of λ_B, as a rule of thumb. Hence, using this rule of thumb and Eq. (52), the spectral ranges of efficient diffraction have been computed for each order and are presented in Table 8.

Diffraction order	Lower blaze wavelength (nm)	Blaze wavelength (nm)	Upper blaze wavelength (nm)	Spectral range
1	1457	2185	3278	SWIR
2	728	1092	1639	NIR/SWIR
3	485	728	1092	VNIR
4	364	546	819	VNIR
5	291	437	655	VIS
6	242	364	546	UV/VIS
7	208	312	468	UV/VIS
8	182	273	409	UV/VIS

Table 8: Spectral ranges of good efficiency up to the 8th diffraction order as predicted from the blazed grating model.

The angles at which the lower blaze wavelength and the higher blaze wavelength diffraction occurs, β_L and β_U respectively, can be computed using Eq. (52). They are identical for each order of diffraction: $\beta_L = 15.73°$ and $\beta_U = 37.6°$. As a reminder the blaze angle is $\beta = 2\phi = 2 \times 12° = 24°$

These diffraction effects must be taken into account as they can degrade the image quality. In their experiment, Rice et al. [38] used a DMD with a micromirror pitch of 13.68 μm leading to blaze diffraction wavelengths ranging from the MWIR region at the first order to the shorter parts of the visible spectrum at order 13.

Although this diffraction effect can degrade the image, it can also be extremely interesting in the case of multispectral or hyperspectral imaging. Indeed, the diffraction properties of the DMD enable it to be used simultaneously as a SLM and as a dispersive element, providing both spatial and spectral differentiation. The idea has been exploited by Ebner et al. [40] to demonstrate the functionality of a hyperspectral MWIR single-pixel microscope. The results showed that the sample throughput drastically improved and the acquisition time for hyperspectral cubes was around 50s, outperforming conventional microscopes by orders of magnitude. The method has proven to be a fast and cost-effective improvement for hyperspectral MIR microscopy and shows great potential for transfer into other disciplines including Earth observation. However, it is important to note that the 50s acquisition time reached in this case could be too large for LEO remote sensing due to the satellite motion mentioned previously. This time may vary considerably when working at other wavelengths, with other detectors, and at various spatial resolutions.

The method should thus be investigated by tuning these parameters in the case of Earth observation.

5.2 Laboratory experiment

5.2.1 Setup

The objective of this laboratory implementation is to achieve a complete CS chain, from signal acquisition to image reconstruction. The point is not to obtain the maximum image quality or compression but to put in place a first experiment to show the feasibility and limitations of the technique. This could then be used as starting point in a future work with the objective to tune the method and extend it to a more space-specific context.

The general scheme of the experiment is based on most standard CS laboratory implementations and is shown in Fig. 33.

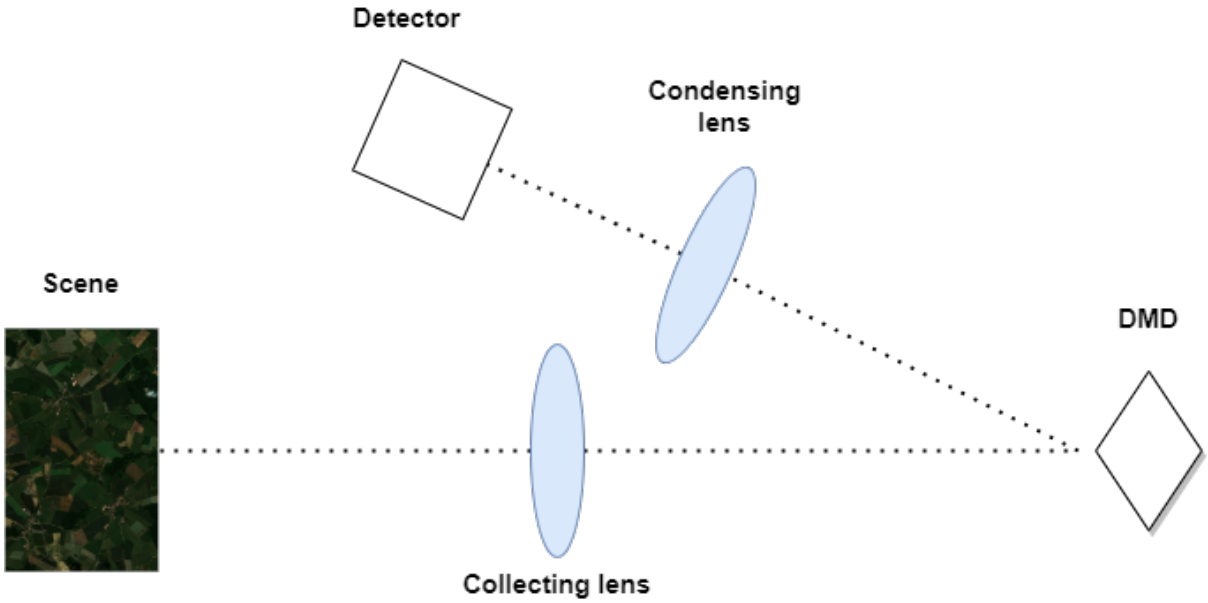


Figure 33: Global scheme of the experiment.

The system may be divided in two subsystems: the collecting part which images a scene on the DMD and the detecting part which focuses the light reflected by the on state mirrors of the DMD on a detector. The first part starts with a scene which is here a digital screen. In continuation of this work, an Earth observation image is chosen to be the scene. Therefore, a self-illuminated digital screen is the best choice to display these kind of images, provided that it is bright enough. The size of the screen has to be chosen adequately so that it can be imaged on the DMD with the available material. To do so, it is interesting to review some basic optics and define some notations. The following figure represents the case of a thin lens in an optical system.

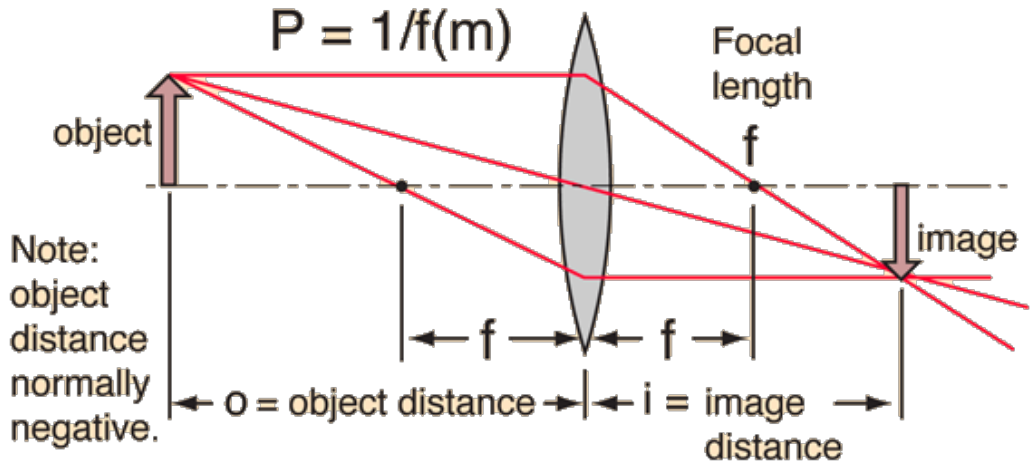


Figure 34: Thin-lens system. [41]

Using these notations, the thin lens equation can be written

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f}. \tag{53}$$

The objective is to determine the lens and the size of the object needed to form a sharp image on the DMD. Two limiting factors must be taken into account. First, the total size of the system is limited by the size of the laboratory table which is approximately 1.7 meters long. Second, as it will be explained in details later, the collection and detection arms of the system are at an angle of 24° to each other. Therefore, the collecting lens should not be too close to the DMD so that it is not in the detector's field of view. To determine the object and image distance to place the object and the DMD, the magnification needs to be considered. It is defined as the ratio between the image size and the object size, but is also equivalent to the ratio of the image and object distance[4]:

$$M = \frac{h_i}{h_o} = \frac{i}{o}. \tag{54}$$

By knowing the magnification value, one can determine the relation between i and o, and compute these distances depending on the focal length of the lens considered using Eq. (53). The DMD dimensions are 14.51mm × 8.16mm. At first, a computer screen was considered to be used as an object, its dimensions are 475mm × 270mm. This yields a magnification of 0.03 regardless of the dimension considered. Hence, $i = 0.03o$, and injecting this into Eq. (53) gives $o = 34.33f$. The next table shows values of the different distances for several focal lengths available.

[4]NB: Considering prior knowledge that both object and image are real, absolute values of the sizes are considered to not bother with sign conventions.

f **(mm)**	o **(mm)**	i **(mm)**	$o + i$ **(mm)**
25	858.33	25.75	884.08
50	1716.67	51.5	1768.17
100	3433.33	103	3536.33
150	5150	154.5	5304.5
300	10300	309	10609

Table 9: Values of the image and object distances for several focal lengths for a computer screen.

It can be seen that the total length of the system is larger than the length of the table for focal lengths of 50mm and higher. As for lower focal lengths the image distance is too short. Indeed, it has been decided to put the lens at at least 10cm of the DMD to avoid being in the camera's FOV. Thus, a smaller screen is needed. In the end, a phone screen was chosen due to its relevant size and brightness. The dimensions of the image displayed on the phone screen is 102mm×64mm. Using this and the DMD dimensions to compute the magnification and injecting it in Eq. (53) as before, one can compute the different distances for several focal lengths:

f (mm)	o (mm)	i (mm)	$o + i$ (mm)
25	248.46	27.8	276.26
50	496.93	55.59	552.52
100	993.85	111.19	1105.04
150	1490.78	166.78	1657.56
300	2981.56	333.56	3315.13

Table 10: Values of the image and object distances for several focal lengths for a phone screen.

Based on the criteria described earlier, both the 100 and 150mm focal length lenses are suitable given their large enough image distance and not too large total distance. In the end, the 150mm lens was chosen because it has a larger aperture, thus collects more light and increase the radiometric resolution of the system.

Now that the image is formed on the DMD, it is necessary to properly design the detection arm of the system. It is composed of the DMD, a detector, and between them a condensing lens, or more precisely a camera lens. The detector is a UI-3240CP Rev. 2 camera [42]. It is a CMOS monochrome detector with sensitivity in the visible and NIR spectra. It has a resolution of 1280 × 1024 pixels for dimensions of 6.784mm × 5.427mm. To determine which camera lens is best in this case, the lens and magnification equations can once again be used, this time with the DMD as the object and the camera as the image. A few values for available focal lengths are presented in the following table.

f (mm)	o (mm)	i (mm)	$o+i$ (mm)
16	50.22	23.48	73.7
35	109.86	51.36	161.22
50	156.94	73.38	230.32
75	235.41	110.07	345.48

Table 11: Values of the image and object distances for several focal lengths to image the DMD on the camera.

Here, there are two main restrictions. The first is that the the object distance must be large enough to not obstruct the optical path of the collecting arm. The second is that the image distance must not be too large since the camera lens is attached to the camera and the distance lens-detector can only be slightly adjusted by adapting the focus of the objective or by adding adaptor rings. Therefore, a 75mm lens is chosen since it fits best these two characteristics. Furthermore, it must be noted that the camera lens is actually a system composed of two groups of lenses and the announced focal length is the effective focal length. However, to determine precisely where to place the camera lens, one has to determine the principal planes of the lens for the object and image sides. The principal plane for the object side is the principal object plane, corresponding to the focal plane of the front lens group. In the same way, the principal plane for the image distance is the principal image plane, corresponding to the focal plane of the rear lens group. These principal planes locations can be roughly estimated by passing a collimated beam through each side of the camera lens, observing where the beam is best focused, and reporting the focal length from this point towards the camera lens. A representation of the camera lens and the estimated principal planes is shown in the next figure.

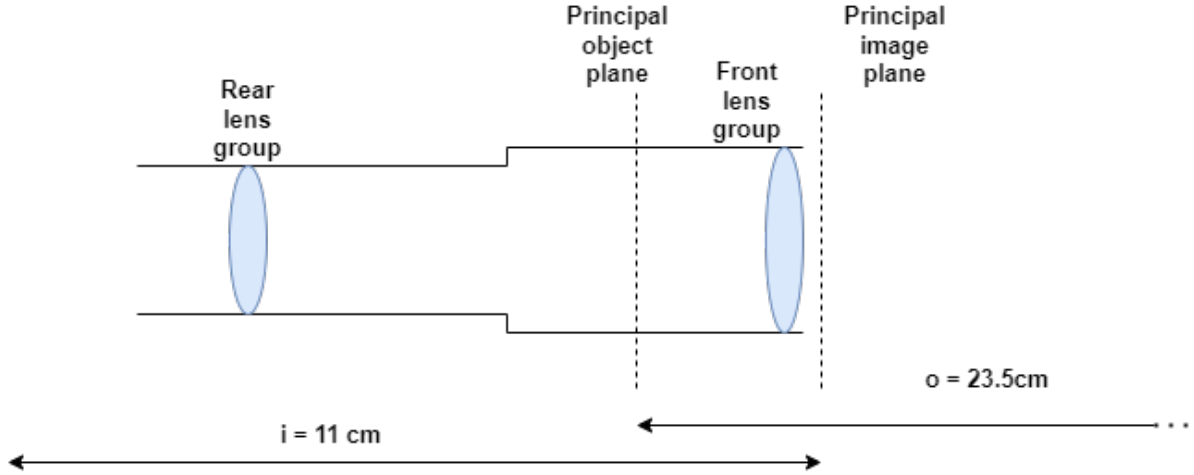


Figure 35: Representation of the camera lens with the estimation of the principal planes.

It is important to note that the distances indicated here are estimations and slight adjustments are made during the camera lens placement to optimize the focus.
Finally, it is necessary to know in which direction to place the camera with respect to the

collecting arm. The goal is to place the camera so that it detects the light reflected by the on state mirrors of the DMD. If the object is placed in the normal direction of the flat state of the DMD, the angle at which to place the camera can be determined using Fig. 36

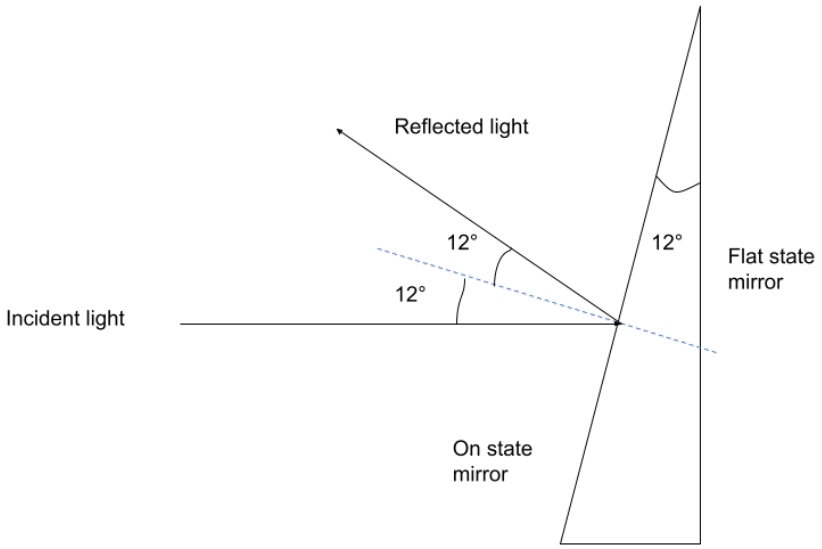


Figure 36: Optical scheme of the reflection occurring at one DMD micromirror.

Indeed, as indicated in Table 7, the tilt angle of the on state mirror is 12°. Therefore, the incident light arriving perpendicularly to the flat state forms a 12° angle with the normal of the on state mirror. Then, the reflection law states that the reflected ray forms an angle equal to the incident angle with the normal. Hence, the total angle between the collecting and detecting arm is 24°. An important note is that the rotation axis of the mirrors is diagonal to the DMD. This means that this 24° should be measured with respect to the plane diagonal to the DMD. As a consequence, the detector will be shifted and elevated with respect to the collecting system. To do so and still having the DMD in the camera's FOV, the camera lies on a ball joint allowing a flexible adjustment of the detector's direction. The final experimental setup is shown in Fig. 37.

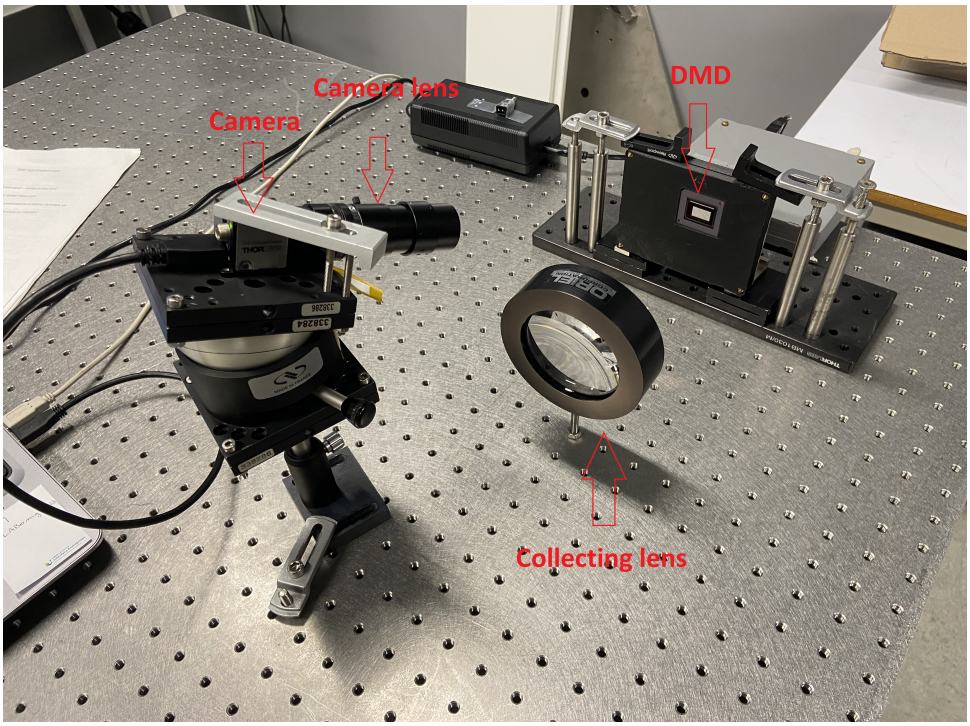


Figure 37: Final experimental setup. The screen used as a scene is not shown here and is further away towards the bottom-left corner.

5.2.2 Calibration

Before starting to take masked measurements, it is interesting to see what the camera detects when a scene is imaged with all the mirrors on the on state. In continuity of this work, an Earth observation image is used as a scene. More precisely, the agricultural fields image (Fig. 23c) is used. As a reminder, the camera is monochromatic thus the resulting image is in grayscale. The result is shown in Fig. 38.

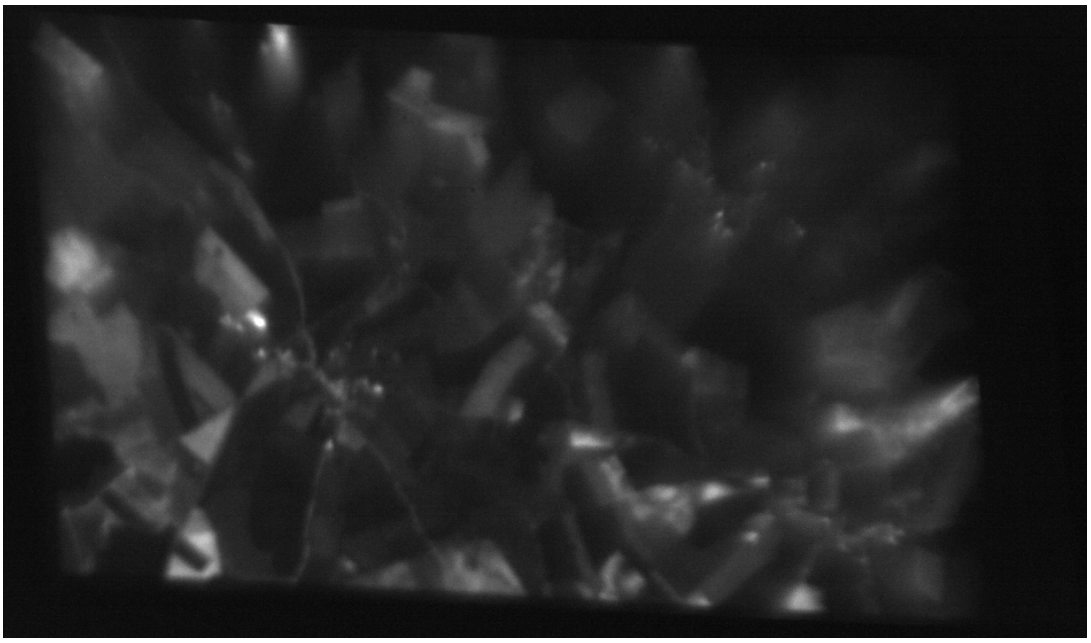


Figure 38: Rendering of the agricultural fields image by the optical system.

It can be seen that the image is relatively dark and it is difficult to distinguish the details. All of the SENTINEL-2 images being dark, the brighter laboratory picture shown in Fig. 37 is used as a scene to have a better view of the system capabilities. The result is shown in Fig. 39.

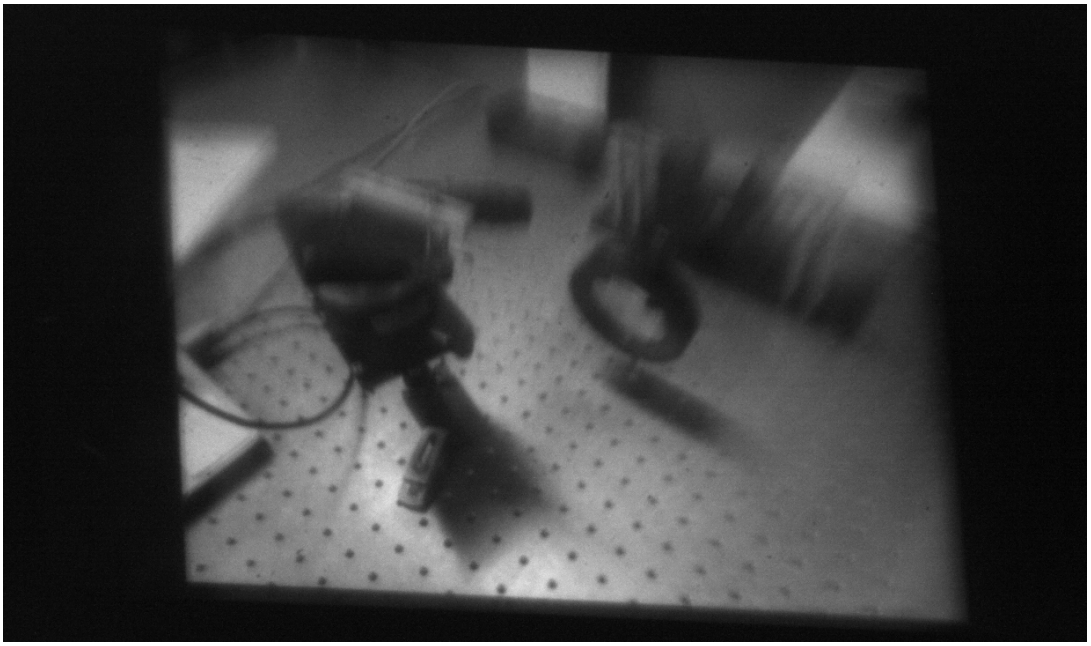


Figure 39: Rendering of the laboratory picture by the optical system.

Since this image is clearer, it will be used in the rest of the experiment. Several things can be noted. First, the image is inclined in the camera frame. This is one consequence of the inclination of the camera due the rotation axis of the mirrors being in the diagonal

plane of the DMD as explained before. The camera has been rotated to correct this effect as best as possible but it is not sufficient to completely correct it. In addition, this leads to edges containing empty pixels (where no light is detected). Note that the frame of the camera has been slightly cropped to remove as much empty pixels as possible, thus not every pixel of the camera is represented. It is interesting to assign a zero-value to these empty pixels on each observation to be sure that they do not degrade the image quality by introducing some noise in the reconstruction due to potential stray light. To do so a flat field image is taken by imaging an uniformly white scene.

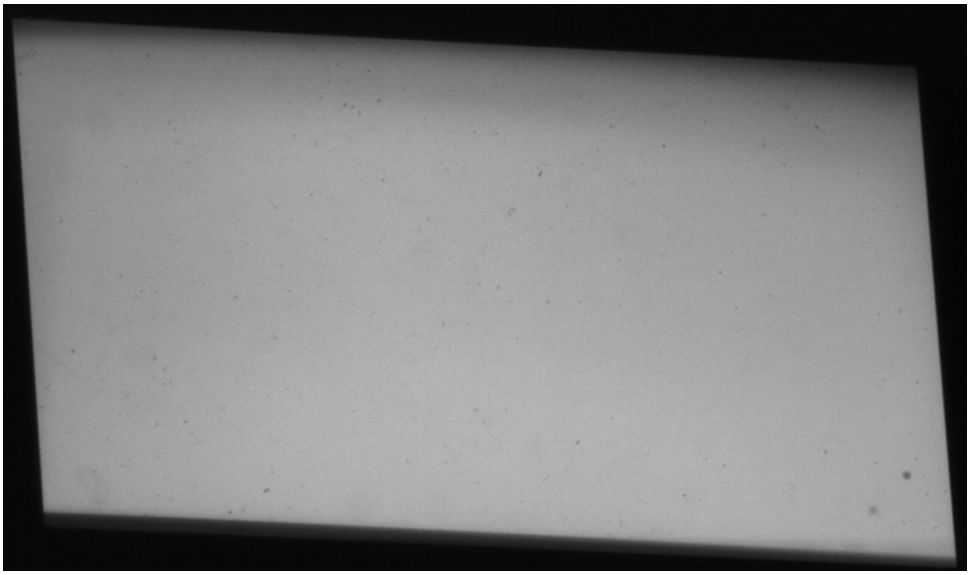


Figure 40: Flat field

It can be seen that some points appear darker most probably due to dust on the lens or the DMD. Nevertheless, it does not have any impact here. To remove the edges, a binary threshold is applied such that for each pixel value $p(x, y)$

$$p(x,y) = \begin{cases} 1 & \text{if } p(x,y) > \text{threshold} \\ 0 & \text{if } p(x,y) < \text{threshold} \end{cases} \tag{55}$$

After applying this with a proper threshold value, the edge mask is formed.

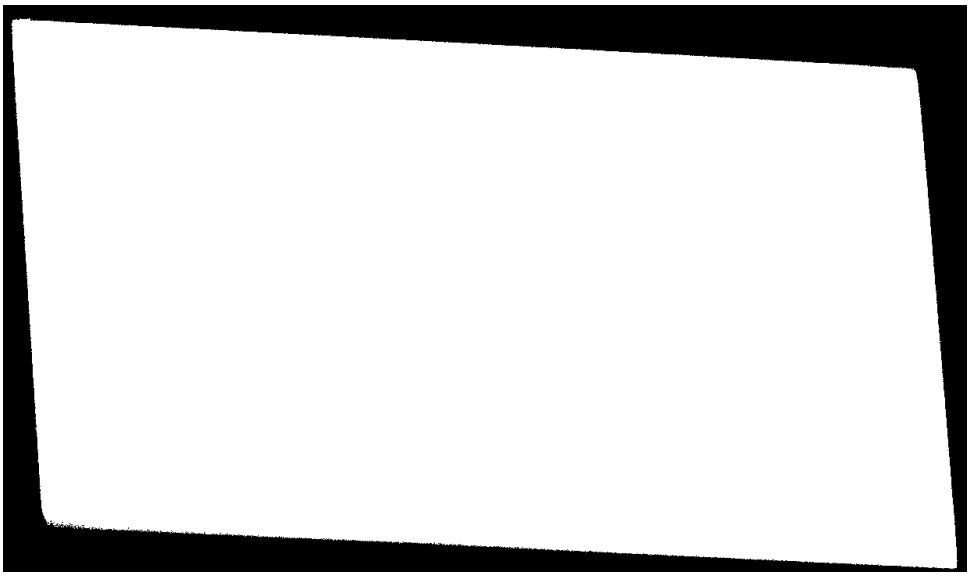


Figure 41: Edge mask

In the end, this is what the image looks like after applying this edge mask.

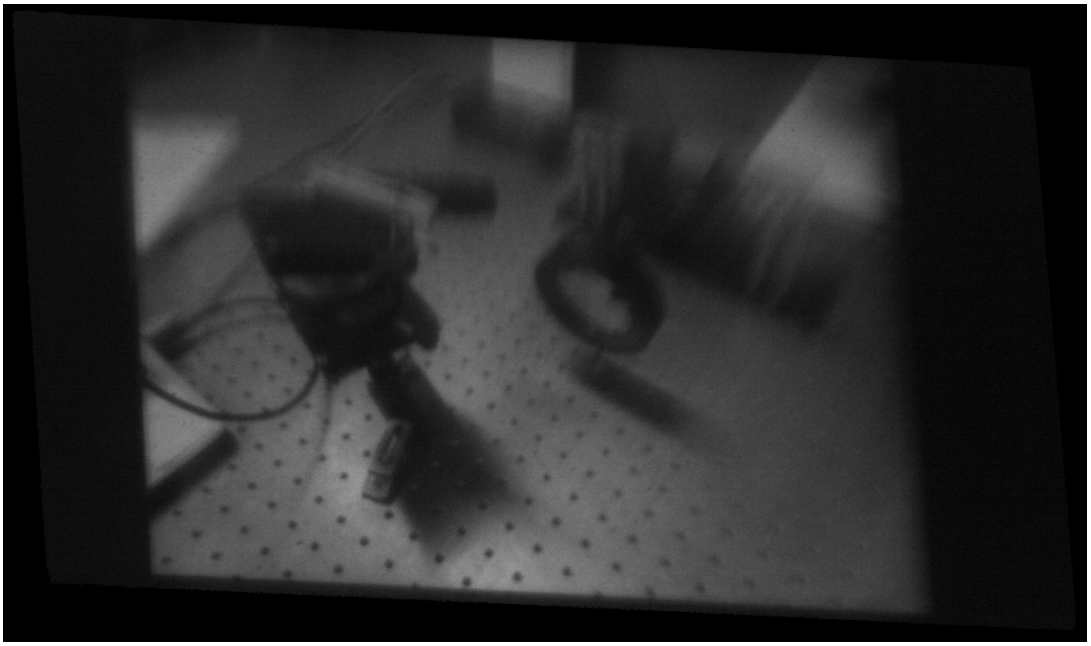


Figure 42: Measurement of the laboratory image after applying the edge mask.

Some pixels appear slightly brighter than the edges on the right and left side of the picture. This is because the image formed by the collecting arm does not cover the whole surface of the DMD. Thus, the slightly brighter pixels correspond to the mirrors of the DMD non illuminated by the image but still sending a low amount of stray light towards the camera.

Last but not least, only a part of the image seems to be properly focused on the camera. Indeed, while the bottom half of the image is relatively clear, the top half is

blurry. This is a direct consequence of the inclination of the camera with respect to the DMD. In close-range imaging, a limiting factor is often the Depth Of Field (DOF) of the system, i.e. the range of distances in a scene that appear acceptably sharp in an image. According to Ray (2002) [43], it can be computed according to equation

$$DOF = \frac{2 \cdot f^2 \cdot u^2 \cdot F_{\#} \cdot COC}{f^4 - F_{\#}^2 \cdot COC^2 \cdot u^2}, \tag{56}$$

where f is the focal length, u the focusing distance (object distance), $F_{\#} = f/D$ the f-number, and COC is the Circle Of Confusion. The COC is a term often used in photography which is originally defined as the finest detail visible by the human eye. It can be viewed as an alternative definition of the resolution. In the case of digital cameras, the COC is usually taken around 1-3 pixels [44]. Hence in this case, the camera pixels having a size of 5.3 μm, a COC of 10 μm is considered as a fine approximation. The focal length is 75mm, the object distance is 235mm, and the f-number is 3.9. This yields a DOF of 0.7mm. This value is very small and explains why the small inclination of the object plane (the DMD) with respect to the camera and the lens induces a partial blurring of the image. Indeed, given the inclination, the top and bottom of the object are at a different distance from the lens. The difference between these distances is very likely to be larger than the DOF of 0.7mm, giving rise to the partial blur.
Generally, one can increase the DOF by reducing the aperture of the lens. That way, the rays that pass through the lens are more parallel and objects focus over a wider distance. Nevertheless, it comes with a decrease of the amount of light collected and thus of radiometric resolution. In this case, reducing the aperture reduces the radiometric resolution too much and does not improve the DOF enough to be a viable solution. Another solution can be found using the Scheimpflug principle which states that by shifting the lens so that its optical axis is no longer perpendicular to the sensor, the DOF can be optimized.

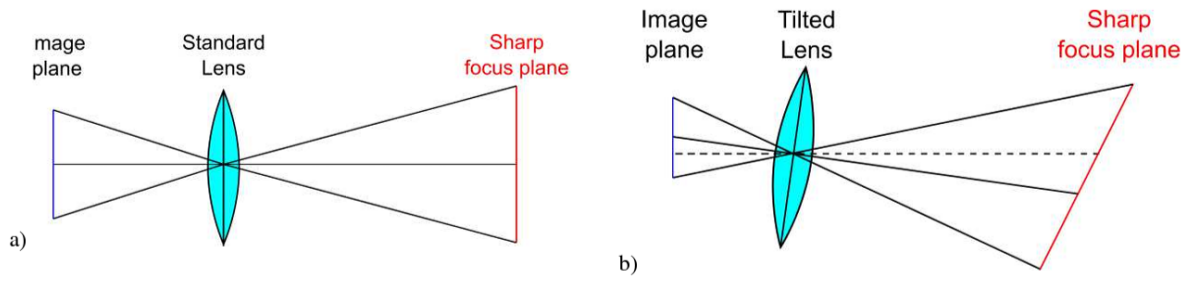


Figure 43: With an ordinary camera lens the plane of focus is parallel to the image plane (a). Scheimpflug principle: plane of focus of an optical system tilts when the lens plane is not parallel to the image plane (b). From [44].

The DOF is said to be optimized when the image plane, the lens plane, and the sharp focus plane meet in the same point [44]. Therefore, having a sharp focus on the whole

image requires to tilt the lens at the right angle so that the object plane corresponds to the sharp focus plane. Practically, it means using a tilt-shift lens, some can be found on the market, but none was available in the laboratory and they are expensive. Another possibility is to do this shifted lens by hand, by decoupling the lens from the camera and tilting it. However, this requires high precision and a complex structure to maintain all the optics at the suitable position and angle. This structure is even more complex due to the fact that the system is already placed at a peculiar angle and position. Therefore, it has been decided that this system is satisfactory for a first implementation of CS in the laboratory since the blurred image can still be masked and reconstructed. Nonetheless, future works will have to solve this focusing problem on the camera. It must be noted that during a private discussion with prof. Laurent Jacques (UCLouvain) he mentioned the possibility to account for the distortion effects in the forward model of the system.

5.2.3 Masking

Now that the rendering of the optical system when all the mirrors are in the on state is known, masks have to be applied to spatially modulate the signal. To apply them properly and be able to inject them in a reconstruction algorithm, an important parameter needs to be considered first. The mask applied on the DMD does not exactly correspond to the resulting mask on the camera. Indeed, the DMD modulates the signal with a certain number N of micromirrors, i.e. N pixels. However, the camera contains a number $\bar{N}$ of pixels, with $\bar{N} \neq N$. Thus, the mask will not hide the same number of pixels on the camera as it does on the DMD. To determine the size and shape of the resulting mask on the camera, the same technique that was used for edge calibration is applied. A flat field measurement is done with the difference that this time a mask is applied on the DMD. But before that, an appropriate mask must be coded and implemented on the DMD. As explained in section 2.1, a random mask should satisfy the Restricted Isometry Property (RIP) as well as the incoherence property. Hence, a binary mask was coded by randomly setting each micromirror state on on or off with a probability of 50%, yielding a mask with a CS ratio of 50%. The result of the flat field measurement with this pattern is shown in Fig. 44.

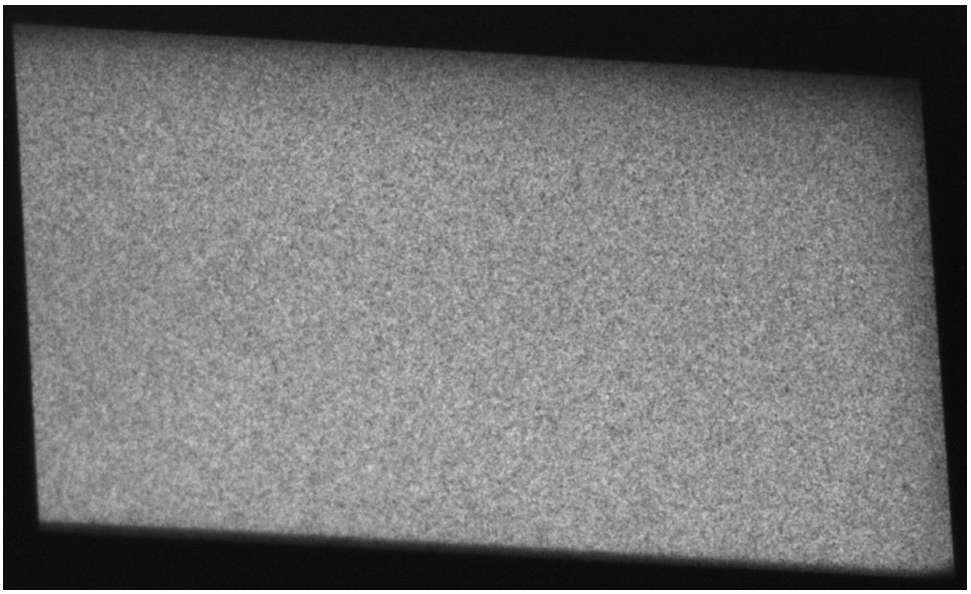


Figure 44: Flat field measurement with a random binary mask with a CS ratio of 50%.

Then, the idea is to apply a binary threshold, in the same way that for the edge mask, to identify which pixels of the camera are attenuated enough to be considered on the off mode. The results of this threshold are presented in Fig. 45. Note that the edge mask has also been applied.

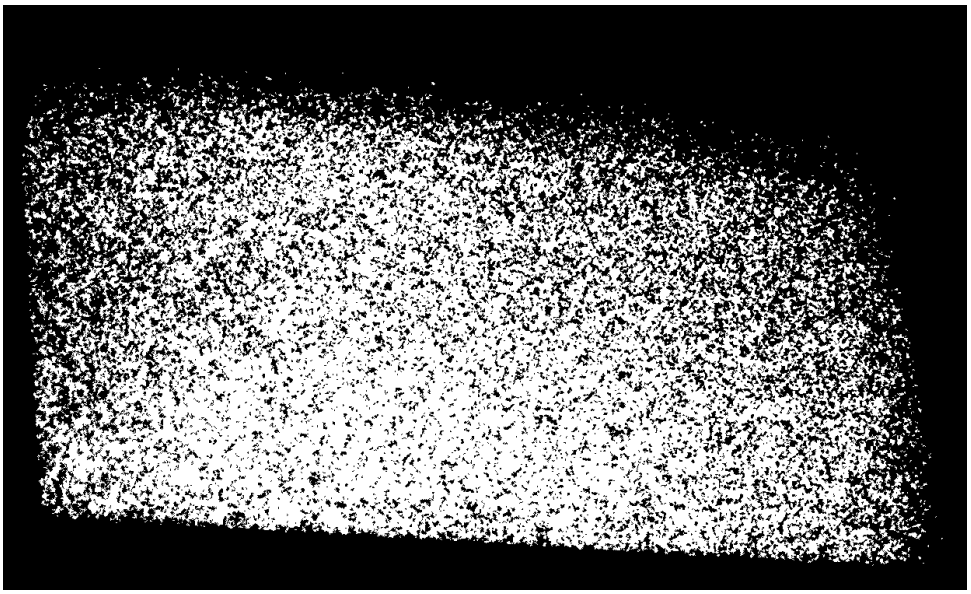


Figure 45: Corresponding mask on the camera with a CS ratio of 50%.

It can be seen that most of the illuminated pixels are located around the bottom and center parts of the image, whereas they are suppose to be uniformly distributed. Thus, it does not accurately represent the shape of the original mask. That is because the micromirrors are small which lead to diffraction effects attenuating the effect of the mask. In addition, the blurring of the image that is observed in Fig. 42 disturbs part of the

signal. These facts combined make it difficult to discriminate between the masked and unmasked region of the image. To counter this effect, a binning of the pixels can be added when coding the pattern. Therefore, a new mask was coded using a binning of 10×10 and a CS ratio of 50%. The same processing that for the previous pattern was applied and the result is shown in Fig. 46.

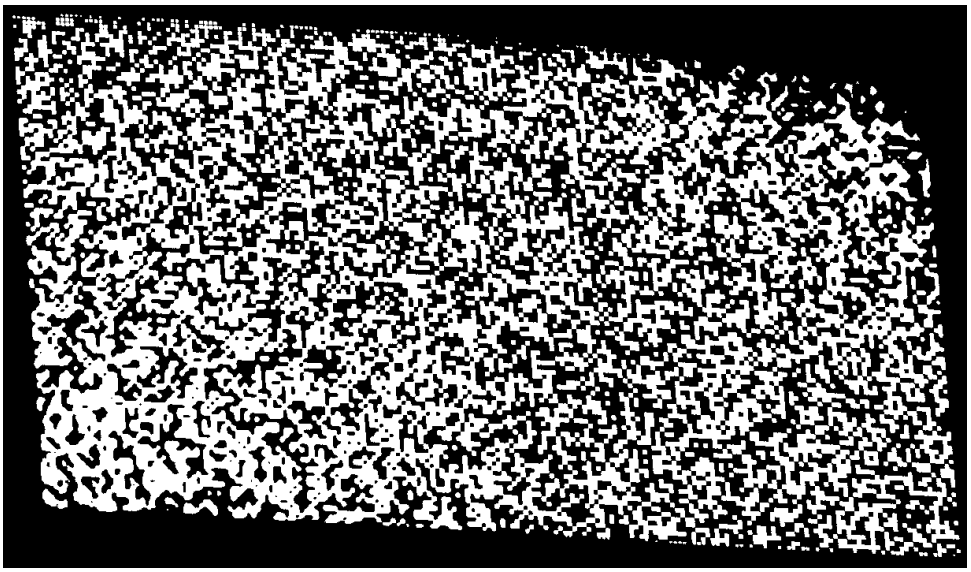


Figure 46: Corresponding mask on the camera with a CS ratio of 50% and a 10×10 binning.

There, the masked pixels are more clearly defined, thus the mask is better represented. However, some distortion effects are still visible in the bottom left and top right corners. Now that the corresponding mask on the camera is known, one can image a real scene with this mask. Once again, the laboratory image is used and the mask described above with a CS ratio of 50% and a binning of 10×10 is implemented on the DMD. The resulting image on the camera is shown in Fig. 47.

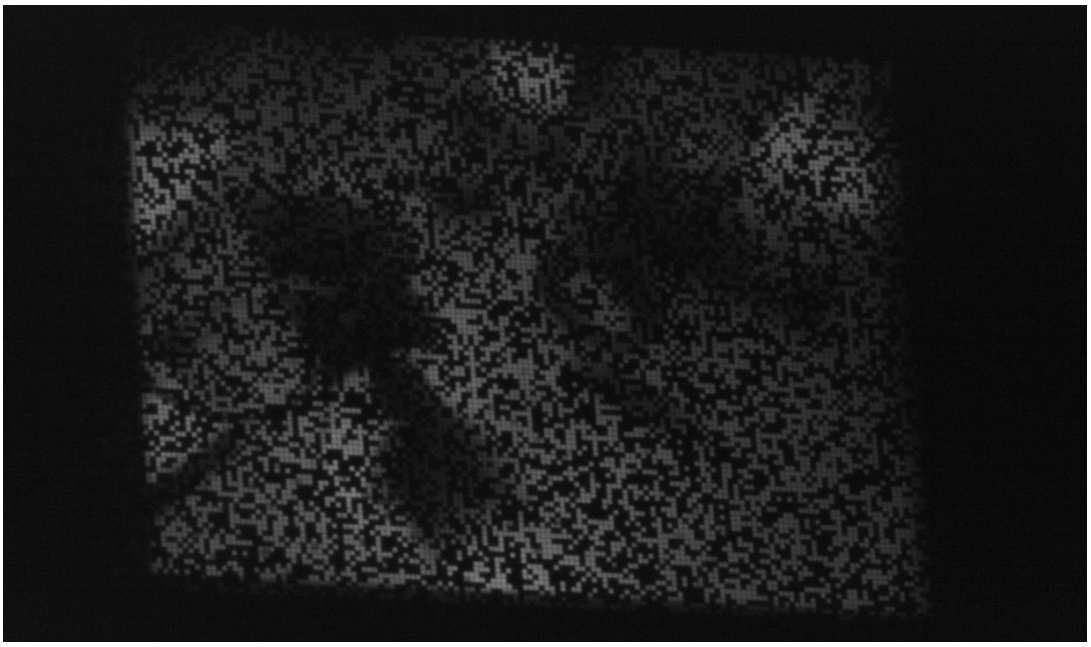


Figure 47: Rendering on the camera of the laboratory image when a mask with a CS ratio of 50% and a 10×10 binning is applied on the DMD.

In order to avoid noise on the pixels supposed to be in the off state, the mask defined in Fig. 46 is applied on the measurement. The resulting image is shown in Fig. 48.

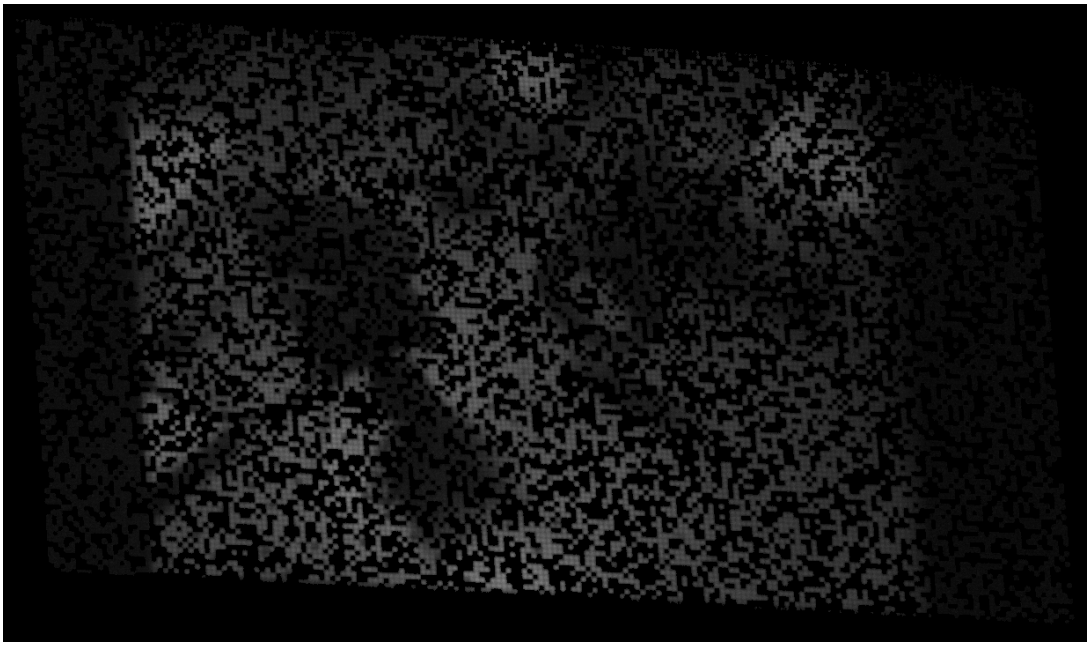


Figure 48: Result of the element-wise product of Fig. 46 and Fig. 47

To summarize, the masked measurement that will be used to build the measurement vector $\boldsymbol{y}$, as well as the mask itself that will be used to construct the measurement matrix $\boldsymbol{\Phi}$ have both been defined. This pre-processing step is necessary to properly know the mask, and the resulting measurement on the camera. It can be repeated with any other mask or scene.

5.2.4 Reconstruction

The first idea for this experiment was to inject the measurements in the ISTA-Net$^+$ reconstruction algorithm, considering the satisfactory results obtained during the simulations on Earth observation data. However, some complications occurred. To understand them, it is necessary to get back to the initial inverse problem. As a reminder, in a CS system the inverse problem to solve is

$$\underbrace{\boldsymbol{y}}_{\in \mathbf{C}^m} = \underbrace{\boldsymbol{\Phi}}_{\in \mathbf{C}^{m \times N}} \underbrace{\boldsymbol{x}}_{\in \mathbf{C}^N}, \tag{57}$$

with $\boldsymbol{y}$ the measurements, $\boldsymbol{\Phi}$ the measurement matrix, $\boldsymbol{x}$ the original signal, and $m < N$. Considering this, it must be highlighted that $\boldsymbol{\Phi}$ does not necessarily exactly correspond to the mask applied on the image. Indeed, in the case of multiple measurements, each line of $\boldsymbol{\Phi}$ corresponds to the mask used to take the measurement. If the mask is two-dimensional, it must thus be reshaped into a one-dimension vector. The same reshaping must be applied to the measurement and signal vector if necessary. For example, in the case of a single-pixel camera, each element y_i of the measurement vector corresponds to one measurement made with a different mask. Hence, each line ϕ_i of $\boldsymbol{\Phi}$ represents the mask used to measure y_i. However, the present application is not a single-pixel camera and several pixels are measured with the same mask. Therefore, the construction of $\boldsymbol{y}$ and $\boldsymbol{\Phi}$ is not straightforward. As a first illustration, one may consider the simple of case of trying to reconstruct the image with measurements made with the same pattern, e.g. with the mask defined in the previous section. A certain number m of pixels are illuminated on the camera, yielding m measurements. Following the same logic, these m measurements should be placed in a one-dimensional vector to form the measurement vector $\boldsymbol{y}$. And, since all these pixels have been measured with the same mask, each line of the measurement matrix $\boldsymbol{\Phi}$ will be identical. The problem is that, during the reconstruction, the ISTA-Net$^+$ algorithm needs to compute the inverse of $\boldsymbol{\Phi}$ at some point. But, if all the lines of $\boldsymbol{\Phi}$ are identical, the matrix is singular and thus not invertible, leading to a mathematical error. The problem still occurs when multiple measurements are made since multiple pixels will still be measured with the same masks, thus some lines will be identical, and $\boldsymbol{\Phi}$ will be singular. Some potential solutions can be investigated: a single-pixel camera could be artificially made by two different ways. First, by applying masks so that only one pixel on the camera is illuminated. This way, only one measurement would be taken by mask and each line of $\boldsymbol{\Phi}$ would be different. However, it is difficult to code such masks since the relation between the pixels on the DMD and the pixels on the camera is not straightforward. Second, by keeping the same kind of masks but integrate all the illuminated pixels. This would yield a single measurement for each masks. Nevertheless, the read-out noise of each pixel being slightly different, it would introduce some additional

noise in the measurement. Moreover, both of these possibilities require taking a lot of measurements. Indeed, given that a binning of 10 by 10 is applied on the DMD, it results in a resolution of $108 \times 192 = 20736$ pixels. For a single-pixel camera architecture to reconstruct the image with this resolution and a CS ratio of 50% would require taking 10368 measurement with different masks. While this is no strange process in CS, it would require to optimize the measurement process and data pre-processing to take so much measurements in an acceptable time.
In addition to these difficulties, it is important to remember that ISTA-Net$^+$ is a DL algorithm. Therefore, several parameters such as the sparsification operator are optimized during the learning process. This optimization depends directly on the masks implemented and on the training data. The algorithm as it is delivered works on sub-blocks of 33×33 pixels. More precisely, it simulates the CS acquisition process by separating the image in blocks of $33 \times 33 = 1089$ pixels. Then, for each block, it reshapes it in a one-dimensional vector and apply one-dimensional masks on it. Therefore, the algorithm is trained with masks and data of this same specific size, which can introduce some errors when new masks of different sizes are used. Moreover, the patterns that have been used to train the algorithm are gaussian masks, which are more representative of transmissive SLMs such as LCDs where each pixel can absorb or transmit a different proportion of the light. Therefore, it is not representative of this experimental case which uses binary patterns.

For all the reasons stated above, the use of the ISTA-Net$^+$ algorithm for this experiment was put aside at the profit of a more simple one called inpainting using IHT and described in section 2.3.3. The laboratory image masked as described in the previous section was injected into this algorithm. The result of the inpainting after 500 iterations and a decaying threshold value form 1 to 0 by steps of 1/500 is shown in Fig. 49. As a reminder, the full image is displayed in Fig. 50 and the masked image injected into the algorithm in Fig. 51.

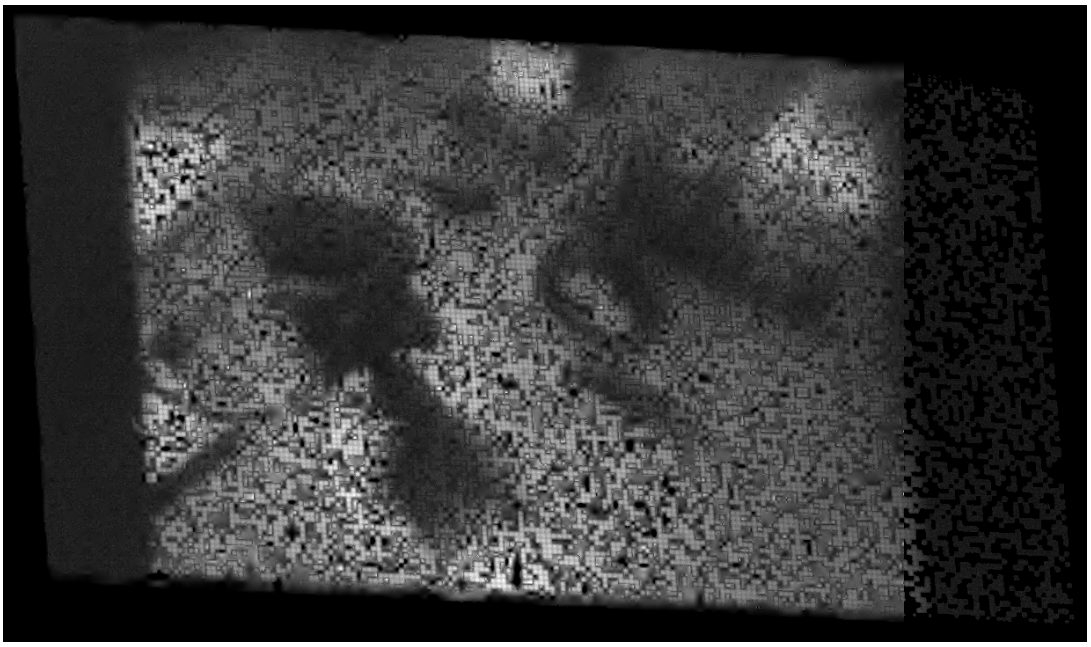


Figure 49: Inpainting using IHT of the laboratory image masked with a CS ratio of 50% and a binning of 10 by 10.

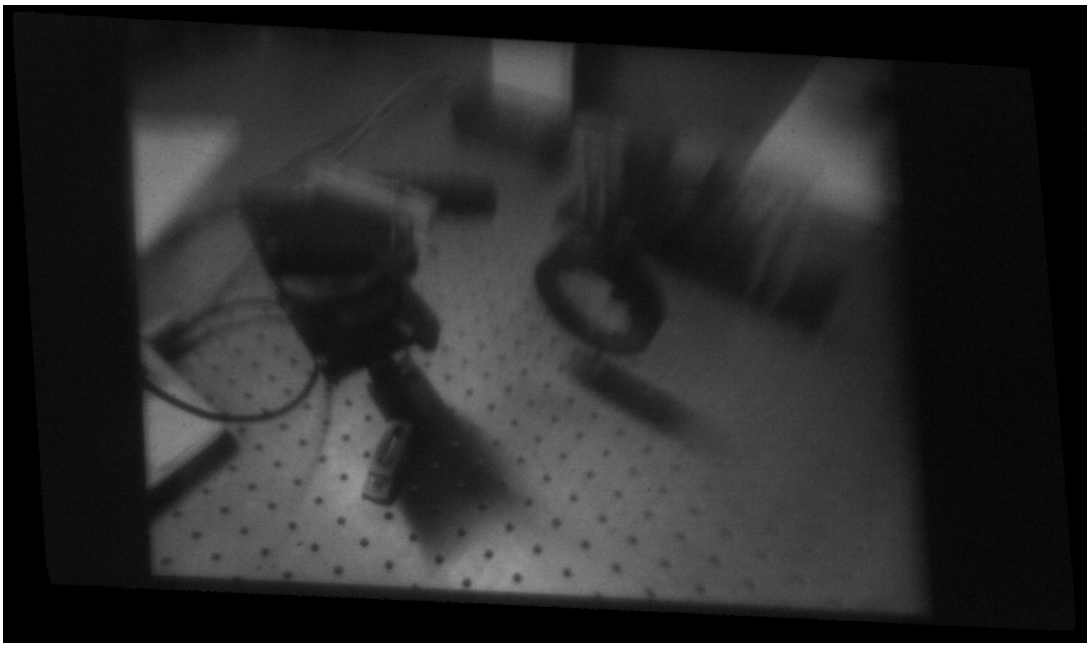


Figure 50: Original image rendering in the optical system.

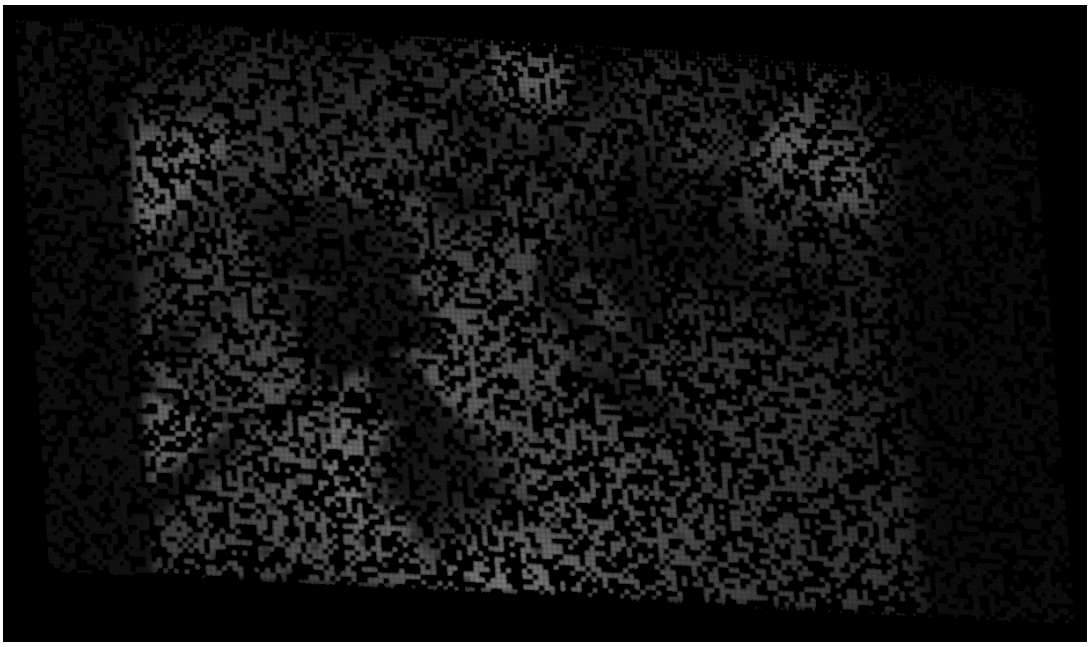


Figure 51: Masked image.

It must be noted that when the mask was passed into the algorithm the edges of the mask were switched in the on state. This way the algorithm would not try to reconstruct an image in this region. It can be seen that the reconstruction is far from perfect. Nonetheless, the image is better than the masked one. This can be quantified with the PSNR and SSIM as described in section 4.3. Comparing the original image and the reconstructed one yields a PSNR of 20.801 dB and a SSIM of 0.615. To have an idea of the improvement with respect to the masked images, the same estimators are used to compare the original image and the masked image. This yields a PSNR of 12.305 dB and a SSIM of 0.369. Hence, the SSIM increases significantly with the reconstruction but the increase is less significant in the PSNR. A possible explanation is that the mask reduces the global signal intensity on the camera, thus even the unmasked pixels can be darker than in the original image. This can add some difficulty for the algorithm to properly reconstruct the original intensity and would explain this relatively low PSNR. Nonetheless, even if the global intensity is reduced, structures and contrasts can still be correctly represented, explaining the stronger increase of SSIM. An additional note is that the algorithm is already optimized to converge towards the best solution, thus changing the number of iterations or the threshold values either leaves the results unchanged or decrease the quality of the reconstruction.

Finally, to test the limit of the inpainting algorithm, one can mask the image with more and wider holes. Practically, this has been done by coding a mask on the DMD with a CS ratio of 25% and a binning of 20 by 20 mirrors. The resulting mask on the camera following the process described earlier is shown in the following figure.

Figure 52: Mask with a CS ratio of 25% and a 20×20 binning on the DMD.

Applying this mask on the laboratory image yields:

Figure 53: Laboratory image masked with a CS ratio of 25% and a binning of 20 by 20 on the DMD.

And finally, the reconstructed image after the inpainting is:

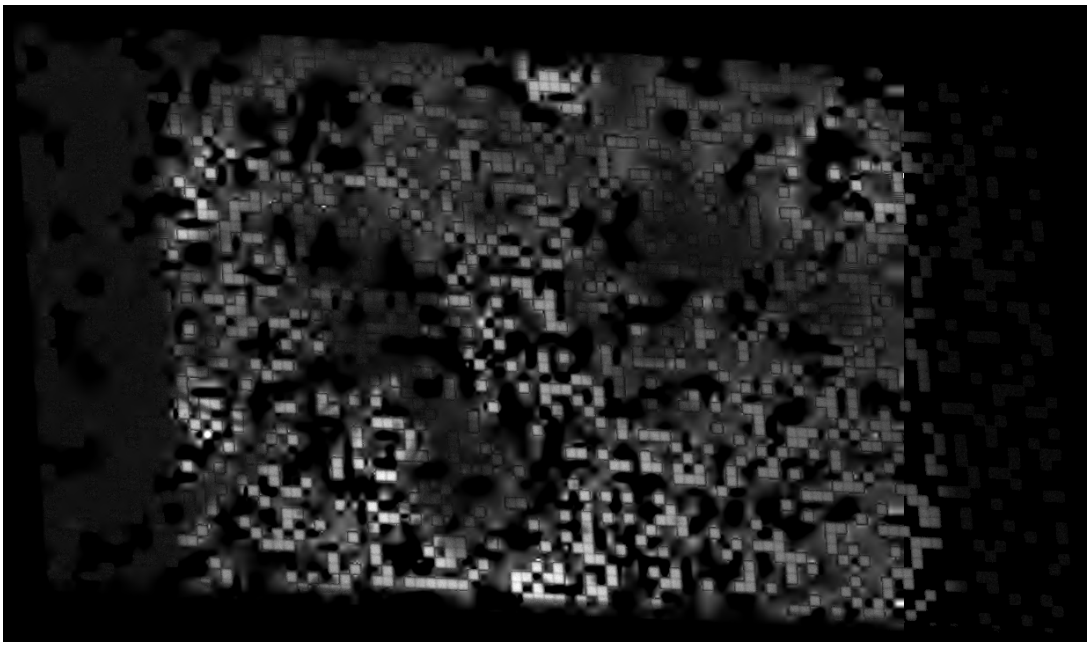


Figure 54: Inpainting using IHT of Fig. 53.

It is obvious that the reconstruction is not good at all because there are too many holes ans thus the reconstruction basis is not complete enough. The PSNR is 21.516 dB and the SSIM is 0.471. In comparison with the reconstruction of the image masked with a CS ratio of 50% the SSIM decreases as expected. However, the PSNR slightly increases. A possible reason is that the larger binning produces less diffraction, thus the intensity of the illuminated pixels is more similar to the corresponding pixels on the complete image. Nevertheless, the increase is too small to be significant.
This shows the limitation of this algorithm to reconstruct images with a low CS ratio. Hence, it can be concluded that while the inpainting using IHT algorithm was sufficient for a first demonstration of the CS technique in the laboratory, a more efficient algorithm must be used to obtain good performances.

6 Conclusion

This master thesis aims to achieve two primary objectives: firstly, to investigate the relevance of CS for an application on Earth observation instruments with an emphasis on the IR spectrum; and secondly, implementing from scratch a first CS experimental chain in a laboratory environment. This work is a continuity of a previous master thesis that studied the efficiency of different algorithms for CS reconstruction. The outcome indicated that DL-based algorithms showed better results than TV-based methods.

The essence of the CS methodology is to reconstruct images from fewer measurements than in classical imaging. Mathematically, this is an ill-posed inverse problem which can only be solved by considering the sparsity of natural signals. This sparsity allows to reconstruct the complete image by solely measuring its significant components in a sparse domain. Practically, this is done by applying random masks on the image via SLMs that hide or attenuate parts of the signal. Subsequently, the masked measurements and the patterns used to obtain them are fed into an appropriate reconstruction algorithm. Such algorithms are generally based on a regularization by minimization of the l_1-norm of the inverse problem. Ultimately, CS offers several advantages, including reduced data storage, transmission, and analysis requirements for images, as well as a decrease in detector sizes due to the reduced number of pixels needed. In addition, CS can be implemented to increase the spatial resolution of imagers, a method called super resolution.
It has been found that CS shows great promise in the area of Earth observation, considering the large quantity of data that is generated by modern multispectral and hyperspectral sensors. Furthermore, the specific case of the IR spectrum is particularly interesting, due to the drastic reduction of detector size that CS could be bring in this spectral range. To do so, some problems must still be solved such as the development of efficient SLM technologies operating at these wavelengths, especially in the TIR range. It has also been found that the application of CS in LEO satellites is limited by the rapid satellite motion. This limitation arises from the need to take several measurements with different masks, which increases the acquisition time. Thus, the motion of satellite can degrade the data quality if the acquisition is too long and the last masks are applied to a different FOV as the first ones.

The ISTA-Net$^+$ DL-based algorithm that was already proven as a fast and effective CS reconstruction algorithm in the previous work has been tested on multispectral Earth observation data. Despite not being specifically trained on Earth observation data, the results obtained are highly satisfactory. Therefore, it can be stated that ISTA-Net$^+$ is suitable to reconstruct compressively sensed Earth observation multispectral images since even better results are expected if the algorithm is retrained with a suitable training set.

An experimental CS setup has been developed within a laboratory environment. It consists in a complete chain starting with the light collection from a scene, the spatial modulation of the signal with a DMD, and the focus of the resulting signal on a camera. The optical system has been successfully implemented, and the reconstruction of images from the measured data has been demonstrated as feasible. Nonetheless, both the optical system and the image reconstruction show several limitations. Firstly, the image on the camera is inclined and partially blurry due to the extremely short DOF of the system. Then, the reconstructed image could not be done with the ISTA-Net$^+$ algorithm and another one dubbed inpainting using IHT has been used. In the end, the reconstructed image, while improving the similarity with the original one compared to the masked image, is far from being a perfect reconstruction. Hence the performance is sufficient for a very first demonstration of the method, but not yet for concrete applications.

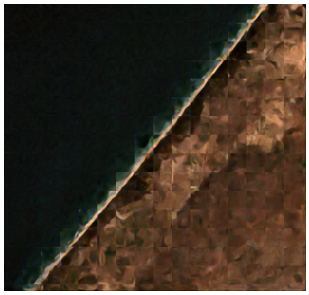
This thesis provides a foundation for future works to dig deeper into the subject and improve the experimental setup. More specifically, the optical system must be optimized to have a sharp and clear image on the detector. Several possibilities have been proposed throughout the work such has the use of tilt-shift lens or a more complex optical components arrangement. Additionally, It would be interesting to adapt the setup for IR imaging. This would require using an IR camera, as well as a DMD with wider micromirrors so that diffraction would not degrade the signal. An adapted IR source would also be needed to illuminate a scene to be imaged.
On the reconstruction side, it is of primary importance to be able to inject the measurements in an efficient, ideally DL-based, algorithm. Either by adapting and retraining ISTA-Net$^+$ to enable direct injection of hardware-compressed measurements into it, or by finding another suitable algorithm.

Appendices

A Reconstruction results

(a) Original

(b) Ratio 1



(c) Ratio 4

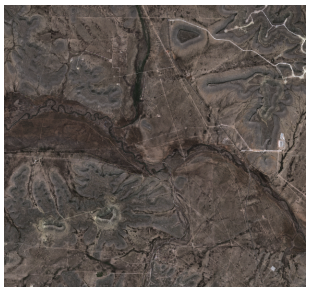
(d) Ratio 10

(e) Ratio 25

(f) Ratio 40

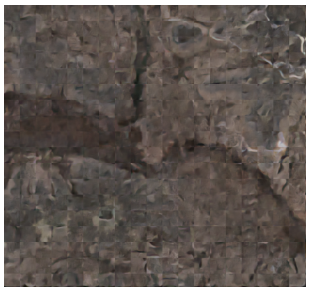
(g) Ratio 50

Figure 55: Results of the reconstructions of the coastal area image.

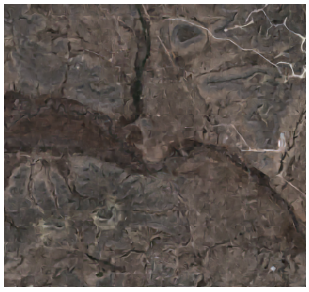


(a) Original

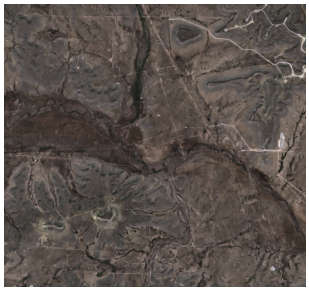
(b) Ratio 1



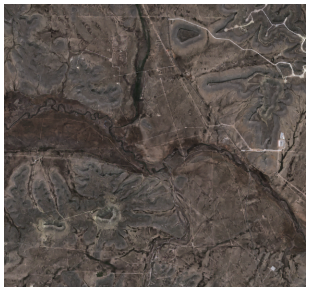
(c) Ratio 4



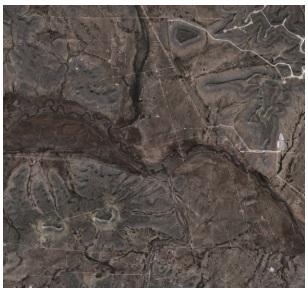
(d) Ratio 10



(e) Ratio 25

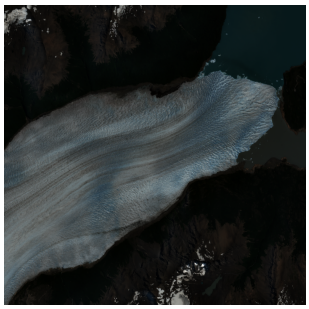


(f) Ratio 40



(g) Ratio 50

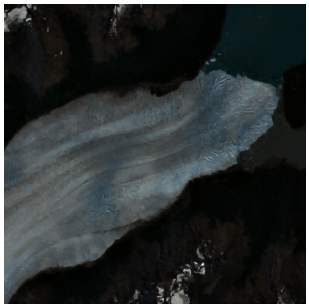
Figure 56: Results of the reconstructions of the desert image.



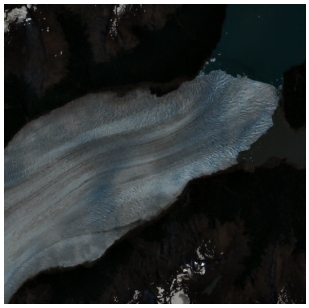
(a) Original

(b) Ratio 1

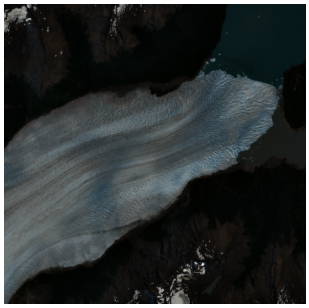
(c) Ratio 4



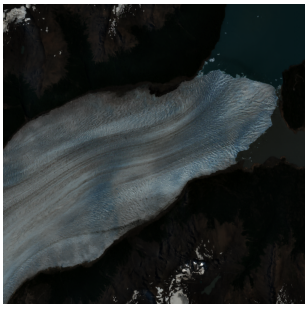
(d) Ratio 10



(e) Ratio 25



(f) Ratio 40



(g) Ratio 50

Figure 57: Results of the reconstructions of the glacier image.

(a) Original

(b) Ratio 1

(c) Ratio 4

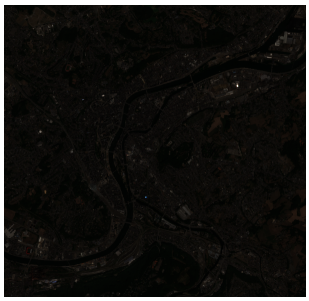
(d) Ratio 10

(e) Ratio 25

(f) Ratio 40

(g) Ratio 50

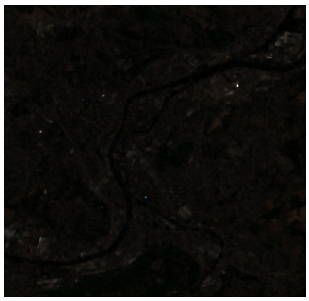
Figure 58: Results of the reconstructions of the lake image.



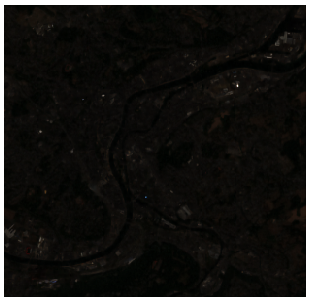
(a) Original

(b) Ratio 1

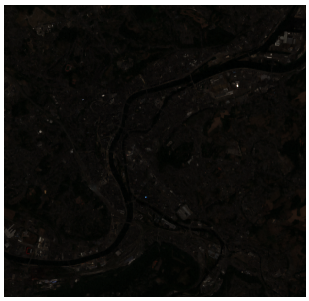
(c) Ratio 4



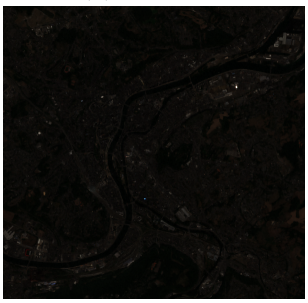
(d) Ratio 10



(e) Ratio 25

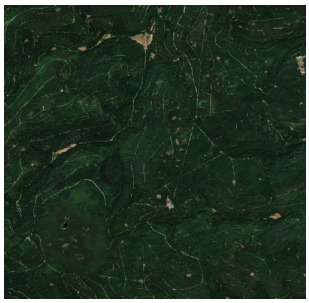


(f) Ratio 40

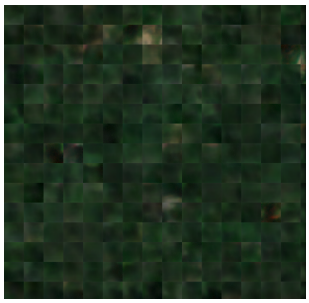


(g) Ratio 50

Figure 59: Results of the reconstructions of the Liège image.

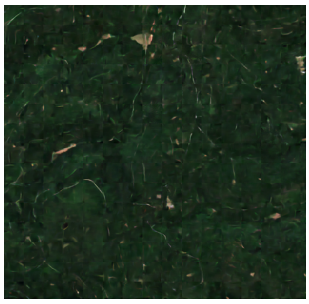


(a) Original

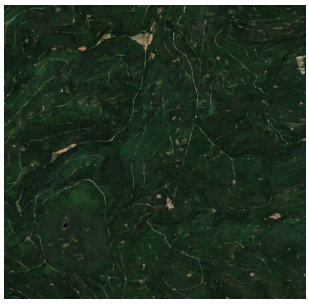


(b) Ratio 1

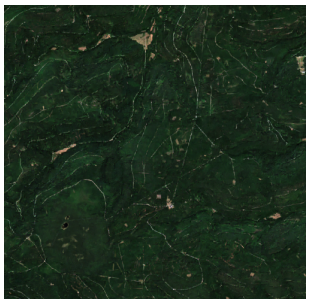
(c) Ratio 4



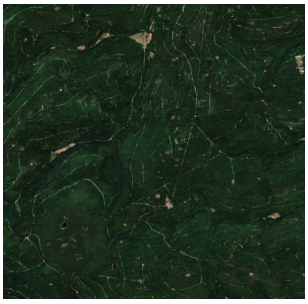
(d) Ratio 10



(e) Ratio 25

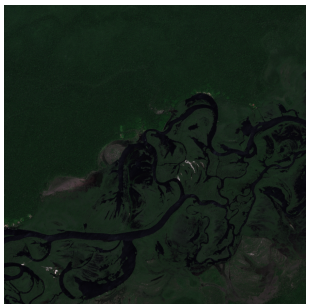


(f) Ratio 40



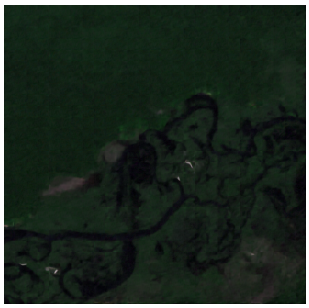
(g) Ratio 50

Figure 60: Results of the reconstructions of the temperate forest image.

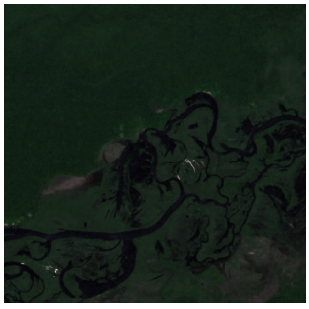


(a) Original

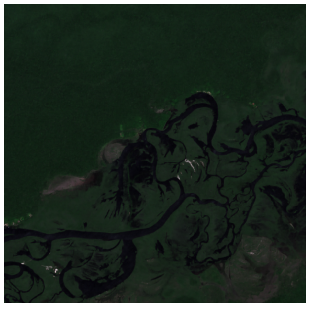
(b) Ratio 1



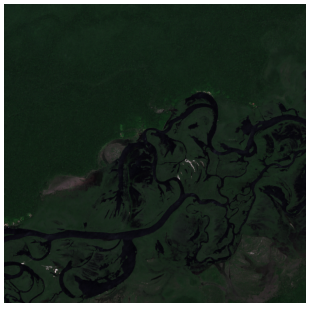
(c) Ratio 4



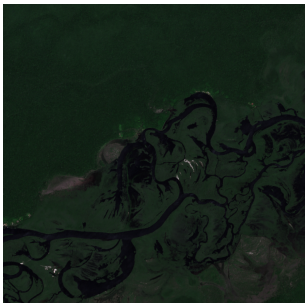
(d) Ratio 10



(e) Ratio 25

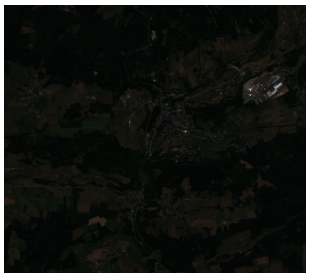


(f) Ratio 40



(g) Ratio 50

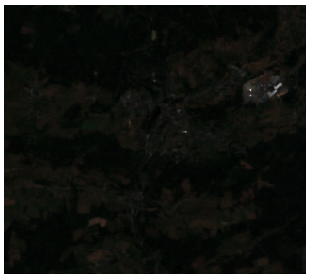
Figure 61: Results of the reconstructions of the tropical forest image.



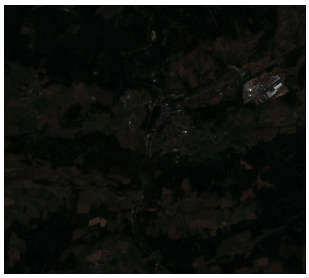
(a) Original

(b) Ratio 1

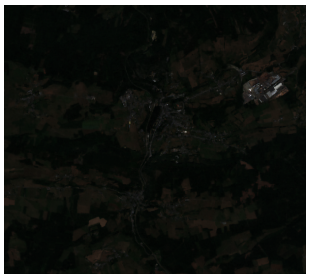
(c) Ratio 4



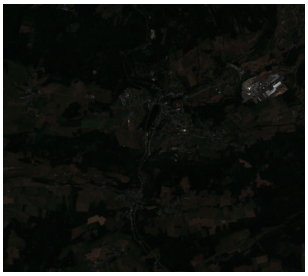
(d) Ratio 10



(e) Ratio 25



(f) Ratio 40



(g) Ratio 50

Figure 62: Results of the reconstructions of the Vielsalm image.

References

1. Foucart, S. & Rauhut, H. *An invitation to Compressive Sensing* (Springer, 2013).
2. Fornasier, M. & Rauhut, H. Compressive Sensing. *Handbook of mathematical methods in imaging* **1,** 187–229 (2015).
3. Baraniuk, R. G. Compressive Sensing [Lecture Notes]. *IEEE Signal Processing Magazine* **24,** 118–121 (2007).
4. Willet, R., Marcia, R. & Nichols, J. Compressed sensing for practical optical imaging systems: a tutorial. *Optical Engineering* **50,** 072601–072601 (2011).
5. Jullien, A. Spatial light modulators. *Photoniques,* 59–64 (2020).
6. Duarte, M. F. *et al.* Single-pixel imaging via compressive sampling. *IEEE signal processing magazine* **25,** 83–91 (2008).
7. Asif, M. S., Ayremlou, A., Sankaranarayanan, A., Veeraraghavan, A. & Baraniuk, R. G. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging* **3,** 384–397 (2016).
8. Slinger, C. W. *et al. Adaptive coded aperture imaging in the infrared: towards a practical implementation* in *Adaptive Coded Aperture Imaging and Non-Imaging Sensors II* **7096** (2008), 41–48.
9. Gehm, M. E., John, R., Brady, D. J., Willett, R. M. & Schulz, T. J. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express* **15,** 14013–14027 (2007).
10. Arce, G. R., Brady, D. J., Carin, L., Arguello, H. & Kittle, D. S. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine* **31,** 105–115 (2013).
11. Gramegna, S. *Study of compressive sensing in view of space imaging applications* Centre Spatial de Liège. MA thesis (Liège University, 2022).
12. Krahmer, F., Kruschel, C. & Sandbichler, M. *Total variation minimization in compressed sensing* in *Compressed Sensing and its Applications: Second International MATHEON Conference 2015* (2017), 333–358.
13. Bentley, P. M. & McDonnell, J. Wavelet transforms: an introduction. *Electronics & communication engineering journal* **6,** 175–186 (1994).
14. Antonini, M., Barlaud, M., Mathieu, P. & Daubechies, I. Image coding using wavelet transform. *IEEE Transactions on image processing* **1,** 205–220 (1992).
15. *Jupyter Notebooks – Inpainting using sparse regularization* Last consulted on May 24, 2023. `https://nbviewer.org/github/gpeyre/numerical-tours/blob/master/python/inverse_5_inpainting_sparsity.ipynb`.

16. Zhang, J. & Ghanem, B. *ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), 1828–1837.

17. You, D., Xie, J. & Zhang, J. *ISTA-Net++: flexible deep unfolding network for compressive sensing* in *2021 IEEE International Conference on Multimedia and Expo (ICME)* (2021), 1–6.

18. Park, S. C., Park, M. K. & Kang, M. G. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine* **20,** 21–36 (2003).

19. Raimondi, V. *et al. A feasability study for a compressive sensing imager in the medium infrared for hotspot detection* in *International Conference on Space Optics—ICSO 2020* **11852** (2021), 2172–2183.

20. ESA. *Newcomers Earth Observation Guide* https://business.esa.int/newcomers-earth-observation-guide#:~:text=Earth%5C%20Observation%5C%20sensors%5C%20onboard%5C%20satellites,the%5C%20orbit%5C%20of%5C%20the%5C%20satellite..

21. USGS. *Radiometric resolution* Last accessed on May 31, 2023. https://www.usgs.gov/faqs/what-radiometric-resolution#:~:text=Radiometric%5C%20resolution%5C%20relates%5C%20to%5C%20how,in%5C%20varying%5C%20degrees%5C%20of%5C%20greyscale..

22. Wilson, M., DeRose, J. & Greb, C. *Microscope Resolution: Concepts, Factors and Calculation* https://www.leica-microsystems.com/science-lab/life-science/microscope-resolution-concepts-factors-and-calculation/.

23. *The Rayleigh Criterion* Last accessed on May 25, 2023. http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/Raylei.html.

24. Jonard, F. Remote sensing [Lecture Notes]. *Université de Liège* (2022).

25. Barrand, N. *Notes for Remote Sensing of the Cryosphere* University of Birmingham. School of Geography, Earth and Environmental Sciences.

26. Mahalanobis, A., Shilling, R., Murphy, R. & Muise, R. Recent results of medium wave infrared compressive sensing. *Applied optics* **53,** 8060–8070 (2014).

27. McMackin, L., Herman, M. A., Chatterjee, B. & Weldon, M. *A high-resolution SWIR camera via compressed sensing* in *Infrared Technology and Applications XXXVIII* **8353** (2012), 48–57.

28. Raimondi, V. *et al. Compressive sensing instrumental concepts for space applications* in *Unconventional Optical Imaging III* **12136** (2022), 66–72.

29. Raimondi, V. *et al.* Designing a Compressive Sensing Demonstrator of an Earth Observation Payload in the Visible and Medium Infrared: Instrumental Concept and Main Features. *Engineering Proceedings* **8,** 27 (2021).

30. Coluccia, G. *et al.* Optical compressive imaging technologies for space big data. *IEEE Transactions on Big Data* **6,** 430–442 (2019).

31. OCS-TECH. *Optical Compressive Sensing Technologies for Space Applications* ESA contract no. 4000116423/15/NL/BJ/gp.

32. Oggioni, L., Sanchez del Rio Kandel, D. & Pariani, G. Earth Observation via Compressive Sensing: The Effect of Satellite Motion. *Remote Sensing* **14,** 333 (2022).

33. Fowler, J. E. *Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging* in *2014 IEEE international conference on image processing (ICIP)* (2014), 684–688.

34. ESA. *SENTINEL-2 Mission Guide* Last accessed: March 29, 2023. `https://sentinel.esa.int/web/sentinel/missions/sentinel-2`.

35. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13,** 600–612 (2004).

36. Ren, Y.-X., Lu, R.-D. & Gong, L. Tailoring light with a digital micromirror device. *Annalen der physik* **527,** 447–470 (2015).

37. *DLP6500FLQ product details and datasheet* Texas Instruments. `https://www.ti.com/product/DLP6500FLQ#params`.

38. Rice, J. P., Neira, J. E., Kehoe, M. & Swanson, R. *DMD diffraction measurements to support design of projectors for test and evaluation of multispectral and hyperspectral imaging sensors* in *Emerging Digital Micromirror Device Based Systems and Applications* **7210** (2009), 114–122.

39. Palmer, C. & Loewen, E. G. *Diffraction grating handbook* (Newport Corporation New York, 2005).

40. Ebner, A. *et al.* Diffraction-limited hyperspectral mid-infrared single-pixel microscopy. *Scientific Reports* **13,** 281 (2023).

41. *Thin-lens equation* Georgia State University. `http://hyperphysics.phy-astr.gsu.edu/hbase/geoopt/lenseq.html`.

42. *UI-324CP Rev. 2 product specifications* IDS. `https://fr.ids-imaging.com/store/ui-3240cp-rev-2.html`.

43. Ray, S. *Applied photographic optics* (Routledge, 2002).

44. Nocerino, E. *et al.* Experiments on calibrating tilt-shift lenses for close-range photogrammetry. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* **41** (2016).