FAKULTA MECHATRONIKY,
INFORMATIKY A MEZIOBOROVÝCH
STUDIÍ TUL

Disertační práce

Single-snapshot compressive hyperspectral imaging using coded aperture

Studijní program:	P3901 Aplikované vědy v inženýrství
Studijní obor:	Aplikované vědy v inženýrství
Autor práce:	**Ing. Jiří Hlubuček**
Školitel práce:	RNDr. Karel Žídek, Ph.D. Ústav nových technologií a aplikované informatiky

Liberec 2022

Prohlášení

Prohlašuji, že svou disertační práci jsem vypracoval samostatně jako původní dílo s použitím uvedené literatury a na základě konzultací s vedoucím mé disertační práce a konzultantem.

Jsem si vědom toho, že na mou disertační práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci nezasahuje do mých autorských práv užitím mé disertační práce pro vnitřní potřebu Technické univerzity v Liberci.

Užiji-li disertační práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti Technickou univerzitu v Liberci; v tomto případě má Technická univerzita v Liberci právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Současně čestně prohlašuji, že text elektronické podoby práce vložený do IS/STAG se shoduje s textem tištěné podoby práce.

Beru na vědomí, že má disertační práce bude zveřejněna Technickou univerzitou v Liberci v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů.

Jsem si vědom následků, které podle zákona o vysokých školách mohou vyplývat z porušení tohoto prohlášení.

12. října 2022 Ing. Jiří Hlubuček

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor for his mentoring, patience, and support. Thank you, Karel, for being the human you are. It was a great pleasure to work under your guidance. You showed me how to do science and inspired me in so many ways.

Further, I would like to thank Jan Václavík and Zbyněk Koldovský for valuable advice and consultations regarding the work in this thesis. I am also thankful to my colleagues from the Group of Computational Spectroscopy and Imaging – Petra, Mariem, Martina, Jiří, Jakub, Lukáš, Vít, and Nirmal, as well as to other colleagues from the TOPTEC centre. Thank you for always being kind to me and for creating a positive work environment.

Rád bych poděkoval své rodině, která mě vždy podporovala a stála při mě. Mým rodičům a prarodičům, kteří mě nasměrovali správnou cestou. Mým sourozencům, kteří pro mě často byli inspirací. I těm, kteří se nedočkali konce této práce.

To the people at Technion in Haifa. Especially Yonina, Oren, Regev, and Satish. It was an interesting and enriching experience to work in your group. I have learned a lot from you, and the knowledge I obtained during the internship highly enhanced my doctoral studies. רבה תודה

To the Photonics Research Group in Castellón de la Plana. Although it was a rather short stay partly in lockdown, I made the most of it. Particularly I would like to thank Enrique not only for mentoring me but also for aiding me with finding accommodation. It was very considerate of you, and I appreciate it. ¡Muchas gracias!

Lastly, I cannot not mention my friends. If it were not for you, I would finish my thesis much earlier. But it would be a lonely and sad journey.

Děkuji.

Annotation

This thesis presents a novel approach to the remote detection of chemical information by compressing the hyperspectral information directly during the measurement. This was enabled thanks to a novel technique developed on a single snapshot called coded aperture snapshot spectral imaging (CASSI). Using a coded aperture allows the implementation of a modern signal processing technique based on an algorithmic strategy – compressed sensing. This method can capture complete hyperspectral information in a single instance without scanning, which yields a significantly higher optical throughput compared to its scanning-based counterparts. Using the CASSI system, it is possible to retrieve the information faster than by a conventional apparatus, utilizing a relatively simple optical setup. In this thesis, the method was extended in order to perform hyperspectral imaging on a broad spectral range in the infrared region.

Keywords: hyperspectral imaging, compressed sensing, coded aperture

Anotace

Tato práce představuje nový způsob získání chemické informace na dálku s kompresí hyperspektrální datakrychle přímo v průběhu měření. Toho je docíleno díky nové technice založené na jediném snímku z detektoru, tak zvané CASSI (Coded Aperture Snapshot Spectral Imaging). Kódovaná apertura umožňuje implementovat moderní techniku zpracování signálu s použitím algoritmu – komprimované snímání. Tato metoda je schopna zachytit kompletní hyperspektrální informaci v jediný okamžik bez nutnosti skenování a dosahuje daleko větší optické propustnosti než její standardní, na skenování založené protějšky. Použitím CASSI je možné získat měřenou informaci rychleji než pomocí tradičních systémů, s použitím relativně jednoduchého optického uspořádání. V této práci byla CASSI metoda vylepšena pro možnost hyperspektrálního zobrazování na širokém spektrálním rozsahu v oblasti infračerveného spektra.

Klíčová slova: hyperspektrální zobrazování, komprimované snímání, kódovaná apertura

Contents

List of Abbreviations

CASSI	**C**oded **A**perture **S**napshot **S**pectral **I**maging
CR	**C**ompression **R**atio
CS	**C**ompressed **S**ensing
D-CASSI	**D**ifferential **CASSI**
DCT	**D**iscrete **C**osine **T**ransform
DMD	**D**igital **M**icromirror **D**evice
DWT	**D**iscrete **W**avelet **T**ransform
FO	**F**irst **O**rder
GPSR	**G**radient **P**rojection for **S**parse **R**econstruction
HDES	**H**yperspectral **DE**tection **S**ystem
HS	**H**yper**S**pectral
HSI	**H**yper**S**pectral **I**maging
HVS	**H**uman **V**isual **S**ystem
IR	**I**nfra**R**ed
LWIR	**L**ong-**W**ave **I**nfra**R**ed
MIR	**M**id-**I**nfra**R**ed
NIR	**N**ear-**I**nfra**R**ed
PSNR	**P**eak **S**ignal-to-**N**oise **R**atio
SAM	**S**pectral **A**ngle **M**apper
SNR	**S**ignal-to-**N**oise **R**atio
SSIM	**S**tructural **SIM**ilarity
TwIST	**Tw**o-step **I**terative **S**hrinkage/**T**hresholding algorithm
VIS	**VIS**ible
ZO	**Z**eroth **O**rder

Introduction

Our vision is an extraordinarily complex and astonishing system, and we depend on it as a primary source of gathering information about our surroundings. Nevertheless, it has limitations in imaging very tiny or distant objects, and we are constricted to only three spectral channels in the visible spectra. Observing nature and developing optical systems over the centuries taught us that there is always something new to discover and far more information around us than we are able to grasp with our senses, whether it is gravitational waves or trillions of neutrinos permeating you at this very second. In fact, the amount of information around us is so vast that for the human brain to process it, it is designed to filter even the limited visual data captured by our eyes. Seeing a broad spectral range, e.g., in the infrared (IR) region with a fine spectral resolution, would allow us to sense chemical composition remotely [1]. This is what hyperspectral imaging (HSI) does. HSI denotes methods that are able to capture both an image and a spectrum of light in each pixel of the scene. Such a dataset provides us with immense information about the measured scene, which cannot be obtained by any other means. Therefore, IR HSI has been a very lively topic in the last decades [1, 2, 3, 4, 5].

Within this field, most studies are focused on the near IR spectral region because it is possible to use common optical elements and germanium detectors or detectors on an InGaAs basis. However, the implementation of HSI in the middle and far IR spectral region is problematic due to the need to use "exotic" optical materials and detectors with high noise levels. Moreover, HSI is very demanding regarding the acquisition time, computing power, processing, and storing the information.

A possible solution to these problems is utilizing compressed sensing (CS) [6]. CS represents data acquisition and processing methods that exploit the essential feature of signals present in nature (i.e., not artificially created) that is not obvious at first sight. Natural signals are sparse or sparse-like – it means that they are compressible, and it is possible to represent them by far fewer measurements than is required by the Nyquist-Shannon sampling theorem [7]. One of the methods that combine HSI with CS techniques is the so-called CASSI (Coded Aperture Snapshot Spectral Imaging) [8, 9, 10]. The CASSI technique is the central method of this thesis.

The two main components of the CASSI setup are a dispersive element and a coded aperture, which is a random pattern – typically a binary one. The random pattern, or in other words, a random mask, encodes the image information for every wavelength and a dispersive element shears the image spectrally on the detector. The resulting image is then used for reconstructing the hyperspectral information back. The unique advantage of CASSI is it does not need to scan the scene. All the hyperspectral data are captured in a single instance. However, since the CASSI method relies on one snapshot on a detector, the compression of the hyperspectral data is enormous. This limits the use of this method on a small number of spectral channels, which often implies the acquisition of a narrow bandwidth [9, 11]. Hence, an improvement of this method is on the spot.

This thesis describes our aim to improve the potential of the CASSI method by using various approaches to improve the compression ratio and reconstruction quality. The approaches are evaluated both in an experimental way and in realistic simulations, which test their utilization for the IR HSI.

The first chapter is devoted to the theoretical fundamentals necessary to understand the experimental part of this thesis. Most notably, there is a short introduction to hyperspectral imaging, compressed sensing, and the CASSI method. The second chapter provides an overview of the state of the art and applications. Also, there are mentioned objectives of this work. From the third chapter onward, starts the experimental part of my work. This includes characterization of our optical system, data processing, and reconstruction. Chapters 4, 5, 6, and 7 describe the modifications of CASSI and evaluate the utilization of the method in the IR spectral region. Chapters 4-6 also give more insight into the articles I have written and are attached at the end of this thesis. The results in Chapter 7, exploring the potential of an enhanced CASSI system for IR HSI, are yet to be published. Chapter 8 summarizes the work carried out within the thesis.

1 Theory: the pieces of the puzzle

In this chapter, I am going to describe the essential aspects needed to understand the CASSI technique, which I have been using throughout the thesis. I will begin with a brief description of hyperspectral (HS) imaging with a focus on the infrared (IR) spectral region. I will continue with an introduction to compressed sensing and the associated method CASSI, showing the fundamentals, characteristics, and limitations. Finally, I will present metrics to assess the HS datacube quality.

1.1 Hyperspectral imaging in the IR spectral region

The primary motivation behind going down the rabbit hole of IR hyperspectral imaging is that every chemical compound has its specific absorption spectrum in the IR spectral region. Hence it is possible to distinguish subjects of different chemical compositions. Absorption of the IR radiation corresponds to the basic vibrational and rotational states of chemical bonds. When IR light interacts with a molecule, its chemical bonds start to vibrate more energetically. Thus it influences absorption at certain frequencies in the spectrum – i.e., wavelengths that are characteristic for each chemical bond. [12, 13]

HSI, also known as chemical or spectroscopic imaging, is a technique that combines conventional imaging with spectroscopic systems in order to obtain both spatial and spectral information about the measured scene. IR spectroscopy, which is the foundation of IR-HSI, is based on the interaction of IR light with molecules of the investigated sample. The results of this interaction could be characterized by absorption, reflection, and emission. Generally, the compound-specific absorption in the near IR spectral region is weak [13]. Therefore, focusing rather on middle and far IR light is needed, i.e., 3-8 μm and 8-15 μm wavelengths, respectively. The fundamentals and evolution of IR spectroscopy can be found in many publications [1, 2, 14].

A spectrum in standard non-imaging spectroscopy represents the integral spectral information about the sample, which depends on the size of the illuminated area. A result of HSI is a three-dimensional matrix of data, the so-called datacube, where one spectrum is assigned to every pixel. The datacube can also be imagined as a stack of papers where each sheet constitutes an image of the inspected sample at a different wavelength.

For a better understanding of how the datacube is acquired, three basic configurations of apparatus for the acquisition of hyperspectral information are listed below:

1. Point mapping (also whisker-broom) – a measured sample is scanned point by point, and for each point, one spectrum is acquired (Figure 1A)
2. Line scan (also push-broom) – a detector captures spectral information for a whole column of pixels simultaneously; a light goes through a slit, and then it is perpendicularly dispersed; hence it is possible to record spectral information along this whole line depending on the spot from which the light emerges; in this way a two-dimensional array

is obtained, which has one spectral dimension and one spatial dimension; in order to get the second spatial dimension of the datacube one must perform scanning of the sample in a direction perpendicular to the imaging line (Figure 1B)

3. Plane scan (also staring configuration) – this arrangement does not need any movement nor spatial scanning of a sample; therefore, it is often denoted as a staring configuration; incoming light from the whole surface of the sample is recorded on a detector as a two-dimensional spatial array for one wavelength at a time; this could be achieved by bandpass filters that are attached to a rotating disc called filter wheel or by changing the wavelength of light by tunable [15, 16] filters (Figure 1C) [12, 17]

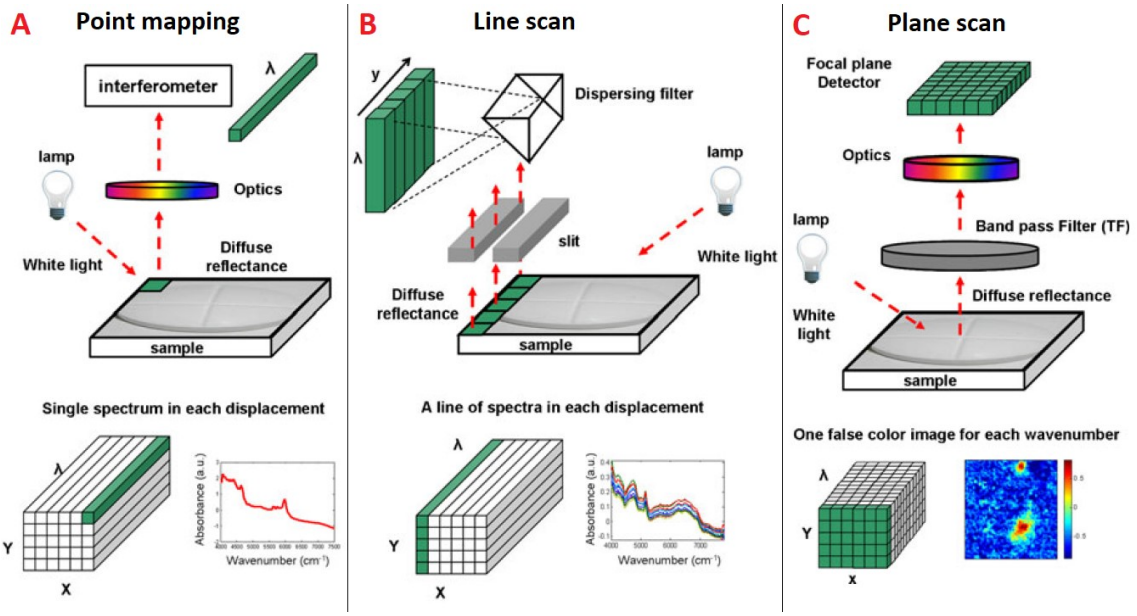

Figure 1 Scheme of different approaches to obtain hyperspectral information and a structure of resulting datacubes. (A) Point mapping. (B) Line scan. (C) Plane scan. Adapted from [17]

A typical hyperspectral camera consists of splitters (tunable filters, diffraction gratings, Fourier transform spectrometers), imaging lenses, a detector, and a device for scanning/controlling motion to obtain very precise spatial information. [12, 17]

The two basic configurations mentioned above that use translation motion of the sample image are point mapping and line scan. These methods have both excellent spatial and spectral resolution. However, from its principle, the acquisition time is of considerable length – up to the order of tens of minutes to hours, depending on the signal level. The time is based on the size of the scanned area, the wavelength range, and the number of scans per pixel [18]. This problem is a huge limitation, especially for the whiskbroom configuration. To provide a particular example, a 256×256 pixels HS image acquired by the point mapping needs an acquisition time of more than 54 minutes, even if a single spectrum is taken within 50 ms.

On the contrary, by plane scan, it is possible to acquire the whole datacube in a matter of seconds or minutes, which hinges only on the number of scanned wavelengths. This can be utilized

especially for the investigation of processes that are unstable over time. The tradeoff here is that, from its essence, this method has worsened spectral resolution. [12]

1.1.1 Black-body radiation and Plancks law

One of the important aspects of the HSI in the IR spectral region is the fact that each object with a non-zero temperature emits some heat as radiant energy. These emitted photons – black-body radiation – represent a background of the measured scene, and, unlike for the VIS optics, these are also emitted by the measurement apparatus itself.

A "black body" is an idealized material with 100% emissivity, i.e., it transforms heat into radiant energy at the maximum rate. Real-life materials emit radiation energy based on their physiochemical properties – they absorb some radiation by molecular interactions. [19] The emissivity at a given wavelength λ is defined by the ratio of the radiance emitted by a material $L(\lambda, T)$ and a blackbody $B(\lambda, T)$ at the same temperature T:

$$\epsilon(\lambda) = \frac{L(\lambda; T)}{\mathrm{B}(\lambda; T)} \qquad \textit{Equation 1}$$

The spectral radiance of a black body is shown in Figure 2 and is determined by Planck's law:

$$B_\lambda(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1} \qquad \textit{Equation 2}$$

where h is the Planck constant, c is the speed of light, and k_B is the Boltzmann constant.

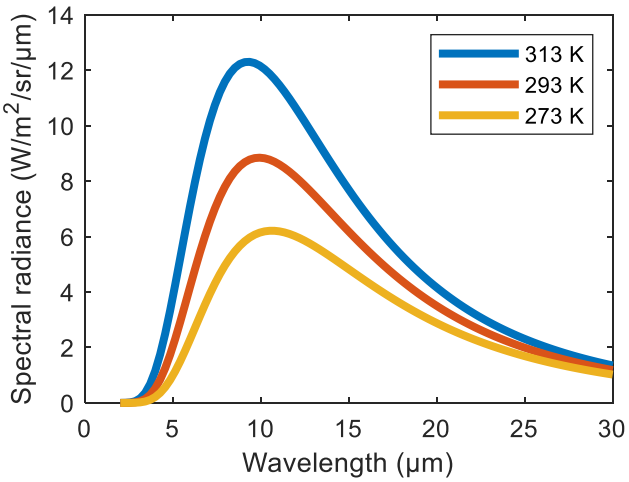

Figure 2 Black-body radiation described by Planck's law for three different temperatures.

1.2 The sky is not always the only limit

It is worth noting that acquiring, storing, and processing the HS datacube is very demanding regarding time, computational power, and data storage capacity. Moreover, in conventional signal processing, one is limited by the Nyquist-Shannon sampling theorem, which states that in order to reconstruct a signal faithfully, the sampling frequency must be at least twice as large as the highest frequency in the signal [7].

Another disadvantage is that acquiring datacubes by scanning results in a loss of light intensity, which depends on the scanning configuration. For a datacube M×N×L, when performing point mapping, we lose light throughput equal to a factor of M×N [20]. Line scan diminishes light throughput by a factor of M or N depending on the scanning direction [21] and by a factor of L in the case of plane scan [22]. Note that the signal intensity is inversely proportional to the spectral and spatial resolution. When we evaluate the photon budget of the HSI, the total amount of light collected from the measured object is given by the imaging optics, and this light intensity is divided into millions of voxels of a datacube. For instance, when we decide to improve twice the spatial resolution of a line-scan HS camera, we need to use a half-width slit that, in turn, reduces the amount of incoming light by 50%. Likewise, for more detailed spectra during a plane scan, we need to restrict the spectral filter to a narrower spectral bandpass, reducing the total intensity of the transmitted light.

1.3 Compressed sensing: going beyond the sampling theorem

The Nyquist-Shannon theorem can be overcome by the compressed sensing (CS) method. It is possible to reconstruct a sparse signal sampled at a rate less than the one restricted by the Nyquist-Shannon criterion through constrained l_1 minimization [23]. This approach was used in the 1970s in reflection seismology for constructing images of layers within the earth [24]. CS is invaluable for cases where there is a need to acquire huge datasets, such as in hyperspectral imaging. An impressive demonstration of the acquisition of an extensive amount of data and the CS uniqueness can be found in [25], where the authors were able to attain as many as 70 trillion frames per second. CS relies on two premises which are *sparsity* and *incoherence*.

1.3.1 Sparsity

The vast majority of natural signals contain only a tiny percentage of important information, i.e., the signal is compressible or sparse – it is possible to write it down very concisely in a correctly chosen basis Ψ with a very small number of non-zero elements. A sparse-like signal denotes a similar case, where one can discard the nonimportant coefficients of the signal, i.e., set them to zero, without much perceptual loss. [7]

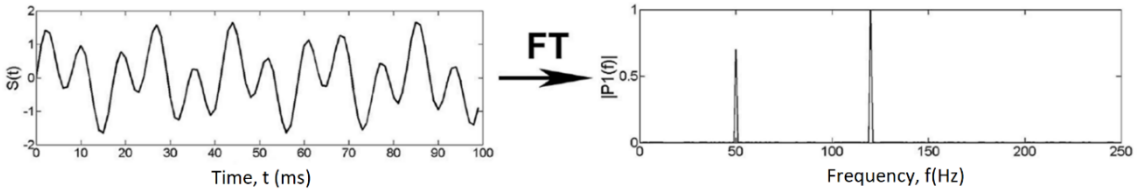

Figure 3 Fourier transform of a signal from time domain to frequency domain. Adapted from [26]

In more mathematical terms, we have a vector f from R^n which could be an *n*-pixel image. We expand f in an orthonormal basis (e.g., a wavelet basis) $\psi = [\psi_1 \psi_2 \dots \psi_n]$ as follows:

$$f(t) = \sum_{i=1}^{n} x_i \psi_i(t) \qquad \textit{Equation 3}$$

where x is the projection of f in the basis, $x_i = \langle f, \psi_i \rangle$. Let's express f as ψx, where ψ is $n \times n$ matrix with $\psi_1, \psi_2, \ldots, \psi_n$ as columns. Now the sparsity implication about discarding nonimportant coefficients should be more obvious if we denote $f_S(t)$ as a vector from the expression above, keeping only the terms corresponding to the S largest values of (x_i). Then from the definition – $f_S := \psi x_S$, where x_S is the vector of coefficients (x_i) which all are set to zero except the S largest ones. Given that ψ is an orthonormal basis, we get $\|f - f_S\|_{l_2} = \|x - x_S\|_{l_2}$ and if the sorted magnitude of the coefficients (x_i) decay quickly, then x_S is a close approximation of x and, consequently, the error of $\|f - f_S\|_{l_2}$ is negligible. [6, 7]

In plain words, the signal can be well described only by the few largest coefficients. In Figure 3, it can be seen that a relatively complex signal shown in the time domain (on the left) has a very sparse representation in the frequency domain (on the right). According to the Nyquist-Shannon theorem, its full description would require sampling the signal with a sampling interval below 4 ms, which would be 250 points each second. At the same time, the same signal can be fully described by two amplitudes and phases of sine oscillations.

Analogically this also applies to an image signal. For instance, a regular camera captures a scene pixel by pixel. Afterward, based on the knowledge of "important wavelet components," it compresses it to JPEG format, which is many times smaller than the RAW data. This is achieved without apparent quality loss. Typically 1 Mpx photo (Figure 4A) can be compressed to 2.5% of its original size and still appears sharp (Figure 4C) using only the 25000 strongest wavelet transform components (Figure 4B). However, the problem is that one is able to compress the scene only after it is captured because, in the general case, we do not know apriori which components will carry the vital information about the image.

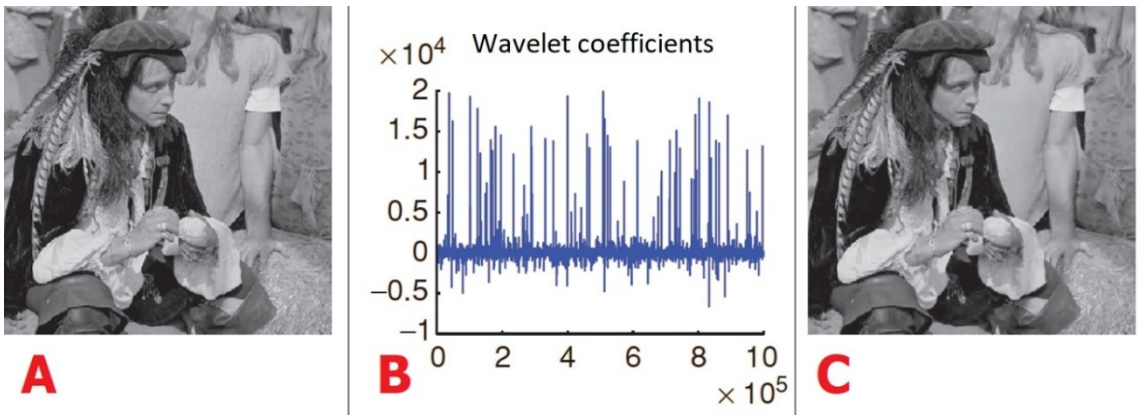

Figure 4 Demonstration of image sparsity using a wavelet transform. (A) An original image, (B) its wavelet coefficients, and (C) a compressed image using only 2.5% of the strongest coefficients. Adapted from [7]

CS brings proof that for the cases where we have no information about the important components of the measured signal (e.g., image), the most efficient way to acquire the information is to sample the signal randomly or incoherently.

1.3.2 Incoherence

The second premise for CS is incoherence. Suppose that we have a pair (Φ, Ψ) of orthonormal bases of R^n. The first basis Φ is used for sensing the signal $f(t)$, and the second basis Ψ is for the representation of $f(t)$. The restriction to orthonormal bases is for the sole purpose of simplicity, but it is not necessary. Then, the coherence between the sensing basis Φ and representation basis Ψ is formulated by:

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \le k,j \le n} \left| \langle \varphi_k, \psi_j \rangle \right| \qquad \textit{Equation 4}$$

It can be said that the coherence expresses the largest correlation between any two elements of the bases Φ and Ψ [6, 7]. The degree of coherence is set by linear algebra that $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$, where number one signifies complete incoherence and $\sqrt{n}$ is maximal coherence. So, if there are any correlated elements contained in Φ and Ψ, then the coherence is large; else, it is small.

Compressed sensing is mainly focused on bases with low coherence. An example of bases that are incoherent, so $\mu(\Phi, \Psi) = 1$, are spikes and sinusoids, i.e., time- and frequency-sampling. Incoherence also expresses the idea that a signal which has a sparse representation in Ψ must be spread out in the domain in which it was recorded. In other words, a peak in the frequency domain is spread out in the time domain and vice versa (see Figure 3). [7]

1.3.3 Compression ratio

One of the essential features of CS is the degree of compression we can achieve. It is described by the compression ratio. However, the literature is ambiguous about the notation. We define the compression ratio as the number of measured points divided by the number of reconstructed points:

$$CR = \frac{Compressed\ size}{Uncompressed\ size} \qquad \textit{Equation 5}$$

Throughout the thesis, when speaking about compression ratio, we will use Equation 5 expressed in percents (CR = 1 = 100 %).

1.4 CASSI: the fastest recording of hyperspectral information there is

The CASSI (Coded Aperture Snapshot Spectral Imaging) method is a CS technique that utilizes a coded aperture and a dispersive element in order to modulate the optical field of the scene, which is then acquired on the detector as a two-dimensional image – the so-called snapshot. A random or pseudo-random mask, typically a binary pattern that resembles a QR code, can be used as a coded aperture. An example of such a random mask can be seen in Figure 8B. The random mask blocks or transmits the light incident on this coded aperture in relation to its binary value.

For randomly generated masks – the so-called Bernoulli matrices - the ratio between blocked and transmitted light is approaching 50%, and the columns of the mask are with a very high probability independent (see Chapter 1.3.2). As it was mentioned in Chapter 1.2, the signal intensity for the standard HSI is inversely proportional to the spectral/spatial resolution. However, the CASSI

method benefits from its signal intensity throughput. With higher resolution, the transmitted light stays constant – a higher resolution mask still blocks the same amount of light as a lower resolution one.

The principle of this method could be described as follows. The light of different wavelengths collected by imaging optics from a measured scene is encoded by a random mask, and then it is spectrally sheared by a dispersive element, e.g., an optical grating. The sheared image is finally captured on a detector. See Figure 5 for a graphic representation of the optical flow in CASSI. As a result, we attain an image of the scene, which consists of randomly-encoded images for different wavelengths laid one over the other with a slight spatial shift dependent on the wavelength. The fact that columns of the random mask are independent makes it possible to distinguish a unique pattern in the spectrally sheared image. Suppose we would use two identical columns in the mask. We could interchange the location in the scene with the spectral position because we would not be able to tell whether the column is at the specific location because of the scene location or because of the spectral shearing.

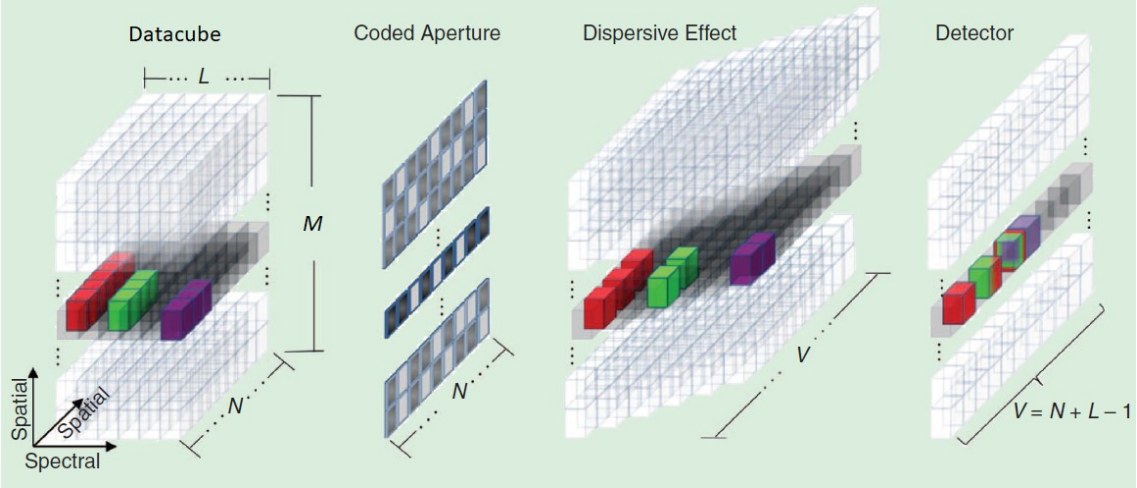

Figure 5 An illustration of the spectral optical flow in CASSI. Adapted from [10]

The most notable advantage of this approach is capturing the complete dataset in one instance – a snapshot, which could be crucial for studying rapidly-changing scenes. Moreover, there are no moving components, otherwise needed for scanning in conventional approaches, and thus a relatively simple design could be made [27].

The reader should be aware that CASSI captures 3D information on a 2D detector – a hyperspectral datacube with dimensions of M×N×L results in an image on a sensor with dimensions of M×(N+L-1). Therefore, the original information is compressed on the detector with compression ratio CR=(N+L-1)/(N×L). For a 256×256 image with 100 spectral slices, this leads to the compression ratio of 90880/6553600 ≈ 1.4%. Compared with the compression example in Figure 4, the compression ratio here is much lower, and the 3D data is way more complex than the image, which results in problems with reconstruction.

However, the CS concept makes it possible to retrieve the complete 3D information back by using a known random mask and assuming that the examined signal is sparse. So we are able to perform a reconstruction of the original scene thanks to a minimization algorithm calculating the following equation:

$$f(D) = \frac{1}{2}\left\|I - \widehat{W}D\right\|^2 + \tau\Phi(D) \qquad \textit{Equation 6}$$

where I is the output of the detector; hyperspectral datacube is denoted as $D(x, y, \lambda)$, where each spectral slice is transformed by an operator $\widehat{W}$. The operator $\widehat{W}$ describes the passage of the light through the system and includes the modulation by random mask $\widehat{M}$, the spectral shearing of the image $\hat{S}$, and the transformation of the datacube D from its sparse basis into the sensing basis $\widehat{H}$. In this thesis, the operator $\widehat{H}$ is used for Haar and Symlet 8 wavelet transforms, but it can generally be any linear transform. By taking this into account, it is possible to formulate an operator $\widehat{W}$ as:

$$\widehat{W} = \hat{S}\widehat{M}\widehat{H}^{-1} \qquad \textit{Equation 7}$$

The second term in Equation 6 is denoted as a regularization term and can have many different forms. Owing to the sparsity of common scenes in a wavelet basis, we will use throughout the thesis the l_1-norm, i.e., $\Phi = \|D\|_1$. The regularization parameter τ then emphasizes the sparsity of the resulting reconstructed signal.

It is worth noting that the data compression is enormous for the CASSI method, which in turn negatively affects the reconstruction of complex hyperspectral datacubes. Thus, an extension of this method is on the spot.

1.5 Assessing the reconstruction quality (similarity of datacubes)

According to the International Standards Organization (ISO), "quality" can be defined as "fitness for use," "customer satisfaction," "doing things right the first time," or "zero defects." [28] This could sound relatively simple, but taking into account the human visual systems' (HVS) peculiarities, it makes judging the quality of an image a challenging task. A thorough reflection on the topic with scientific data evaluation can be found in [29].

To assess the quality of reconstructed datacubes, at first, we were using the HVS. Nevertheless, with more data to process and more complicated scenes, the neural network behind the HVS was no longer satisfactory. Therefore, we created artificial data and calculated the lowest attainable difference between the ground truth and the reconstructed datacube by least squares. We denote this difference as Δ and define it as a minimum of least squares with optimization of scaling factor a:

$$\Delta = min_c\left\{\frac{1}{n}\sum_{i=1}^{n}(a.y_i - \bar{y}_i)^2\right\}, \qquad \textit{Equation 8}$$

where n is the number of datacube voxels, y_i and $\bar{y}_i$ are original and reconstructed datacube values at the ith point, respectively.

However, with more knowledge about the field of CS and HSI, we started to evaluate the quality by more commonly used metrics: PSNR (peak signal-to-noise ratio), SSIM (structural similarity) index, and SAM (spectral angle mapper).

PSNR. It is worth noting that in order to calculate PSNR, we scaled the whole datacube by a single factor, not slice-by-slice, which is sometimes used in the literature and yields higher PSNR. We calculated PSNR as:

$$\mathrm{PSNR} = 10\log_{10}(\mathrm{peakval}^2/\mathrm{MSE}), \qquad \textit{Equation 9}$$

where *peakval* is the maximum value of the original datacube, and as MSE, we use Δ.

SSIM. The SSIM index is a multiplicative combination of three terms – the luminance term l, the contrast term c, and the structural term s.

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \qquad \textit{Equation 10}$$

The three terms could be written as:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \qquad \textit{Equation 11}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \qquad \textit{Equation 12}$$

$$s(x,y) = \frac{\sigma_{xy} + \mathrm{C}_3}{\sigma_x\sigma_y + \mathrm{C}_3} \qquad \textit{Equation 13}$$

where μ_x, μ_y are the local means; σ_x, σ_y are standard deviations; and σ_{xy} is cross-covariance for images $x,\ y$. If we select $\alpha = \beta = \gamma = 1$, and $C_3 = C_2/2$, the index simplifies to:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + \mathrm{C}_1)(2\sigma_{xy} + \mathrm{C}_2)}{(\mu_x^2 + \mu_y^2 + \mathrm{C}_1)(\sigma_x^2 + \sigma_y^2 + \mathrm{C}_2)} \qquad \textit{Equation 14}$$

SAM. SAM is the spectral similarity between the reconstructed and original spectra. The spectral error is calculated as:

$$sam = cos^{-1}\left(\frac{\sum_{i=1}^{c} t_i r_i}{\sqrt{\sum_{i=1}^{c} t_i^2}\sqrt{\sum_{i=1}^{c} r_i^2}}\right) \qquad \textit{Equation 15}$$

where t is the reconstructed spectra in the hyperspectral data and r is the original spectra of length C. The final SAM score is the mean value of sam.

So to summarize the metrics, in an ideal case – meaning the reconstructed datacube would be identical to the original one, we would get Δ and SAM equal to zero, while SSIM would be one. In the case of PSNR, the higher, the better.

2 State of the art

2.1 Current challenges in hyperspectral imaging

In contrast to the standard RGB imaging, by hyperspectral imaging (HSI), we obtain spectral information in addition to the 2D image. The spectral information can be exploited in various ways – for instance, to identify the chemical composition. From the principle of this method, it is evident that one acquires a massive amount of information (see Chapter 1.1). This naturally affects the acquisition times, which could be in the order of seconds to hours to acquire the complete information, depending on the required spatial and spectral resolution and the method used [17].

An integral part of most hyperspectral instruments is the moving part required to scan the imaging scene, which is one aspect of the complexity of these instruments. It is also necessary to mention the low efficiency of radiation utilization. For example, in a line scan (Figure 1B), a large part of the light intensity is filtered out by the slit used, or in a plane scan (Figure 1C), the light is filtered out using a narrow bandpass filter. So, the better the spectral resolution one wants, the worse the use of light is.

Recording, storing, and processing hyperspectral information is very demanding in terms of acquisition time, storage capacity, and computing power. For example, in the realm of earth remote sensing, there are freely accessible data from the AVIRIS instrument [30]. It is an optical sensor that provides calibrated images of the spectral radiance in 224 spectral channels. Typically, the file size of these data is several GB. Therefore, the traditional classification techniques cannot be directly applied to the HS data, and modification is needed [31]. In addition, the need for using complex optics, and thereby the high purchase price, plays a role here.

Moreover, as we measure a scene pixel by pixel for very fast phenomena, we are limited by the acquisition time. A possible solution to most of these problems is the use of compressed sensing techniques, specifically the CASSI method (see Chapter 1.4). For example, Gao et al. [32] accomplished to capture events happening on the order of tens of ps – a reflection of a laser pulse on a mirror (Figure 6).

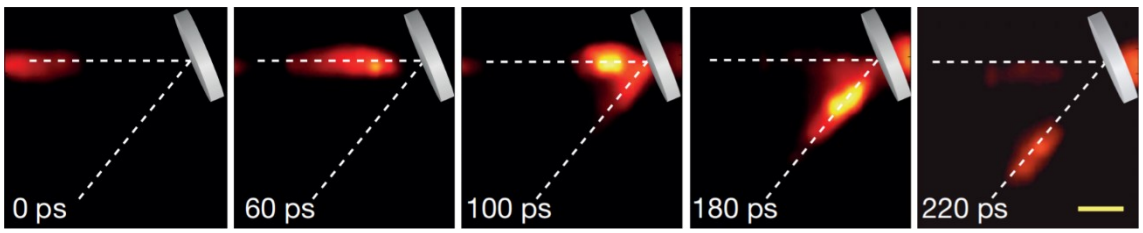

Figure 6 Photos of a laser pulse being reflected on a mirror, scale 10 mm. Adapted from [32]

2.2 CASSI

Thanks to CASSI, one does not need to scan the scene because the method is based on only one image from the camera, i.e., a single snapshot. It is captured in a single moment, and this image is much smaller in size than the entire hyperspectral datacube. Then it is possible to reconstruct the scene back thanks to the knowledge of the random mask and that common images are sparse. By not having to use a slit or narrow bandpass filters as in the standard hyperspectral approaches (see Chapter 1.1), the utilization of light is significantly improved, which is absolutely critical for real-world applications. Also, naturally, there are no scanning artifacts. In Fourier-transform infrared spectroscopy, the optical throughput gained compared to the standard spectrometers using slits is called the Jacquinot advantage [33]. Similarly, using snapshot imaging in HSI, the throughput improvement in comparison to the scanning-based systems is referred to as the snapshot advantage [34].

However, for a wide spectral range, taking only one snapshot brings in a large compression between the data recorded on the detector and the hyperspectral datacube we want to reconstruct. This issue is typically overcome by limiting the sensing to a narrow spectral width (100-190 nm) [9, 11], which is discretized to a relatively small number of wavelength regions (25-28 spectral slices). It is, in fact, very counterproductive. For example, if we are trying to detect the absorption spectrum of a chemical substance, we need a much finer wavelength resolution for its characterization. For instance, measuring on a spectral range of 8-14 μm, 28 spectral slices would provide a resolution of only approximately 214 nm. In this way, any sharp absorption peaks of a chemical compound would be undetected. The standard CASSI also exhibits limitations in image quality for complex scenes and the time needed for the hyperspectral datacube reconstruction, which can reach as long as 14 minutes [9] since the problem in Equation 6 is severely underdetermined.

Attempts to improve the reconstruction quality and the compression ratio are usually implemented using multiple snapshots of the same scene [35, 36, 37, 38, 39], while some of them also try to optimize the coded aperture [37, 38, 39]. But then, the CASSI method loses its uniqueness in recording the hyperspectral data in one instance. Moreover, some advanced modulator to change the random mask pattern is required, or the optimized masks are spectrally selective. There was also an effort to use a higher-order discretization model image for reconstruction [40]. However, the used complex model describing the detector increases the computational demands highly, as it calculates with 170 spectral bands instead of 8.

A promising way to enhance the CASSI performance is to capture a non-diffracted image of a scene that provides more knowledge about the measured scene [11, 41, 42, 43, 44]. Nevertheless, attaining the non-diffracted image requires splitting an incoming beam, and a second detector is needed, which makes the optical system even more complex. The light splitting can reach as much as 50% intensity loss in the measured spectrally sheared image [41].

In addition, due to the use of a random mask that encodes the image for each wavelength, the imaging quality of the system is critical in the whole measured spectral range. It is needed to attain the image without distortion or chromatic aberration to obtain good-quality reconstruction in the standard CASSI experiment. Otherwise, these discrepancies would lead to wrongly encoded information, and it would subsequently cause problems in reconstruction. It is easily feasible in the visible spectral region, where various complex lenses corrected for aberrations are available off-the-shelf. However, in the IR spectral region, this is more challenging. Complex IR optical systems are costly, and their adjustment is highly demanding.

Another criterion, which is essential for using CS in HSI, is the time required to reconstruct hyperspectral data. This information is usually intentionally not stated in the literature. An example of a rare case where the computational times are present is Ref. [9]. Here the authors declare that for 100 iterations of the reconstruction algorithm, a time of 14 minutes on a desktop PC was required (datacube with dimensions of 128×128×28 pixels, spectral range 540-640 nm). The authors do not explicitly state how many iterations were needed for the overall data reconstruction. Nevertheless, assuming the reconstruction was restricted to 100 iterations, the time required is still not suitable for real-life processing. The same group in Ref. [45] demonstrated the ability to capture 248×248×33 datacubes at video rates (30 fps). The catch is that the postprocessing took several hours of computing time to reproduce the video datacube sequence.

Based on the above-listed shortcomings, Chapter 2.4 sets out the objectives of this thesis.

2.3 Application of HSI and CS

In this chapter, only a few selected applications of hyperspectral imaging and compressed sensing will be mentioned. Since these are very broad topics ranging from agriculture to medical applications, it highly exceeds the scope of this work.

Thanks to the conventional HSI, it is possible to noninvasively determine the quality of food [5, 46] and drugs [12]. It is utilized in many applications ranging from scientific research, such as imaging the chiralities of single nanotubes [47] or volcanology [3], to real-world problems involving medical imaging [48], food analysis, and safety inspection [49, 50], forensic sciences [51, 52], criminology [53], art conservation [54], or agriculture [55].

These applications typically employ the IR spectral region, as each chemical has a particular absorption spectrum in this region. A broad range of application fields leads to extensive research in IR HSI. Compared to the HSI "mainstream," the exploitation of CS in HSI is far more seldom. It focuses mainly on the VIS-NIR spectral region because it is easily manageable in the sense of optical elements and detectors' availability, alignment, etc. The VIS-NIR applications will not be mentioned in this brief summary since we are mainly interested in longer wavelengths.

Our main interest lies in MIR and LWIR, which are commonly used abbreviations for the spectral ranges of 3-8 μm and 8-15 μm, respectively. Note that the delimitation of these ranges slightly varies throughout the literature. To mention a few from the MIR and LWIR spectral regions, Ref.

[56] from 2021 provides a proof-of-concept optical setup operating on 3-5 μm, which is able to capture a hyperspectral datacube of 64×48×100 with the acquisition time within one minute. It uses a digital micromirror device (DMD) for spatial encoding, which has to be modified to operate in the MIR spectral region. To the best of my knowledge, this is the only existing work on MIR hyperspectral imaging based on compressed sensing. Regarding LWIR, there are very few publications [57, 58, 59], but note that the list might not be exhaustive. References [57], [58], and [59] all present a LWIR hyperspectral imager using single-pixel detection technology, which collects all three dimensions on a single detection element. However, the experimental results are very limited. To this day, I was not able to find any work regarding LWIR CASSI.

2.4 Objectives

Here we will set the main objectives and goals that triggered the work in this thesis. The overall goal is to localize and distinguish between different chemical substances in the IR spectral region. Therefore, we aim to obtain HSI in a broad spectral range covered with many (up to 100) spectral frames. This is in contrast to previous CASSI reports, which used narrow spectral width and limited spectral resolution. The studied HS camera operating in the VIS spectral range served as a model system for testing the real-life CASSI datacube retrieval and for an outlook of IR CASSI imaging.

This is to be done by keeping the following points:

(1) We aim to retain only one snapshot or, to be more precise, capture all the information in one instance since it is the main advantage of CASSI. Also, we want to keep the optical system as simple as possible, i.e., not using a second detector nor any advanced light modulators.

(2) The demand for reconstructing real-life complex datacubes implies that we need to improve the compression ratio of the measurement.

(3) We aim at using a relatively simple optical system suffering from certain optical aberrations, which can potentially also be created in the IR spectral region.

As we show in the following sections, we reached the set objectives by modifying the optical setup while retaining the simplicity of the system. And more importantly, it leads to a significant improvement in compression ratio and consequently also the retrieved datacube quality. The simulations of our modified setup demonstrate the feasibility of such a system in the IR spectral region.

3 Experimental part: putting it together

After introducing the theoretical fundamentals in Chapter 1, we can now proceed to the work to which I have devoted the last years. First, I will describe our experimental setup, its parameters, and the modifications that were necessary to push the limits of the CASSI method. Then, I will discuss the data processing, simulations, and reconstructions that were performed using Matlab. As a reconstruction algorithm, we use TwIST (Two-Step Iterative Shrinkage/Thresholding) [60] for image restoration during the reconstruction to minimize Equation 6. The reconstruction procedure is described in detail in subsection 3.2.2.

3.1 Hyperspectral detection system

The hyperspectral (HS) camera based on compressed sensing built within this work is an extension of a compact and robust hyperspectral detection system (HDES) for the visible and near-IR spectral region [27], which was designed and manufactured in TOPTEC research center by Jan Václavík. HDES was built through extensive optimization, mostly based on off-the-shelf optics with three custom-made optical elements. It features a low f-number (F/3.9) and has a concentric mounting (see Figure 7). It works on the CASSI method. Hence, the system is missing any moving or scanning part while it is able to acquire the image and spectral information in a single snapshot. Moreover, the compactness of the system consists in the possibility of taking the apparatus outside a laboratory, contrary to the commonly used CS table-top setups.

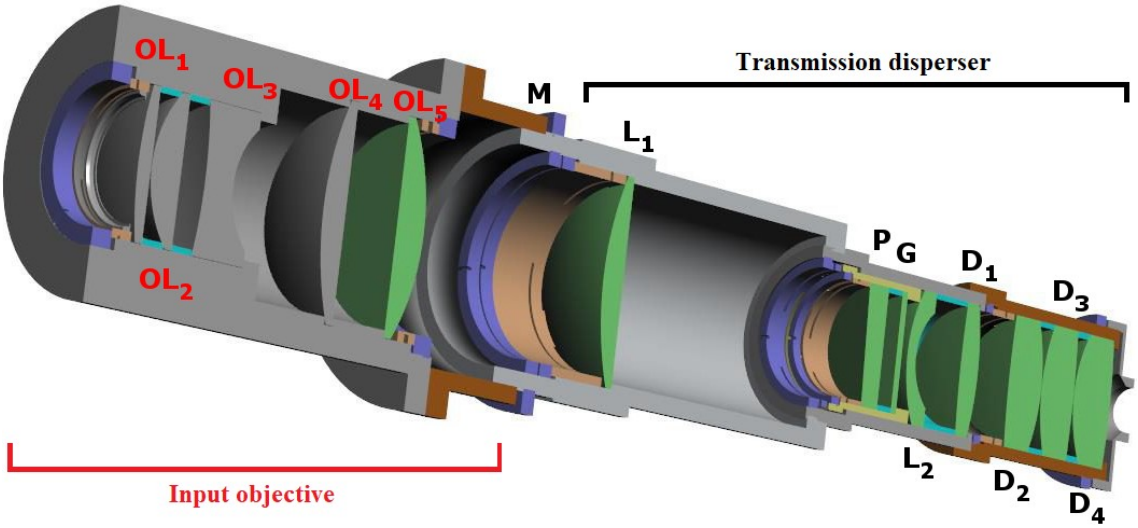

Figure 7 A cut through the HDES camera. Input objective elements (red) and transmission disperser elements (black) separated by a random mask M. OL and L are lenses; D denotes doublet; P and G are prism and grating, respectively. Adapted from [27]

The functionality was demonstrated during the initial experiments by the spectro-temporal reconstruction of a scene based on random matrix encoding. In Figure 8, you can see the measured scene of three laser spots (panel A), the used random mask (panel B), the spectrally shifted snapshot of the scene (panel C), and three selected reconstructed datacube slices at different wavelengths (panel D). We can see that for a simple scene with sharp spectral features, the standard CASSI system works fine. However, real-world scenes are nowhere near as simple as this tested one. Chapter 5 shows that a simple scene illuminated by a broadband light source is already beyond the ability of a standard CASSI system to perform a faithful reconstruction.

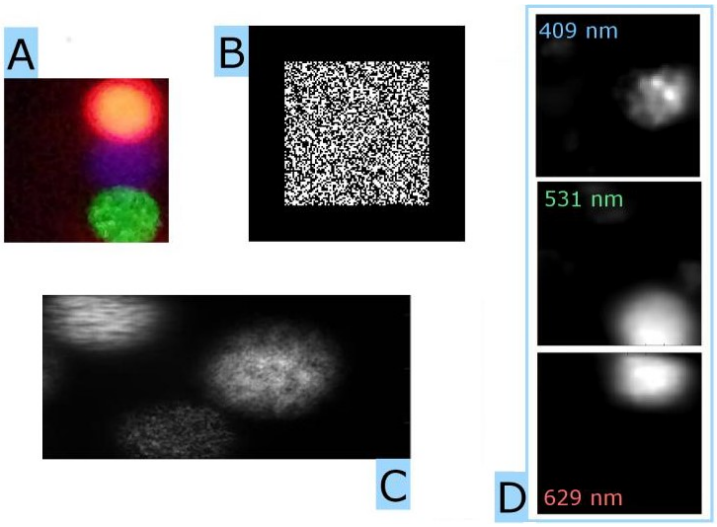

Figure 8 Example of a simple scene (A) reconstruction by using a random mask 90×90 pixels (B). Detected 2D image (C) is used to reconstruct back the image for each spectral frame (D). Adapted from [27]

3.1.1 The original optical system

The optical design of this HS system was carried out mainly with a focus on the availability of the components and their price. Specifically for this system, custom-made lenses OL_3, L_2, and the prism P were produced – see Figure 7 or Ref. [27], where a more detailed description and a list of the optical elements with their specifications can be found. The construction of the optical system can be divided into two detachable parts separated by a coded aperture (random mask). The front part is the input objective imaging the scene on the random mask, while the rear part is the transmission disperser tracing and dispersing the encoded scene to the detector (CMOS IDS camera, UI-1495LE-M).

The input objective is composed of five lenses, was optimized for good performance in low-light conditions, and has a wide field of view of ±10 degrees. The transmission disperser comprises five commonly available optical elements – lenses and doublets, together with a transmission grating G (300 lines/mm) and the previously mentioned custom-made prism P and lens L7.

The encoded image in the CASSI method is shifted for each wavelength before it hits the detector, which was ordinarily in previously published setups secured by a prism or a grating. Either way, this approach causes optical axis folding, which is linked to secondary problems with the adjustment of the system and its stability. Hence, a combination of the prism P and grating G is

justified as it allows for concentric construction and enables a robust single-tube housing for the optics.

3.1.2 Modification

Our effort to make the CASSI method perform better comprises the following changes made to the optical system. The detector was replaced with the Manta G-507 camera (Sony IMX264, resolution 2464×2056). The resolution of the random mask pattern, as well as its physical dimensions, was revised. We created two complementary random masks (64×64 px, 13.55×13.55 mm) on top of each other (see Figure 9). The double mask was prepared by etching a thin chromium layer on a glass BK7 substrate via photolithography. The masks contain a set of guiding points located outside of the mask region, which can serve for the alignment of mask and dispersion elements.

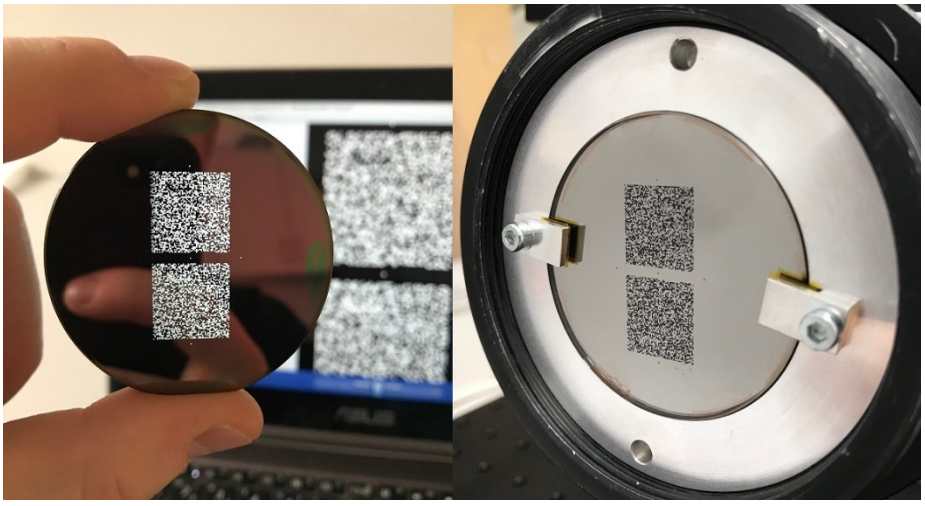

Figure 9 Complementary random masks prepared via photolithography and their implementation in the optical system – see Figure 10 for the random masks location denoted as M.

The input objective part was replaced with a double lens L_D and a field lens D_F (Thorlabs AC508-200-A-ML, f=200 mm, Ø2"). The double lens was created by cutting two lenses (f=100 mm) into a rectangular shape of size 75×50 mm and glued together (see L_D in Figure 10). For the list of all elements and their specifications, see Supplemental information in [61].

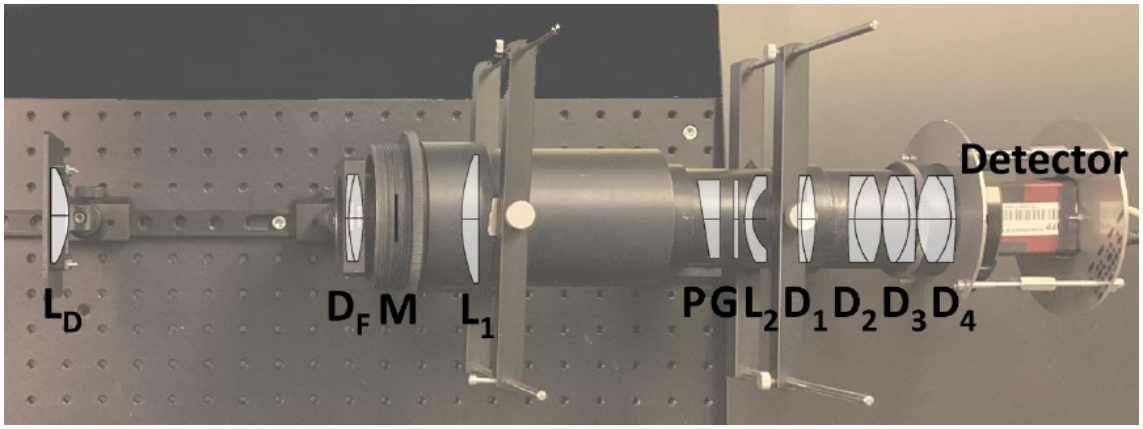

Figure 10 Scheme of the modified optical setup. L denotes lens; D indicates doublet; M is the random mask; P and G are prism and grating, respectively. Adapted from [61]

More details and reasoning behind these system modifications can be found below in Chapters 5 and 6.

3.1.3 Spectral efficiency

In the system, we use a blazed diffraction grating with optimized efficiency of first-order diffraction for 500 nm. Therefore, we expect that the spectral efficiency differs for the FO and the ZO, i.e., the FO light intensity on the detector will be different from the ZO light intensity in dependence on wavelength. This is an important property to measure since we wanted to combine the image information from the FO and ZO diffraction.

We measured the spectral efficiencies of FO and ZO diffraction – referred to as η_{FO} and η_{ZO}, respectively. We used a monochromator (Chromex 250 IS) to illuminate a single spot on the random mask with a quasi-monochromatic light with varying wavelengths from 440 nm to 900 nm for 17 different wavelengths in total. Selected spectra from the monochromator can be seen in Figure 11 on the left. Their correct FWHM (full width at half maximum) is around 4 nm, but the spectral shape is not entirely regular. The width of the spectral lines in the order of nanometers does not have any significance for determining spectral efficiency.

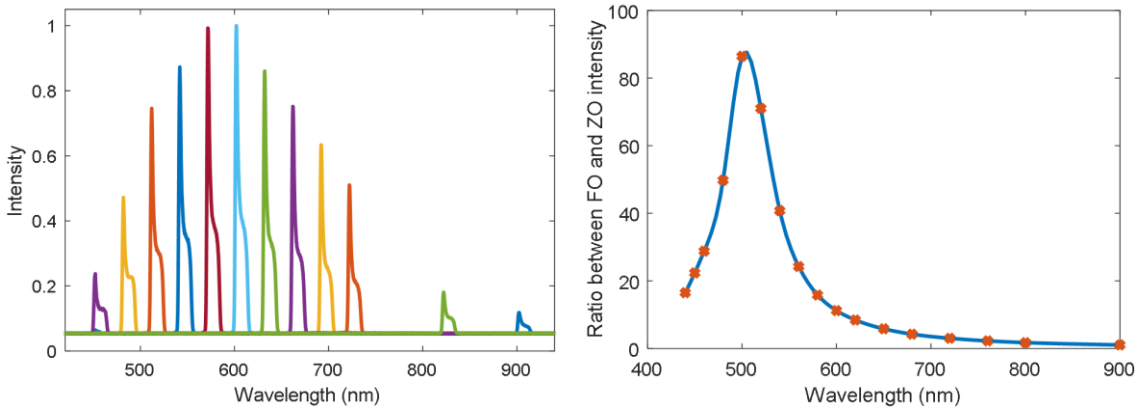

Figure 11 Selected spectra of quasi-monochromatic wavelengths used for the FO vs. the ZO intensity calibration (left) measured by Flame spectrometer from Ocean Optics. Spectral dependency of the relative FO and the ZO intensity (right) – red marks are measured points, blue line is a fit to the data by spline.

For each wavelength, we obtained an image of the illuminated spot on the detector. In fact, for each wavelength, there are two spots on the detector – one is the FO, and another is the ZO, naturally. Consequently, by the ratio of the mean intensities between those two spots, we are able to assess the dependency $\eta_{FZ} = \eta_{FO}/\eta_{ZO}$, which is plotted in Figure 11 on the right. The η_{FZ} was used in the reconstruction algorithm for the accurate formation of the detector image in Equation 6 as well as for the initial guess – see Chapter 3.2.3. The largest intensity difference is around 500 nm, where the FO is more than 80 times stronger than the ZO, which is the result of the grating efficiency optimization. Then the ratio rapidly decreases up to 900 nm, where it is approximately equal to one.

3.1.4 Analysis of aberrations

During the estimation of the spectral efficiency of the FO and the ZO image, we illuminated a single spot on the random mask for different wavelengths – see Chapter 3.1.3. If we superimpose the detected FO images of these spots, we attain a set of distinct dots because each wavelength is projected onto a different position on the detector. It provides us with a convenient visualization of the aberrations present in our system – see Figure 12. It shows the aberrations of FO images (spots) of both upper and lower random masks. The wavelength of the spots decreases from left to right.

We observed the wavelength-dependent vertical shift of spectral slices on the detector and changes in the size and acutance of the corresponding spectral slice. Through careful analysis, it is noticeable that the greatest discrepancies manifest in wavelengths below 500 nm. The images of these

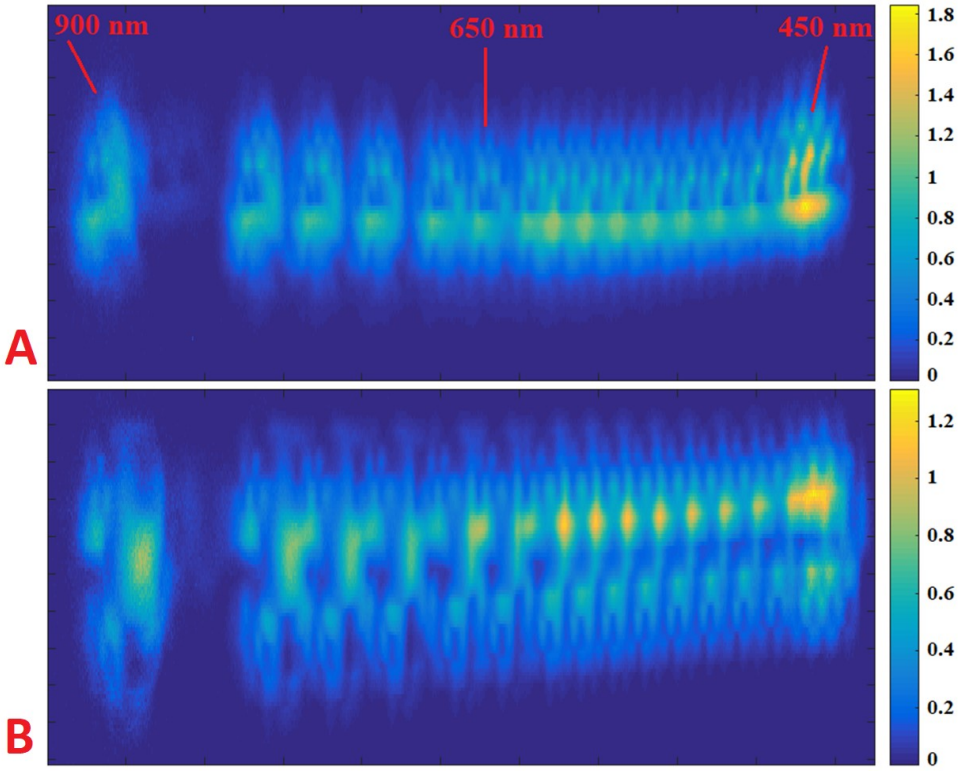

Figure 12 Visualization of aberrations of the system – superimposed normalized FO images of a single spot on the random mask illuminated by different wavelengths ranging from 440 to 900 nm. (A) Upper FO images. (B) Lower FO images. Images resolution is 100×1050 px. Adapted from [62]

wavelengths are not horizontally aligned but are shifting on the detector upward. This shift is more prominent in the upper FO image – see Figure 12A. For the range of 440-500 nm, the shift was 0.5-1 mask pixel, which would inevitably corrupt the reconstruction due to the incorrectly encoded information. Moreover, we found out that the second-order diffraction image for wavelengths under 500 nm interfered with the first-order image for wavelengths above 800 nm. Therefore, to solve both these issues, we used the OG-515 cutoff filter to get rid of the wavelengths under 500 nm.

Note that in the spectral range of 520-720 nm, where the light source used is the most intense, the change in the vertical shift is less than 0.25 mask pixels. This is not expected to disrupt the

reconstruction, as we excluded the border pixels between the lines of the mask during the data processing procedure – see Chapter 3.2.1.

Based on this analysis, we choose suitable parameters for the random mask – 64×64 px with a total side length of 13.55 mm. The overall size of the mask is given by the imaging optics and the camera size. The fineness of the mask pattern must be kept so that the aberrations do not impose wrong encoding of information, as the intensity from one mask line on the detector would be mixed with the neighboring ones.

3.2 Data processing and reconstruction

The captured image by our HS camera cannot be fed to the reconstruction algorithm as such – there has to be some data processing preceding the datacube retrieval. The array detector used has a high resolution (2464×2056 pixels), so the detected image has to be downsized to match the random mask resolution. One mask pixel corresponds to approx. eight pixels on the detector. For the extension of CASSI in Ref. [62] (Chapter 5), where we used a single mask, the first-order image 64×186 px and 64×64 px zeroth-order image have to be extracted. For the extension in Ref. [61] using two masks above each other (Chapter 6), a 128-line image is obtained – i.e., 128×186 px FO image and 128×64 px ZO image. In this sense, when we take into account the space between the two random masks, a detector with approximately 145×275 px would be sufficient to capture the complete information.

3.2.1 Calibration

First, a calibration procedure was done based on homogenous illumination of the random mask by monochromatic light (Nd:YAG laser, 532 nm). The captured image is cropped loosely based on aim pointers in the mask proximity, which are vertically displaced from the mask to not interfere with the image. Then the precise cropping is done through correlation with the random mask. In the image, the illumination in the vertical axis is not homogenous. This flaw was compensated by a flat-field correction, where the correction curve was obtained from the image as a sum of rows.

Secondly, the FO and ZO cropped image of the laser-illuminated mask was rescaled to coincide with the random mask resolution. Because of the limited acutance of the image, the pixels at the border of mask pixels contain information from both neighboring mask pixels. Therefore, we leave out those pixels during rescaling in order to enhance the contrast in the resulting image. This omission was applied in both directions for the ZO image but only in the horizontal direction for the FO image since it is the direction of spectral shearing.

By the calibration step, we attained the positions of the FO and ZO images and their scaling factors with very high precision. This allowed us to rescale in a corresponding way the actual measured data.

3.2.2 Reconstruction

The rescaled image of the detector, denoted as I, was fed to the reconstruction algorithm, whose core is the TwIST algorithm [60], and which minimized Equation 6. The used regularizer $\Phi(D)$ that ensures sparsity of the reconstruction in the wavelet domain was, in our case, l_1-norm – $\Phi(D) = \sum|D|$. Another part of TwIST is a denoising function $\Psi(D)$ which, as the name suggests, denoises the datacube D. To be more specific, we used soft thresholding as the denoising function:

$$Y = sign(X).(|X| - T)_+ \qquad \textit{Equation 16}$$

where

$$(x)_+ = \begin{cases} x & if \quad x \geq 0 \\ 0 & otherwise \end{cases} \qquad \textit{Equation 17}$$

In plain words, soft thresholding puts small wavelet coefficients equal to zero, i.e., getting rid of the wavelet components in the image, which include noise. In Equation 6, the operator $\widehat{W}$ describing the passage of the light through the system is dependent on the extension of CASSI we use. For the standard CASSI, it is formulated by Equation 7, where the spectral shearing operator $\hat{S}$ can be written as:

$$\hat{S} = \sum\nolimits_{\lambda} \left[\eta_{FO}(\lambda)\hat{T}_{FO}(\lambda)\right] \qquad \textit{Equation 18}$$

where $\eta_{FO}(\lambda)$ is the spectral efficiency of the FO intensity – the ratio of the light coming to the detected spectrally sheared image, $\hat{T}_{FO}(\lambda)$ is the spectrally-dependent translation of the image to the detector area. While for the extension by ZO image in Chapter 5, the operator $\hat{S}$ changes to:

$$\hat{S} = \sum\nolimits_{\lambda} \left[\eta_{FO}(\lambda)\hat{T}_{FO}(\lambda) + \eta_{ZO}(\lambda)\hat{T}_{ZO}\right] \qquad \textit{Equation 19}$$

where $\hat{T}_{ZO}$ represents the spectrally-independent translation of the ZO image and $\eta_{ZO}(\lambda)$ is the spectral efficiency of the ZO intensity (see Chapter 3.1.3).

Finally, for the double lens system in Chapter 6, the operator $\widehat{W}$ in Equation 7 changes to:

$$\widehat{W} = \hat{S}\widehat{M}\widehat{D}\widehat{H}^{-1} \qquad \textit{Equation 20}$$

where $\widehat{H}$ is a wavelet transform – Haar or Symlet 8, $\widehat{D}$ denotes imaging the scene with the double lens, $\widehat{M}$ is the modulation by the random double mask. The operator $\hat{S}$ remains the same as in Equation 19.

The operator $\widehat{W}$ is one of the two fundamental procedures in TwIST – it transforms the datacube to the detector image. The other one is $\widehat{W}^T$ which does the opposite – it transforms the detector image into the datacube. TwIST can function utilizing these operators as two matrices, where $\widehat{W}^T$ is the transposition of the $\widehat{W}$ counterpart. Nonetheless, using matrices is unreasonable since the matrices would need to be immensely large with a vast majority of elements equal to zero, and the minimization process would be, consequently, very slow. Therefore, we define the operators as two functions.

We have comprehensively tested a number of approaches to improve image retrieval using both experimental and synthetic datasets. This consisted of the application of direct cosine transform in the spectral dimension to increase the sparsity of the HS datacube, using algorithms based on gradient projection for sparse reconstruction and alternative directional method of multipliers, or employing different types of wavelet transforms. All of the above tests exhibited better results in specific types of scenes while generally performing worse. We will discuss the comparison more closely in Chapter 6.6.

3.2.3 Initial guess

An important step in a datacube reconstruction, before the iterations even begin, is making an initial guess of the datacube, which is done based on the available data. We developed an initial guess procedure for the standard CASSI as well as for its extensions. It is worth noting that the initial guess can be relatively faithfully retrieved only for the extensions because it captures the spectrally integrated image of the scene. The best initial guess resemblance regarding the SSIM index exhibited the double projection approach *Double* mode (see Chapter 6, Table 2). Therefore, here we present how the initial guess for this mode was created in several steps. We denote a HS datacube spectral slice as $\Gamma(\lambda)$ for wavelength λ.

1) The spectral slice was selected from the detected FO image I on the corresponding horizontal position, and it was multiplied by the random mask: $\Gamma'(\lambda) = \widehat{M}\widehat{T}_{FO}^{-1}(\lambda)\, I$. Note that I consists of an upper and lower FO image; hence $\Gamma'(\lambda)$ consists of an upper and lower image, as well. We denote them as $\Gamma'(\lambda)_u$ and $\Gamma'(\lambda)_l$, respectively.
2) Then we summed the upper and lower spectral slice $\tilde{\Gamma}(\lambda) = \Gamma'(\lambda)_u + \Gamma'(\lambda)_l$.
3) The spectral weight of the HS slice was determined as a sum of all pixels of the slice $\tilde{\Gamma}(\lambda)$: $w(\lambda) = \sum_{x,y} \tilde{\Gamma}(\lambda)$.
4) The ZO image Z extracted from the detector was normalized to its mean value $\bar{Z}$ and applied to the spectral slice: $\Gamma(\lambda) = \frac{Z}{\bar{Z}}(\tilde{\Gamma}(\lambda) + Z)$.
5) The resulting initial guess $G(\lambda)$ was calculated by subsequent processing of the $\Gamma(\lambda)$ HS slice with denoising $\widehat{N}$ via the Rudin-Osher-Fatemi denoising model.
6) The denoised slice was multiplied by its spectral weight: $G(\lambda) = w(\lambda).\widehat{N}\Gamma(\lambda)$.
7) The datacube guess G was rescaled so that the mean value of the $\widehat{W}G$ was equal to the mean value of the detector image I.

4 Evaluation of using standard coded aperture imaging in the IR region

This chapter provides a short overview and more insight into the paper Evaluation of using coded aperture imaging in the mid- and far-infrared region [63].

The central goal of this paper was to test the feasibility of using the standard CASSI method in the IR spectral region. We performed numerous reconstructions of artificial hyperspectral scenes, which included a spill of a chemical agent. Implementation of CASSI operating in the IR would enable less expensive and simpler construction of HSI devices.

The simulations included a simple and complex scene, illumination by black-body radiation, and the effect of noise on the reconstruction quality. Several different sizes of the random mask (32x32 – 512x512 px) were evaluated as well as a different number of spectral slices (117-470) and varying concentrations (100-2000 ppm-m) of the chemical compound. The number of spectral slices reflects the spectral resolution, while varying the concentration impacts the intensity at specific spectral slices in proportion with the absorption spectrum of the chemical.

Throughout Section 4, we employed the standard CASSI method.

4.1 Data preparation

In terms of data preparation, we multiply the 2D scene $S(i,j)$ by the 1D radiation spectrum of the light illuminating the scene with or without Planck's law (black-body radiation) – $R(\lambda)$. In this sense, we obtain the 3D datacube $D(i,j,\lambda) = S(i,j).R(\lambda)$. The chemical compound was positioned in the central part of the scene. It means that we multiplied the central part of the datacube $D(k,l,\lambda)$ by a transmission spectrum of the studied compound $C(\lambda)$ for the "contaminated" scene pixels k,l: $D'(k,l,\lambda) = D(k,l,\lambda).C(\lambda)$. In other words, the original datacube was modified for the pixels with indices k and l with the transmission spectrum of the compound. At the same time, it stayed unaffected for the other pixels, as it is depicted in Figure 13C.

The datacube $D'(i,j,\lambda)$ enters the CASSI system, where it is encoded by the random mask $M(i,j)$ for every wavelength – $D''(i,j,\lambda) = D'(i,j,\lambda).M(i,j)$. Then the final detector signal $I(m,n)$ is created by shifting every slice of the encoded datacube $D''(i,j,\lambda)$ by one pixel-column compared to the previous slice, while the slices overlay each other. It simulates the spectral dispersion, and it can be written as $I(m,n) = \sum_\lambda D''(m,n+\lambda,\lambda) + N(m,n)$ where $N(m,n)$ denotes noise. The interested reader should not be confused with the different notation used in the paper [63], as it was changed due to consistency in this work.

4.2 Reconstructions

The reconstructions were performed using the TwIST algorithm [60], which minimizes Equation 6. It recovers the datacube from the fed detected image, and subsequently, we can extract the absorption spectrum of the chemical agent. In order to correctly determine the chemical agent and its concentration, the relative intensity and position of the peaks are important factors.

4.3 Results and discussion of Chapter 4

We obtained a reasonable agreement between the original and reconstructed spectrum – see Figure 13A. The data shown are for a datacube dimension of 128×128×470. The reconstructed spectra were obtained as a mean value of the central part on the position where the chemical was originally located. From Figure 13D, it can be seen that the highest intensity stripe in the upper part of the original image in Figure 13B was partially restored. We can as well, at a very rough guess, estimate where the chemical compound is located – compare the dark part with Figure 13C. Yet, the resemblance is very coarse, and we cannot talk about any resolution of finer details. It is worth noting that the data shown in Figure 13A-D are without any noise, and even 1% of the noise level (approx. 43.5 dB) added to the detected image seriously impacts the reconstruction (Figure 13E-F).

In other words, our results suggest that the standard CASSI could potentially determine the type of chemical agent and solely roughly localize it only for the ideal case, i.e., without any noise. Therefore, to make CASSI work in the IR spectral region, there is a necessity for an upgrade of the method.

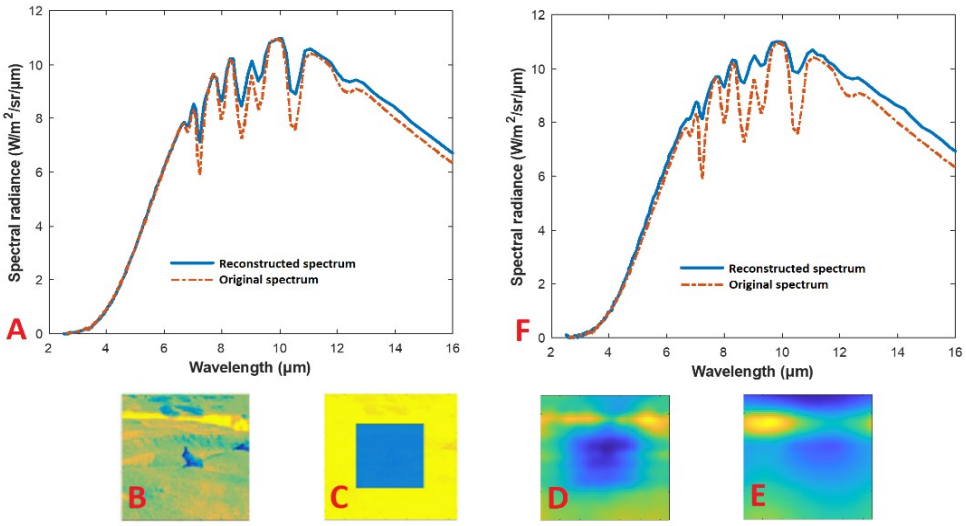

Figure 13 (A) Original (red) and reconstructed (blue) spectra of data without noise. (B) Original scene. (C) A slice of the original datacube with the most prominent absorption in the central part. (D) A slice of the reconstructed datacube without noise. (E) A slice of the reconstructed datacube with 1% of noise. (F) Original (red) and reconstructed (blue) spectra of data with 1% of noise. Adapted from [63]

5 Extension of CASSI by a zeroth-order image

This chapter provides a short overview and more insight into the paper Enhancement of CASSI by a zero-order image employing a single detector. [62]

In order to improve the reconstruction fidelity, there was a need to find a way to gain more information about the measured scene. As mentioned previously (Chapter 2), one way to improve the CASSI method is to capture the non-diffracted image on a second detector [11, 41, 42, 43, 44]. However, this modification makes the imaging system more complex, and since it typically includes a beam splitter, it causes a substantial loss of light intensity, which could reach as much as 50% [41].

A significantly better approach to keep the simplicity of the system is to utilize the zeroth-order (ZO) diffraction arising from the diffraction grating. This image is inherently present in the systems employing a diffraction grating, but its use has not been reported for the CASSI experiment. Nevertheless, for the realization of this idea, some changes to the HDES system [27] had to be made – see the next subsection.

The results presented in this chapter, as well as in the paper [62], are derived from the upper image created by the double lens (see Chapter 6 and Figure 18).

5.1 Modifications of the optical system for capturing the ZO

For the sake of capturing the whole ZO image on the same detector as the first order, the physical dimension of the random mask had to be adjusted as well as the detection area of the used camera – we employed a large detector (Manta G-507, see Chapter 3.1.2). Due to using a wavelet transform in the reconstruction algorithm (see Chapter 3.2.2), the resolution of the mask was chosen to be 2 to the power of n (2^n). Another constraint of the physical mask dimensions was the aberrations present in the system that emerge from the combination of the simplicity of the optical setup and broad spectral range.

Based on the aberration analysis – Chapter 3.1.4 – it was possible to choose suitable parameters for the random mask, which is 64×64 px with a side length of 13.55 mm. Figure 12 shows that the area around 450 nm has a large vertical shift of approximately one mask pixel. This would inevitably impose problems with reconstruction. For this reason, the OG-515 cut-off filter was used in the measurements to block the spectrum under 500 nm.

5.2 ZO enhanced CASSI measurements

Using our modified CASSI system, we measured several different testing scenes illuminated by a monochromatic light source (Nd:YAG laser, 532 nm), a red diode, or a broadband quartz tungsten-halogen lamp (Thorlabs). In Figure 14, on the left, there is an example of two detector images. The upper one is for an opaque cross illuminated by a green laser and a red diode, and the bottom one is for the same cross but illuminated by a broadband light source. The spectra of the used light sources are shown in Figure 15C-D. The FO of diffraction is on the left, while the ZO of diffraction is on the right. Note that in the ZO image, a random mask pattern is apparent for both scenes. Whilst in the FO image, it can be identified only for the scene illuminated by green laser light and a red diode – Figure 14A. In this case, the spectral shearing is not so prominent, contrary to a broadband light source – Figure 14B.

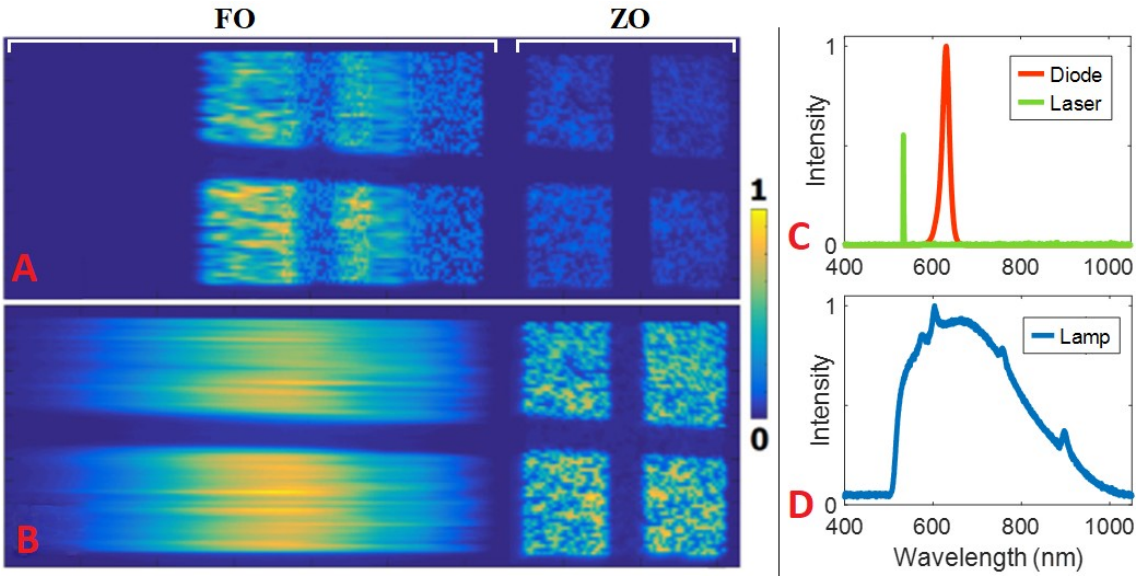

Figure 14 (A) Detector image of an opaque cross illuminated by a green laser and a red diode, and (B) illuminated by a broadband light source. The FO of diffraction is on the left, the ZO of diffraction is on the right. Images resolution is 600×2260 px (C) Measured spectra of the green laser and the red diode, and (D) the broadband light source.

5.3 The effect of using the ZO

The TwIST algorithm [60] used for reconstruction enabled us to feed an initial guess (Chapter 3.2.3) of the datacube, so we do not have to start from a trivial guess implemented in the algorithm [62]. We tested the use of the ZO on the quality of the reconstruction by employing the ZO (i) in the initial guess only; (ii) in the reconstruction itself (in the operator $\widehat{W}$, Equation 6) only; (iii) in both the initial guess and the reconstruction; and (iv) without using the ZO at all. Acknowledging that TwIST is an iterative algorithm, we can say that it is beneficial to set the initial guess as similar as possible to the real data both in terms of reconstruction time and quality.

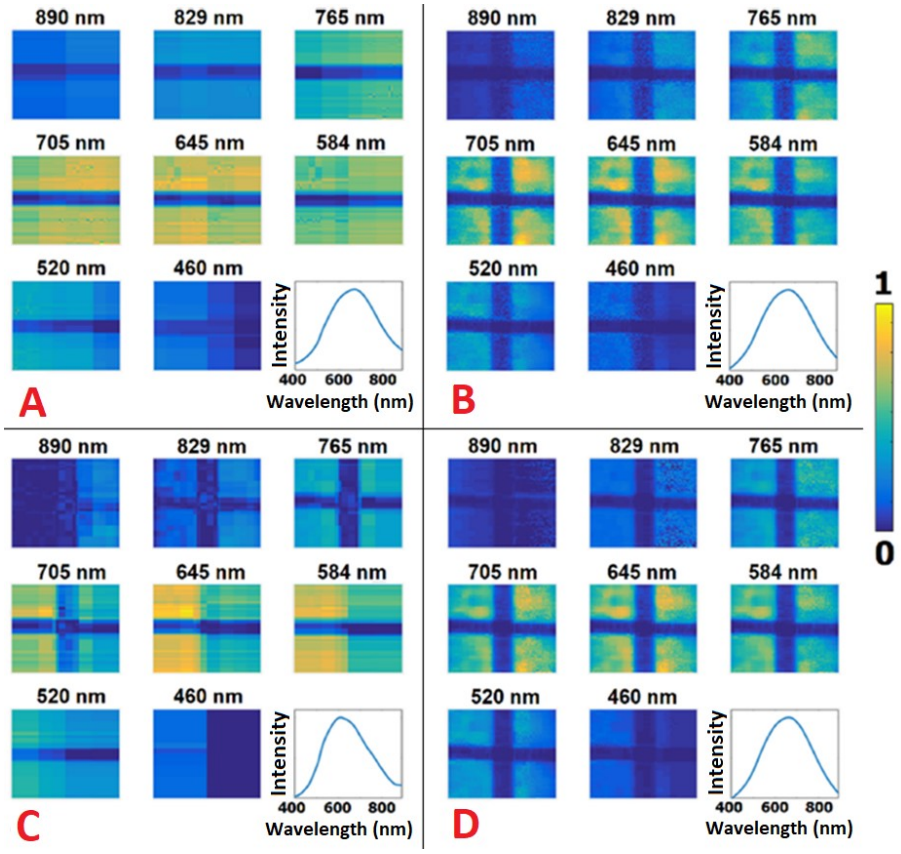

Figure 15 Reconstructions of the scene from Figure 14B; each selected spectral slice is normalized to the maximum datacube value, color bar is on the right. (A) Not using ZO; (B) using ZO in initial guess only; (C) using ZO in operator $\widehat{W}$ only; (D) using ZO both in initial guess and operator $\widehat{W}$. Adapted from [62]

The results summarized in Figure 15 for the cross illuminated with a broadband light show that capturing and utilizing the ZO has a tremendous effect on the spatial quality of the reconstructed datacube. Each quadrant in Figure 15 depicts selected spectral slices at different wavelengths and an overall spectrum of the reconstructed scene for four different approaches to using the ZO image. Panel A shows reconstruction by the standard CASSI method; panel B has a ZO-assisted initial guess along with a standard CASSI reconstruction; panel C comprises using the ZO only in the iterative part of the TwIST reconstruction; panel D includes the ZO knowledge in both the initial guess and the datacube reconstruction.

It is noticeable that without using the ZO – Figure 15A, the CASSI method struggles to restore vertical features in the scene. It is caused by spectral shearing (see the FO image in Figure 14B), which significantly diminishes the restoration of the perpendicular lines of the scene. This effect is more prominent for the measurements of broader spectral regions with many spectral slices where the compression ratio is low. Reconstructed datacube slices in Figure 15C show the significance of having a reliable initial guess. The algorithm was not able to restore the opaque cross correctly in all the slices. The use of ZO in the initial guess is highly improving the spatial quality of the reconstructed datacube, as can be seen in Figure 15, panels B and D.

But even for the most accurate reconstruction in Figure 15D, where ZO is used for both the datacube initial guess and retrieval, the overall spectrum below 500 nm does not resemble the actual one. In this spectral region, the intensity should be equal to zero because of the used cut-off filter.

To some extent, the spectral similarity can be enhanced by using a higher regularization parameter τ, which promotes sparsity (see Equation 6), as we describe in Ref. [62]. Note that in this case, the reconstruction quality could be further improved by limiting the reconstructed range to 500-900 nm. Nevertheless, the area in this region is a good indicator of the retrieved datacube quality.

5.4 Simulations

It is very challenging to assess the reconstruction quality of the experimental data, as we do not have the ground truth to compare to the reconstructed datacube. It is complex to carry out reasonable reference measurements as each hyperspectral camera features a different field of view and spectral response.

At the same time, we gained low detector residuals even for the reconstructed datacubes that clearly did not match the original scene. For instance, the detector residuals of the reconstruction in Figure 15A were comparable to the residuals of the one in Figure 15D, while the quality of the two reconstructions is very different. Therefore, the residuals are not a good measure to assess reconstruction quality in our case. Hence a further evaluation was necessary.

Therefore, we created synthetic detector images from known datacubes (ground truth) faithfully resembling the real detector images. This was possible owing to the rigorous analysis of the aberrations of our system mentioned in Chapter 3.1.4. In particular, we incorporated the effect of the wavelength-dependent vertical shift of spectral slices on the detector and wavelength-dependent spectral slices' acutance.

To quantify the impact of using the ZO, we calculated the lowest attainable difference between the ground truth and reconstructed datacube as a minimum of least squares with optimization of scaling factor a – we denote it as *delta* Δ, see Equation 8.

In Table 1, we can see that ZO usage is beneficial in all cases. Nevertheless, the enhancement level depends on the properties of the imaged scene. It has a more significant effect for scenes with higher complexity, i.e., broadband light illumination, a higher number of spectral features, etc. This is caused by the fact that the simple scenes can be well retrieved by the standard CASSI system.

Scene A is the simplest of the three scenes in Table 1. Hence, the effect is not as large as in Scene C, which is more complex (see Ref. [62]). It is worth noting that while the difference in *delta* between the standard CASSI and the ZO-assisted one might be minor, the ZO utilization provides us with a much more robust reconstruction regarding the change in the reconstruction parameters.

Table 1: Delta results of the original and reconstructed datacube for different scenes [62]

	Scene A	Scene B	Scene C
Not using the ZO	2.11e-03	1.00e-03	9.71e-04
Using the ZO	2.06e-03	9.26e-04	6.04e-04

5.5 Conclusion of Chapter 5

The presented extension of the CASSI method was constructed with a limited number of optical elements based mainly on off-the-shelf optics. It can be employed for systems exhibiting low compression ratios and suffering from aberrations, especially if there is a need to preserve the system's simplicity. The proposed modification of CASSI is unique in the sense that it enables to capture a spectrally dispersed image of a scene as well as a nondispersed one on the same detector.

We measured different scenes on a broad spectral range (500–900 nm) and observed the effect of including the nondispersed scene image in the reconstruction as well as optimizing the reconstruction parameters. For instance, the regularization parameter τ (Equation 6), which promotes sparsity, has a significant effect on the reconstruction quality. Low values of τ enhance spatial quality, while high values improve spectral similarity.

The modification led to improved overall reconstruction quality and an approximately five-fold reduction in computational time. Note that the improvement of results is not only because of the higher compression ratio but also because of obtaining more spatial information – we can set the initial guess very close to the original scene. These effects are more prominent for complex scenes.

The real data findings were confirmed by simulations utilizing rigorous analysis of aberrations, which provided us with quantification of the quality of the results.

6 Extension of CASSI by double projection and differential coded aperture

An important factor limiting the quality of retrieved data via compressed sensing is the so-called compression ratio, i.e., the number of measured data points with respect to the number of elements of the reconstructed dataset. A hyperspectral datacube with dimensions of 64×64×123 px will have a spectrally sheared imprint of 64×186 px on the detector. The compression ratio (CR) is, in this case, 11904/503808 ≈ 2.4%. Using the upgrade from the previous chapter, the CR increases to 16000/503808 ≈ 3.2% since the detector image is now extended by the 64×64 px zeroth order. A significant improvement in the reconstruction quality was achieved. However, the aberrations still limit the reconstruction of complex scenes.

Another way to amplify the performance of CASSI is to take multiple snapshots of the same scene [35, 36, 37]. Yet, in this way, the CASSI system needs some advanced modulator (such as DMD – Digital Micromirror Device) in order to change the random mask pattern. Furthermore, it loses its main advantage of capturing the whole scene in a single instance. Hence, the question is: how to take more snapshots while keeping the simplicity of the system?

6.1 How to take more snapshots while keeping the simplicity of the system?

Let's take the following example into account. In the case of using a DMD for light modulation, there are two branches corresponding to two digital states (positions) of the micromirrors. When a random mask is invoked by a DMD, it creates two reflections corresponding to two complementary random masks. Typically, only one branch is being used, which corresponds to the modulation by a binary mask {0,+1}. That said, there are some applications that successfully exploited both reflected branches, e.g., balanced detection [64]. With the use of a balanced photodetector, which corresponds to the subtraction of the two signals, we gain the random mask {-1,+1}. Using such types of random masks leads to image retrieval with a significantly higher SNR.

Moreover, in a standard CASSI system, using a single mask with {0,+1} elements blocks 50 % of the information at the pixels, where the random mask binary information is 0. Considering the extension from the previous chapter, where the FO and the ZO are missing half of the pixels in the image, thanks to the complementarity of the masks, we now obtain the information from all the pixels and, most notably, the whole nondispersed image of the scene. This would be immensely helpful, especially for the spatial resolution of the initial guess.

This clearly sets our objectives. To get better datacube retrieval, we want to utilize complementary masks. But, at the same time, we want to retain the simplicity of the system, i.e., not using any advanced modulators. A solution is to project the scene in parallel with two lenses through

two different random masks at the same time. The proof of concept is presented in the paper Improving Compression Ratio in CASSI [65]. This chapter provides a short overview and more insight into the paper Differential Coded Aperture Single-Snapshot Spectral Imaging [61].

Suppose we double the information by capturing two projections of the same scene. In that case, we decrease the CR to 23808/503808 ≈ 4.7%, and by also using the ZO, we get 32000/503808 ≈ 6.4%, which is approximately a 2.7-fold improvement in CR compared to the standard CASSI. Note that by using complementary random masks, we can extract more information about the scene thanks to special data processing that has a further improving effect.

6.2 Modifications of the optical system for double projection

The detector size of our hyperspectral camera (Manta G-507, resolution 2464×2056, pixel size 3.45 μm) was chosen so that there would be enough detector area to capture the two spectrally dispersed projections together with the two nondispersed ones. In our optical setup, we use a detector with high resolution. However, a detector with approximately 145×275 px would be sufficient. Note that the spatial resolution of a measured scene is restricted by the resolution of the random mask used, not by the detector itself.

We acquired two complementary random masks (see Figure 9) via photolithography of a chromium layer on a BK7 substrate. The bottom mask is an inversion of the upper one, and both are 64×64 pixels. In order to correctly project the images emerging from the double lens, there is a space of nine pixels left between them, which translates to 1.9 mm in physical dimension. Around the masks, there are guiding pointers that are vertically displaced, so they do not interfere with the detected image. The pointers serve for calibration, cropping of the detected images, and alignment of the system.

The scheme of the optical system can be seen in Figure 10. Doubling the image was achieved by the double lens L_D. Right in front of the transmission disperser (see Chapter 3.1.2) is a field lens D_F, which steers the rays coming from the double lens under a greater angle. For proper imaging, there is also an aperture in the imaging plane of the double lens. The aperture crops the height of the imaged scene – otherwise, the two images would be overlapping. Analogously to the previous chapter, the spectra were cropped by the cut-off filter for wavelengths below 500 nm.

6.3 The effect of random mask complementarity

As stated before, we expect that the mask's complementarity will positively affect the reconstruction quality. To determine the effect of the masks' complementarity also quantitatively, we performed a set of simulations. Clearly, the experimental realization (creating various random masks via photolithography, their implementation to the system, system alignment, calibration, etc.) would be very complex in this case.

For the results further discussed in this section, we considered the case when we also used the ZO for initial guess and datacube retrieval. Since it was proven in Chapter 5 that the ZO

significantly improves the reconstruction quality, we carried out the comparison for the best available configuration.

In the case of complementary masks, the two ZO images could be directly summed, which provided us with a complete picture of the scene without spectral shearing. If we sum two noncomplementary ZOs, we get double values on the positions where the "1" pixels are concurrently in the upper and lower mask, while we will be missing values on the positions with 0s in both masks. Letting the reconstruction algorithm work with this kind of data, naturally, the results would be worsened because of the badly assigned values on approximately 50% of pixels, i.e., approx. 25% of double values, and approx. 25% of no value. We must therefore choose a different approach for relevant comparisons.

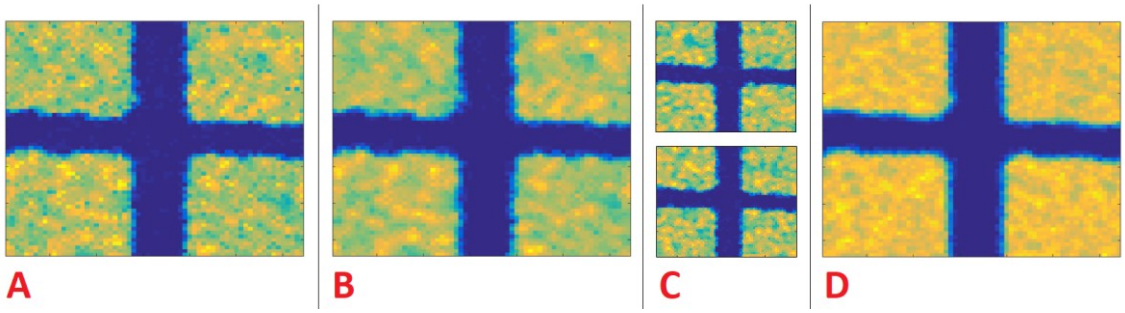

Figure 16 Demonstration of the ZO benefits of complementary random masks. (A) Summing noncomplementary ZO images for two random masks and then calculating missing pixels. (B) Calculating missing pixels for noncomplementary ZO images separately and then summing them. (C) Upper and lower complementary ZO images with calculated missing pixels. (D) Summed images from C.

It is possible to add the top and bottom noncomplementary ZO images together and then divide by two the pixels where there are double values and calculate the pixels with the missing values as an average of the eight neighboring pixels. However, this approach does not work very well due to system imaging imperfections that were included in simulations (blur, noise, etc.). You can see the inhomogeneities caused by this approach in Figure 16A.

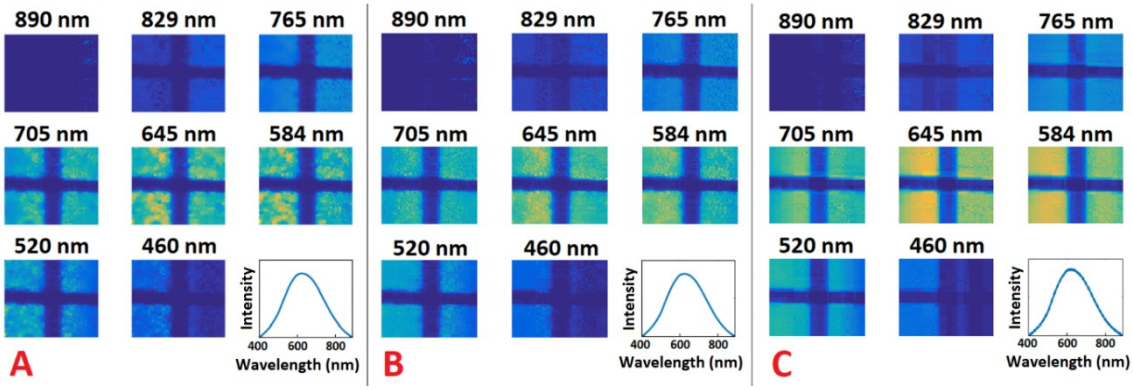

Figure 17 Comparison of simulated data reconstructions of (A) two random noncomplementary masks without calculating the missing pixels in the initial guess, (B) two random noncomplementary masks including calculating the missing pixels in the initial guess and (C) two complementary masks using their sum, Double approach – see Table 2.

Considerably better results were attained by directly calculating the missing pixels as the average of the adjacent pixels for each image separately, then using the mean value of these images – Figure 16B. In Figure 16C are the top and bottom images with calculated missing pixels for complementary masks and their sum in Figure 16D. Here we can clearly see that the resulting ZO image is far more homogeneous, and we can expect it to have a better effect on the spatial resolution of the retrieved datacube.

The quality of the retrieved datacubes can be compared visually in Figure 17. At the same time, we can use *delta* – defined by Equation 8, for the quantification of the reconstruction quality (lower number equals better reconstruction). The values reached 7.23e-04, 6.56e-04, and 6.38e-04 for (i) a reconstruction of noncomplementary masks without pixel calculation (explained above), (ii) a reconstruction of noncomplementary masks with pixel calculation, and (iii) a reconstruction of complementary masks, respectively. The complementary masks featured consistently better properties for different simulated scenes. Even though the difference is not that prominent, we need to take into account also the time required for reconstruction, which is crucial for potential real-time imaging applications. It is notably prolonged for the case of noncomplementary masks since we need to perform extra calculations to obtain the complete ZO image, as opposed to complementary masks, which require only a simple sum of two matrices.

To illustrate this difference, we provide the computational times required on a standard laptop for the initial datacube guess and reconstruction for the results in Figure 17. In the case of complementary masks, the initial guess took an average of 0.6 s. For noncomplementary masks with pixel computing, it took an average of 13.2 s, which is 22 times longer.

Suppose we use the same principle of calculating the missing pixels directly in the reconstruction algorithm. In that case, we have to calculate each slice of the data cube separately, and in our case, for 123 slices, such an approach is very lengthy. We can, for instance, compare the time required for 13 iterations of the TwIST algorithm [60] for the complementary and noncomplementary modes. The algorithm was limited to 13 iterations since, under different initial conditions, the reconstruction may require a different number of iterations. The TwIST algorithm converged in as little as 11 s after 13 iterations using complementary masks – see Figure 17C for the result. For the noncomplementary mask with calculating the missing pixels, the reconstruction time reached 154 s - Figure 17B. The result in Figure 17A took comparable time as the one in Figure 17C but is visibly worse as well as it has a higher *delta*. From the example above, we see that complementary masks outperform noncomplementary ones, both in terms of reconstruction quality and, above all, in terms of speed.

6.4 Differential CASSI measurements and approaches to data processing

The previous subsection showed that complementary masks are the best option for a double-mask CASSI system. Following these findings, the experimental implementation was carried out only for the complementary random masks. In particular, we used two random binary complementary masks of 64×64 pixels.

An example of such a double-mask measurement can be seen in Figure 18B. It is possible to approach the data reconstruction in several ways compared to the standard multi-snapshot CASSI, owing to the mask's complementarity. We denote the standard multi-snapshot approach as *Double*, which means that the two detector images are optimized during the datacube retrieval in parallel. The standard single-snapshot CASSI is represented in this chapter as *Single*. Nevertheless, in our complementary measurements, we can subtract the upper and the lower image, which would correspond to a measurement with a mask composed of ±1s – we indicate it as *Diff*. It is worth noting that CS algorithms work better for {+1,-1} matrices than for {+1,0} matrices because of their compressed sensing performance [64].

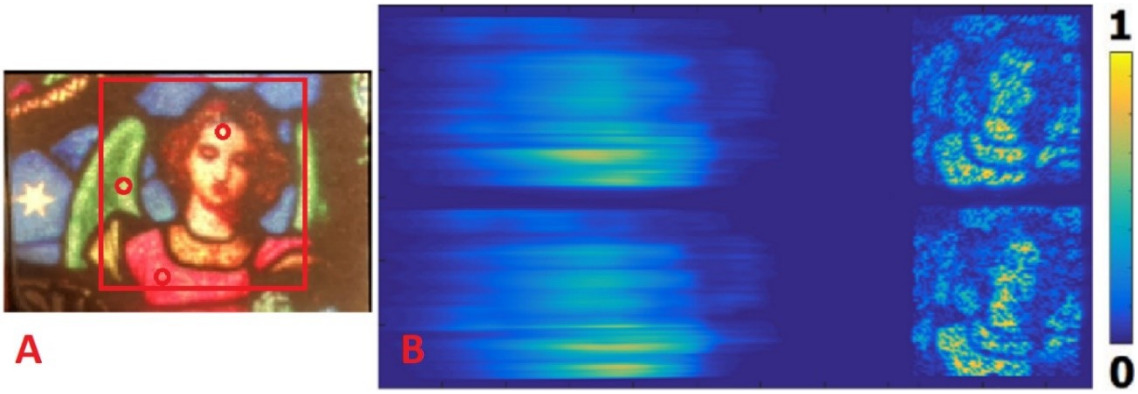

Figure 18 (A) Example of a complex scene with many different spectral features. The red square marks the imaged area. The red circles are used for spectra comparison. (B) The scene from (A) detected by our CASSI system. Note that the same color bar applies to Figure 19, Figure 20, and Figure 22. Addapted from [61]

However, by using only this differential image, we would lose the information about the magnitude – this occurs when one subtracts two similar datasets, which are shifted with respect to each other. Therefore, a better way to process the data is to calculate not only with the difference between the snapshots but also with their sum – the *Diffsum* approach. It is a combination of the differential character of the random mask while it preserves the information about the image intensity scaling. The matrix notations overview of detector images $\overline{D}$ for all the approaches is in Table 2.

Table 2: Different approaches to data processing. $\overline{D}$ denotes detector image. A and B are upper and lower FO images, respectively.

Data processing approach	*Matrix notation*
Single	$\overline{D} = [A]$
Double	$\overline{D} = [A; B]$
Diff	$\overline{D} = [A - B]$
Diffsum	$\overline{D} = [A - B; A + B]$

The final detector image D fed to the reconstruction algorithm is $D = [\overline{D}\ \overline{ZO}]$, where $\overline{ZO}$ is a sum of upper and lower ZO images.

6.5 The effect of data processing approaches

Here, we will demonstrate the difference between the data processing approaches on two scenes – an opaque cross and a stained glass foil illuminated by a spectrally broad light. More examples can be found in Ref. [61]. Figure 19 shows selected slices of the reconstructed datacube for different data processing approaches. As can be seen, all approaches are able to reconstruct spatial information well since we employ the ZO image in all the cases. Nevertheless, the approaches differ mainly in the reconstructed spectrum of the light. The *Single*, *Diff,* and *Double* approaches face a problem with spectrum reconstruction below 500 nm, where the intensity should be zero due to the use of the yellow cut-off filter. In terms of the reconstructed spectrum, we can safely say that *Diffsum* is the best approach.

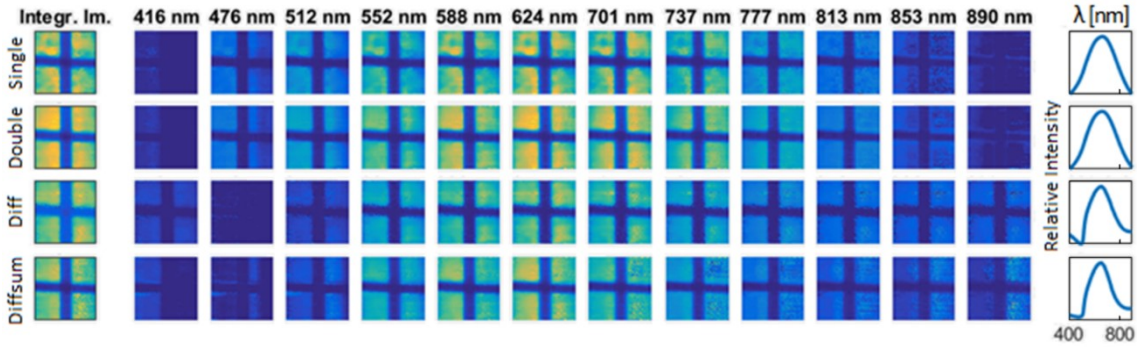

Figure 19 Reconstructed spectrally integrated image, individual spectral slices, and a spectrum of an opaque cross illuminated with broadband light by using four processing approaches (see Table 2). Adapted from [61]

To show the ability of the system to retrieve more complex scenes, we performed measurements of a stained glass foil (see the scene in Figure 18). It is worth noting that the good spatial information obtained is mainly thanks to the extension of CASSI by the ZO image, which was described in the previous chapter. Nevertheless, we can see the improvement as we extend into the double mask approaches, which exhibit much better homogeneity compared to the *Single* approach. See the selected enlarged slices in Figure 20 for comparison. It shows that *Diffsum* exhibit more uniformity as opposed to *Single*.

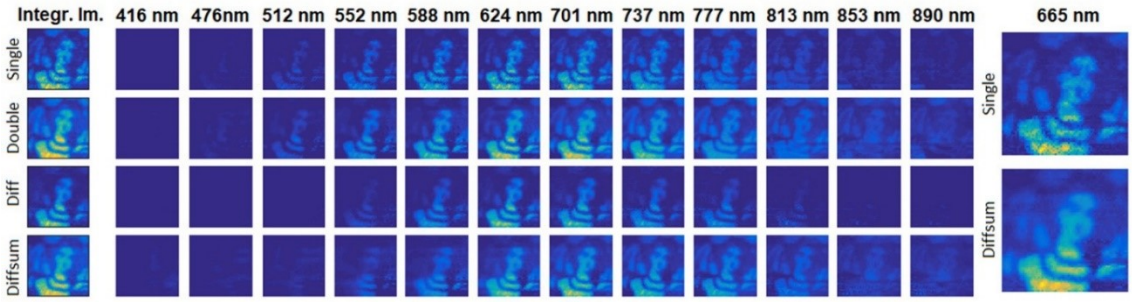

Figure 20 Reconstructed spectrally integrated image and individual spectral slices of an stained glass foil with broadband light by using four processing approaches (see Table 2). Two spectral slices of Single and Diffsum were enlarged for comparison. Extractea from [61]

In Figure 21 are depicted the measured and reconstructed spectra of three points of the stained glass foil scene. The colors of the curves correspond to the colors of the selected points in the scene. The results indicate that, although obtaining a good spatial resolution was possible, the scene is too complex to reconstruct spectra reliably. It is worth noting that the spectra were not adjusted for the spectral efficiency of the system. Nevertheless, the closest similarity was achieved using *Diffsum*.

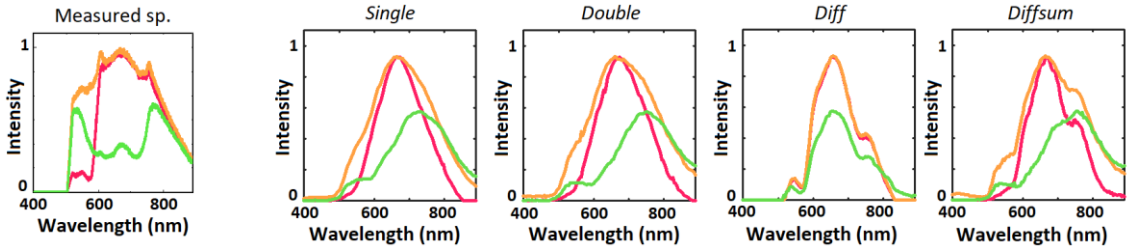

Figure 21 Measured spectra at the points marked in Figure 18A (on the left). Reconstructed spectra at the same points for four different reconstruction approaches (on the right).

6.6 Simulations

The artificial data generation and simulated retrieval were carried out for the same reason and in the same manner as the ones in the previous chapter. However, we extended our evaluation into more complex metrics to evaluate various aspects of the methods – PSNR (peak signal-to-noise ratio), SSIM (structural similarity) index, and SAM (spectral angle mapper), since these are more commonly used throughout the literature.

Table 3 Evaluation metrics – PSNR, SSIM index, and SAM error of retrieved datacubes for different scenes [61]

Approach	*Scene A (cross)*			*Scene B (filters)*			*Scene C (feathers)*		
	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
Single	17.79	0.50	0.39	21.78	0.65	0.53	21.29	0.55	0.58
Double	18.79	0.59	0.39	21.90	0.74	0.52	21.66	0.55	0.57
Diff	16.26	0.35	0.23	21.35	0.66	0.54	19.90	0.46	0.64
Diffsum	19.67	0.59	0.36	22.11	0.72	0.54	21.85	0.56	0.55

An overview of the results is provided in Table 3. Note that we calculated PSNR by scaling the whole datacube by a single factor, not scaling slice-by-slice. The quantitative comparison shows that the double mask approaches, namely *Diffsum* and *Double,* surpass S*ingle*. As we discussed above, *Diff* lacks information about the intensity magnitude. Hence it mostly exhibits the worst results. On the contrary, *Diffsum* consistently leads to the best results for various types of scenes, and it was also stable with respect to the selection of the reconstruction parameters.

As was mentioned in Chapter 3.2.2, we thoroughly tested a range of approaches, including different algorithms, to enhance image retrieval. Yet, for the mask resolution of 64×64 px, we were getting the best results using TwIST.

To illustrate our statement, we provide a comparison of reconstructions between GPSR (Gradient Projection for Sparse Reconstruction) and TwIST in Figure 22. On the left are eight selected spectral slices and an overall spectrum of ground truth data; in the middle, there is a reconstruction using TwIST; and on the right, there is a reconstruction using GPSR. The presented results were achieved after extensive optimization of parameters for both methods. We concluded that the TwIST algorithm provides better or comparable results with GPSR. Hence, we can safely claim that the presented results cannot be easily improved just by using a different reconstruction algorithm.

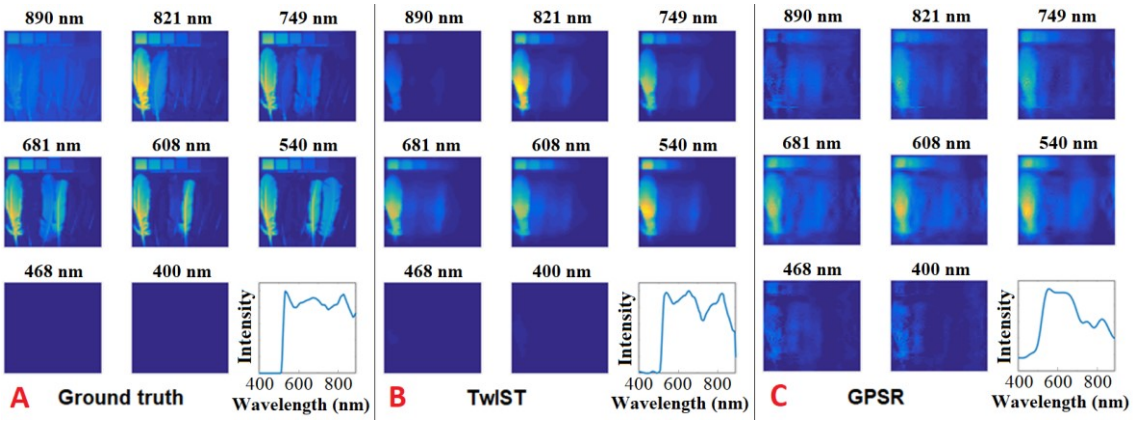

Figure 22 Comparison of ground truth with reconstructions using different algorithms.

During the real measurements, we were confined to a 64×64 px random mask due to the magnitude of the aberrations present in our system. At last, we evaluated the use of different mask resolutions. Table 4 presents the resulting metrics of retrieved datacubes utilizing higher-dimension random masks. For the *Diffsum* approach, PSNR is improved by ~0.8 dB for 128×128 px mask and further ~0.8 dB for 256×256 px mask, while SAM was first worsened and then reached approximately the same value for the highest mask resolution tested. The improvement in PSNR for the *Single* approach was less notable, and spectral metric SAM got even inferior.

Table 4 Evaluation metrics of retrieved datacubes utilizing random masks of higher dimensions [61]

Approach	*128×128 px – Scene C (feathers)*			*256×256 px – Scene C (feathers)*		
	PSNR	SSIM	SAM	PSNR	SSIM	SAM
Single	21.76	0.57	0.65	22.42	0.61	0.61
Double	22.34	0.60	0.62	23.08	0.63	0.56
Diff	19.78	0.48	0.60	20.95	0.54	0.52
Diffsum	22.67	0.60	0.58	23.49	0.63	0.54

6.7 Conclusion of Chapter 6

We showed that by a simple modification of the optical setup, we were able to improve the compression ratio in CASSI systems 2-fold and as much as 2.7-fold using also the ZO image compared to the standard CASSI. Simultaneously, the modification retained the CASSI's main advantage – a single snapshot.

We demonstrated the improvement in the reconstruction quality on a broad spectral range of 500-900 nm. Utilizing the *Double* approach, which is equivalent to multi-snapshot CASSI, it reached an increase of ~1.0 dB in PSNR. We also quantified the superiority of complementary masks over noncomplementary ones. With the use of the ZO image, we were able to make the initial estimate of the datacube very similar to the measured scene, which, on average, decreased the total number of reconstruction iterations needed.

Moreover, we proposed a new approach to data processing which we denoted as *Diffsum* or differential CASSI (D-CASSI) since it utilizes a matrix of {+1,-1} as a random mask. This was possible thanks to the mask's complementarity. PSNR, in this case, soared ~1.9 dB compared to *Single*.

We backed our measurement of real scenes with rigorous simulations, which enabled us to quantify the results. It also allowed us to control the reconstruction parameters better, explore the possibility of using different reconstruction algorithms, compare our proposed approach to standard CASSI covering the whole double projection area, and investigate utilizing random masks of higher dimensions.

7 Evaluation of the CASSI extensions in the IR region

The ultimate goal of the thesis was to evaluate the effectiveness of the CASSI extensions made within this work for hyperspectral imaging in the IR spectral region. Therefore, we created a set of artificial scenes and detector images as it would be produced by an IR CASSI system. As opposed to the visible spectral region, it is also necessary to consider the radiation of a black body, which could be calculated by Planck's law – see Chapter 1.1.1. Compared to Chapter 4, here we implemented both extensions of CASSI, which shift the abilities of the method to a different level.

The random mask dimensions in our optical setup were constrained by the aberrations present in our system, as it was designed mainly on off-the-shelf optics due to the vision of building an analogous setup in the IR. However, the results for higher resolution masks in Ref. [61] showed that using a 256×256 px random mask would lead to better CASSI performance. Therefore, we exploited this knowledge and used it in the following simulations. There is also a practical rule mentioned in Ref. [66], which says that for exact recovery, it is necessary to have about four incoherent samples per unknown nonzero term. I.e., the number of samples is equal to 4 times the sparsity level. It also underlines the idea behind using a higher-resolution mask.

7.1 Data preparation

Hyperspectral datacubes with 256 px in spatial dimensions and 50-100 spectral frames were created using an arbitrarily chosen image from an IR camera (see Figure 23) [67]. The scene was contaminated with a chemical agent (isopropanol or acetone) of various concentrations ranging from 1000 to 8332 ppm-m. The absorption spectra of chemicals were obtained from the National Institute of Standards and Technology [68]. Then, for convenience, it was transformed into transmission spectra as $transmission = 10^{-\alpha.concentration.pathLength}$, where α can be calculated as $\alpha = -log_{10}(spectrum)/(c_0.pathLength)$. The spectral range was set to 8-14 μm since it is the most used range to detect and characterize chemicals – they have their specific signature in this range, and it is also the transparency "windows" in the Earth's atmosphere [13, 69].

To create the hyperspectral datacube $D(i,j,\lambda)$, the scene $S(i,j)$ is first multiplied by black-body radiation described by Planck's law $P(\lambda)$: $D'(i,j,\lambda) = S(i,j).P(\lambda)$. Then a part of the datacube $D'(k,l,\lambda)$ for selected scene pixels k,l is multiplied in spectral dimension by the transmission spectrum of a chemical compound $C(\lambda)$: $D(k,l,\lambda) = D'(k,l,\lambda).C(\lambda)$. The datacube $D(i,j,\lambda)$ enters the CASSI system, where it is doubled by the double lens $\widehat{D}$, modulated by the random mask $\widehat{M}$, and then divided into the diffraction orders by grating and spectrally sheared $\hat{S}$ – see Equation 19, so it creates the detector image $I(m,n)$.

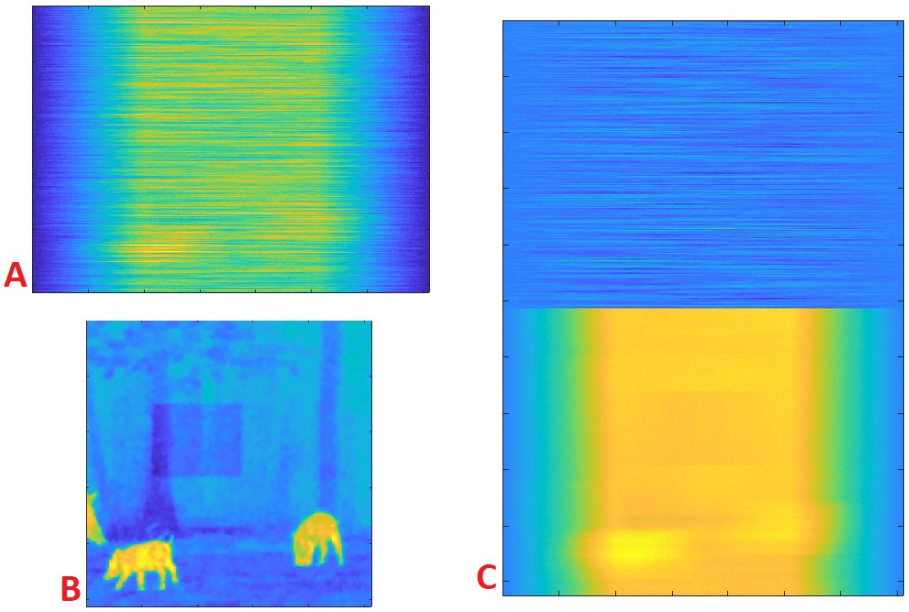

Figure 23 (A) The artificial FO created from an arbitrarily chosen IR camera image – upper FO. (B) The artificial ZO of the same scene – upper and lower ZO images summed together. The rectalngle with visibly lower intensity is an area of the chemical absorption. (C) The differential (upper part) and sumed (lower part) detector image used for reconstruction by the Diffsum approach.

In the practical sense, it means doubling each slice of the datacube and then multiplying with the random mask – i.e., a matrix with ones and zeros. Afterward, the intensity of the encoded datacube is divided in a ratio that 70% of the intensity goes to the first-order (FO) diffraction, 20% to the zeroth order (ZO), and the rest to the other orders, i.e., 10% is lost. This would correspond to a real-world scenario, as diffraction gratings do not typically feature 100% efficiency. For simplicity, these intensities are constant throughout the spectral region. The FO detector image is constituted by shifting each subsequent slice of the encoded datacube by one pixel-column and summing it all together – the slices are overlaying each other. The ZO detector image is made by the integration of all the slices. Finally, Gaussian noise $N(m,n)$ is added, and the resulting detector image is:

$$I(m,n) = \hat{S}\hat{M}\hat{D}D(i,j,\lambda) + N(m,n) \qquad \textit{Equation 21}$$

7.2 The new data processing approach

Note that the detector image I has two spectrally sheared FO images and two spectrally integrated ZO images. As it was previously discussed, the *Diffsum* approach leads to the best results regarding datacube retrieval. The approach uses the difference between the two detected FOs and their sum – the sum keeps the magnitude of the signal intensity. In contrast, the two detected ZOs are always utilized as a sum. Similarly, as the difference corresponds to measurement with {+1,-1} mask, the sum would be related to a measurement of {+1,+1} mask – it correlates with passing all the light without any encoding.

While it is important to retain the information about the absolute image intensity, using a {+1,+1} mask does not make sense for the CASSI method as the encoding by a random pattern is its core essence. Hence, we also performed datacube reconstructions in which, together with the difference of the detected images, we included the upper encoded image by {+1,0} mask. In this way, we created another approach, which we denote as *Diffone*. Using the matrix notation consistent with Table 2, the matrix representing the detector $\overline{D}$ would be written as $\overline{D} = [A - B;\ A]$, where A and B are the upper and lower FO images, respectively.

The reconstructions showed that although the idea behind encoding the image has solid grounds, the difference contains useful information and makes up for the sum. In other words, only the magnitude is needed, and the *Diffone* approach exhibited slightly worse results. In Figure 24 are plotted PSNR results of reconstructions for different regularization parameters tau (Equation 6). Each dot or circle represents one result of the reconstructed datacube under various reconstruction parameters – see Chapter 7.3. The blue circles denote the *Diffsum* results, while the red symbols are *Diffone* results. It is obvious that the highest PSNR values are achieved by the *Diffsum* approach, while it seems that *Diffone* is less dependent on the chosen reconstruction parameters. However, it is possible to identify the parameters that work best and utilize them further.

A possible explanation is that *Diffsum* compensates better for the noise in the image. The noise is pronounced both in {+1,-1} and {+1,+1} mask as opposed to *Diffone*, where the information about the inhomogeneities in the lower image encoded by {0,+1} mask is missing.

For the sake of simplicity, the data that follow are only for the *Diffsum* approach as it exhibits the best results and features the most robust datacube retrieval.

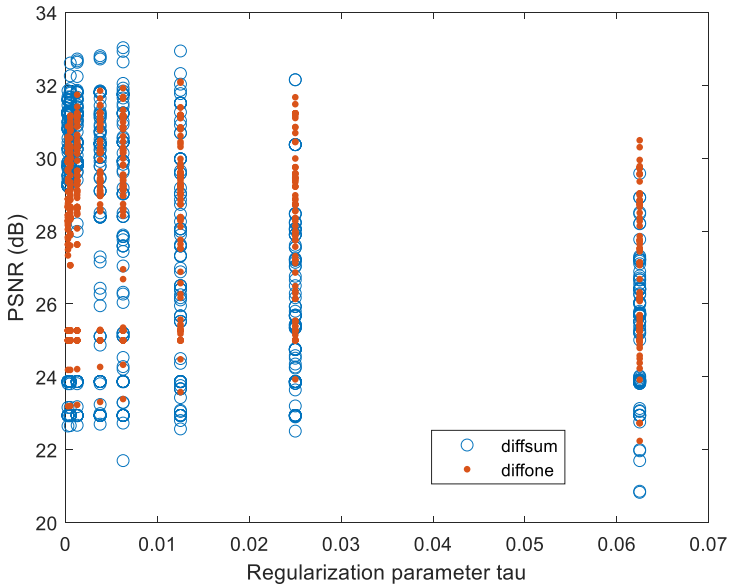

Figure 24 PSNR of reconstructed datacubes by Diffsum (blue circles) and Diffone (red dots) aprroach for different reconstruction parameters.

7.3 Reconstruction and scene parameters testing

First, a set of different parameters were tested on a datacube 256×256×50 with distinct properties to identify the major parameters to tune during the optimization of datacube retrieval. The tested parameters and scene properties included:

- Noise in the detected image with a different amount of SNR in dB (23.5, 34.8, 44.8, and without noise)
- The concentration of the chemical agent (isopropanol of 1000 ppm-m or 3333 ppm-m)
- Transforming the spectral slices with various wavelet transforms (Haar and Symlet 8)
- Option to use DCT (discrete cosine transform) of the datacube in the spectral dimension
- Different values of regularization parameter tau (put stress on the sparsity).

The reconstructions that showed the best PSNR results for noisy data were achieved using Symlet 8 wavelets together with DCT in spectral dimension. The results of noisier data were more prominent to be tau dependent. The maximal difference of PSNR between the best results achieved using different tau was 1.9 dB. The most significant reconstruction quality change was observed between SNR 23.5 dB and 34.8 dB. Hence, the next logical step was to create data with finer SNR division around those values.

7.4 Let's make some noise: the noise and spectral dimension size effect

Figure 25 shows a plotted dependency between reconstructed data PSNR and noise level present in the detected snapshot, which was defined by SNR. It is obvious that the higher the noise level (lower SNR) there is, the worse the results are. However, for the potential future experimental setup, we would like to find the lowest SNR for which it is still possible to maintain a reasonable reconstruction quality. This is possible owing to the character of the dependence in Figure 25, which does not change gradually but rather features a threshold value of PSNR, under which the reconstruction quality rapidly deteriorates. From this point of view, SNR 29.8 dB, which translates to approximately 3.9% of noise in our case, was chosen as it is the inflection point from which the fitted curve becomes almost constant.

Therefore, we fixed the level of noise at SNR = 29.8 dB and studied the effect of varying the number of spectral slices (50, 60, 70, 80, 90, 100). This simulation was created to investigate the effect of the datacube spectral dimension size. The results show – see Figure 26 that the PSNR is gradually decreasing with the number of spectral slices, but this reduction is not so prominent as for SNR. A higher number of spectral slices could provide better spectral resolution, which is necessary for distinguishing sharp spectral features. Nevertheless, it is also connected to a greater compression which in turn negatively affects the reconstruction quality. In the case of a datacube with dimension

256×256×100, the CR is 312832/6553600 ≈ 4.8%, while the same datacube with 50 spectral slices has CR 287232/3276800 ≈ 8.8%.

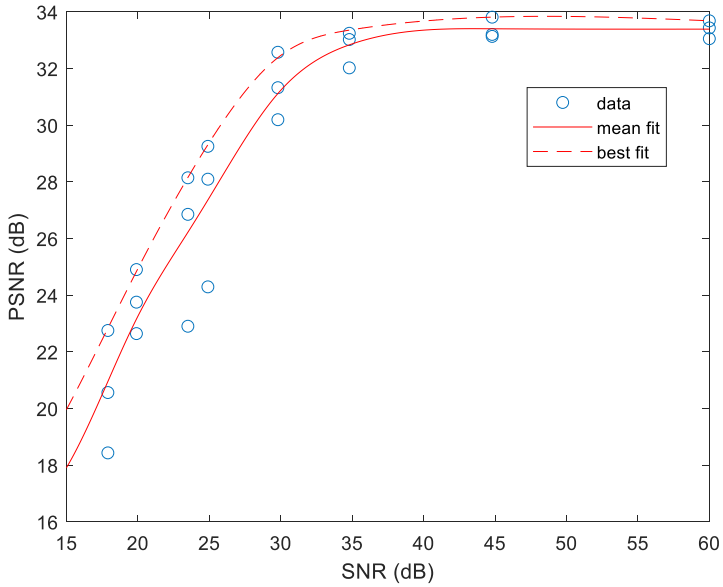

Figure 25 Dependency between SNR of the data and reconstructed PSNR. Fitted by spline. Full line – all data fit, dashed line – best results for each SNR fit. Note that SNR 60 dB represents data without noise, as it would not make sense to put these values to infinity. Hence, it is only for guiding.

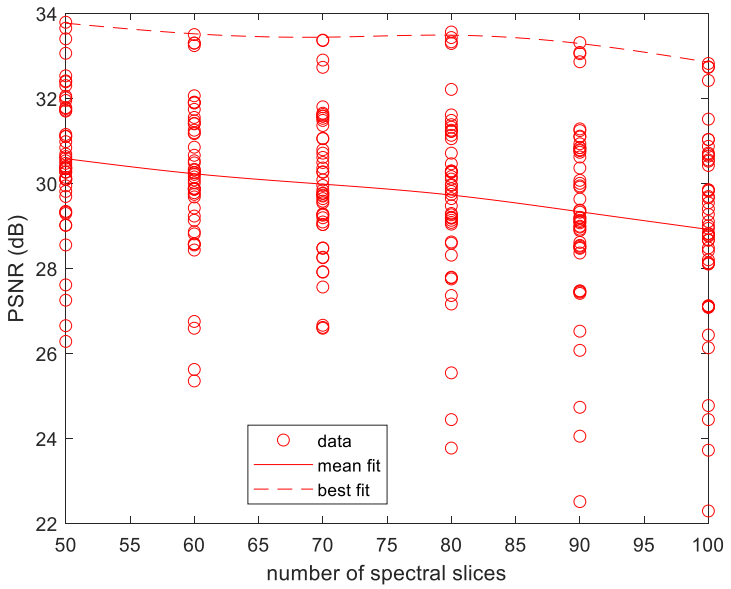

Figure 26 Dependence of the PSNR of the reconstructed datacube on number of spectral slices. The results were fitted by a spline curve. Full line – all data fitt, dashed line – best results for each number of spectral slices fit.

7.5 Promoting sparse solutions via multiple regularization weights

In the reconstruction algorithm, we can attain a sparse dataset by using discrete wavelet transform (DWT) in spatial dimensions and DCT in a spectral dimension. Hence, we work with a datacube that

is converted by both these transforms, but the sparsity is promoted by a single regularization parameter tau, which is common for both spectral and spatial dimensions. We investigated the option of putting stress on the sparsity in spectral and spatial dimensions separately, i.e., using two regularization parameters. In order to test this approach, a minor change in the reconstruction phase called soft thresholding (see Chapter 3.2.2) was made. During soft thresholding, the datacube is in both the above-mentioned bases and the wavelet/DCT coefficients that are lower or equal to tau are set to zero. This promotes sparsity.

The modification of soft thresholding consists of (i) applying inverse DCT to the datacube – now, the datacube is in a wavelet domain only. (ii) Using the first tau for soft thresholding in spatial dimensions. (iii) Performing inverse DWT and then forward DCT – now, the datacube is in the DCT domain only. (iv) Using the second tau for soft thresholding in spectral dimension. (v) Performing forward DWT – datacube is back in both transform domains.

Numerous simulations were performed using combinations of different reconstruction parameters for the modified soft thresholding – we denote it as *Double-tau* soft thresholding. Datacubes of 50-100 spectral slices were considered to provide a comparison with standard soft thresholding. In Figure 27, we can see the PSNR results of *Double-tau* soft thresholding for a different number of slices – blue circles. The mean and the best fit of the data for standard soft thresholding from Figure 26 are added for comparison as red full and dashed lines, respectively.

We did not observe a significant improvement by using the *Double-tau* thresholding despite scanning a large number of parameters for optimization. The mean fit is comparable for both

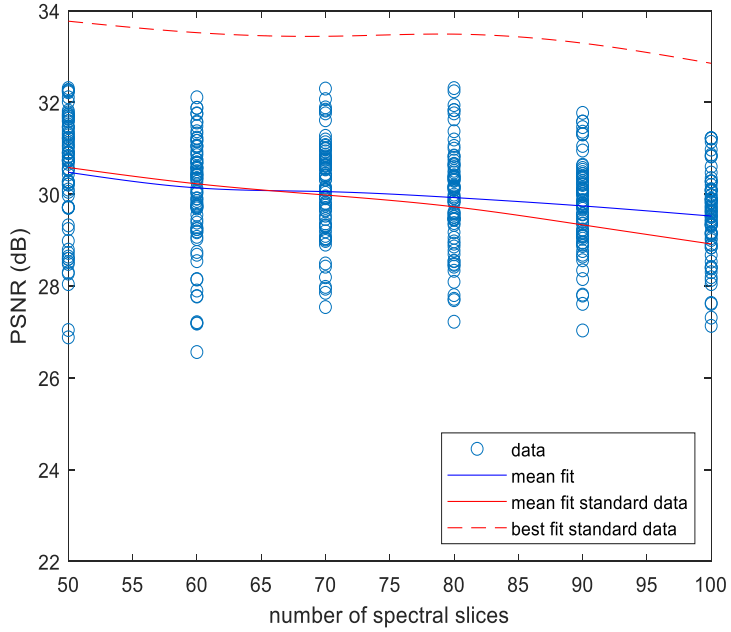

Figure 27 Dependence of the PSNR of the reconstructed datacube on number of spectral slices for Double-tau soft thresholding. Blue circles are results obtained using Double-tau soft thresholding. Blue line is mean fit of the data. Red lines are the fits using standard soft thresholding from Figure 26 – for comparison.

approaches, but the best achievable results of standard soft thresholding greatly surpass the *Double-tau* one. Therefore, we kept the basic soft thresholding in the following simulations.

7.6 Combining spectrally- and spatially-oriented reconstructions

After many trials of improving and enhancing reconstruction quality mentioned in the chapters above, we ended up with two sets of datacube reconstruction parameters – one with a focus on good spatial information (indicated as $SET_{spatial}$) and the other one with a focus on yielding correct spectra (indicated as $SET_{spectral}$). Below we present results for three scenes denoted as Scene A – Figure 28A, Scene B – Figure 28B, and Scene C – Figure 28C [67, 70]. The properties of the data in this chapter were set as follows: datacube 256×256×50 px, the concentration of isopropanol in the scene 1000 ppm-m, noise at the detector SNR = 29.8 dB.

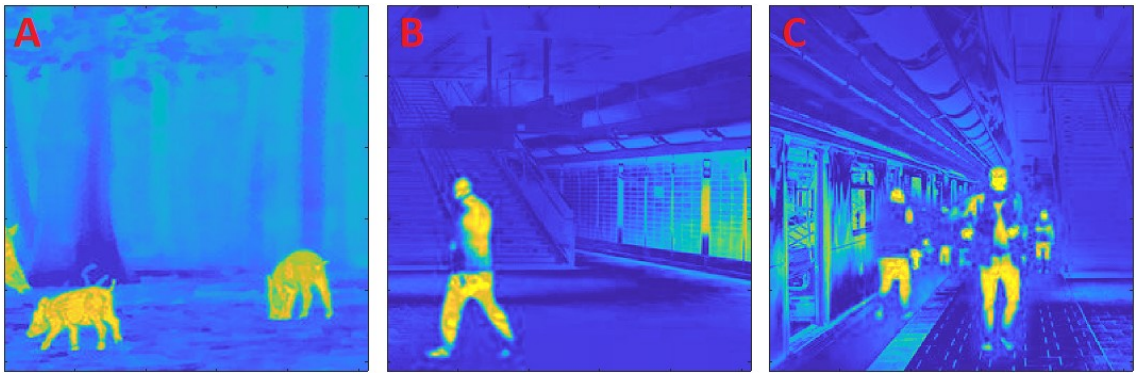

Figure 28 Original scenes used for testing in Chapter 7 denoted as (A) Scene A, (B) Scene B, and (C) Scene C.

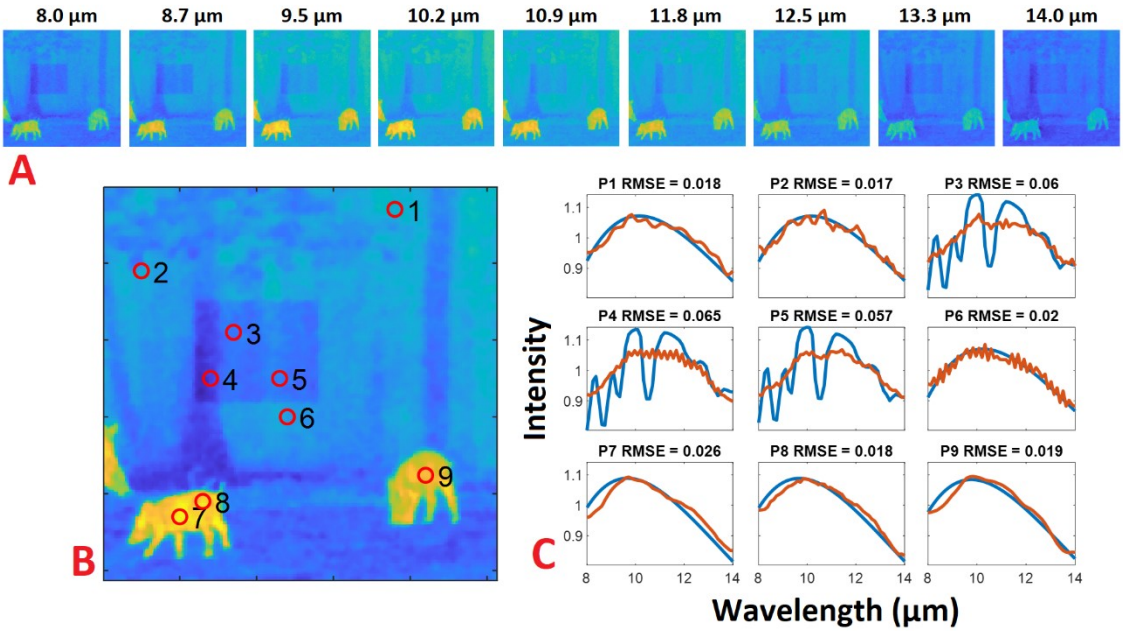

Figure 29 Reconstruction of Scene A using parameters of $SET_{spatial}$. (A) Selected reconstructed slices. (B) Integrated reconstructed slices. (C) Original (blue) and reconstructed (red) spectra from the points marked in B.

One of the important points in evaluating the simulations was the assessment of spectral quality. We found out that the overall SAM score (see Chapter 1.5) does not necessarily reflect the fidelity of the reconstructed spectra of the chemical. Hence, we also considered the mean SAM score only from the area where the chemical is present – we denote it SAM_{chem}. This metric reflects much more faithfully the quality of the retrieved spectral features in the IR region.

This discrepancy can be seen in Table 5, providing the results of Scene A reconstructions. The reconstruction $RecA_{spatial}$ was performed using $SET_{spatial}$, while $RecA_{spectral}$ was reconstructed using $SET_{spectral}$. $RecA_{spatial}$ achieved a better SAM score than $RecA_{spectral}$, but at the same time, SAM_{chem} is much worse.

The main distinction between the two abovementioned sets of parameters is the use of DCT in the spectral dimension. This way, DCT highly promotes correct spatial information but corrupts the spectra. Figure 29 and Figure 30 show the best achievable results for the two sets. In each figure, there are selected reconstructed spectral slices (panel A), an integrated image of all the reconstructed slices (panel B), and spectra from nine selected points P1-P9 of the scene (panel C). The location of the points can be seen in panel B. The evaluation metrics of these reconstructions are in Table 5, denoted as $RecA_{spatial}$ and $RecA_{spectral}$, respectively. It can be seen that Figure 29 is superior in terms of spatial correctness (higher PSNR and SSIM), but the spectra in points P3-P5 are absolutely missing spectral features of the spilled chemical (higher spectral error SAM_{chem}). In contrast, in Figure 30, we obtain more reliable spectra, but the images resemble seeing the scene with severe myopia.

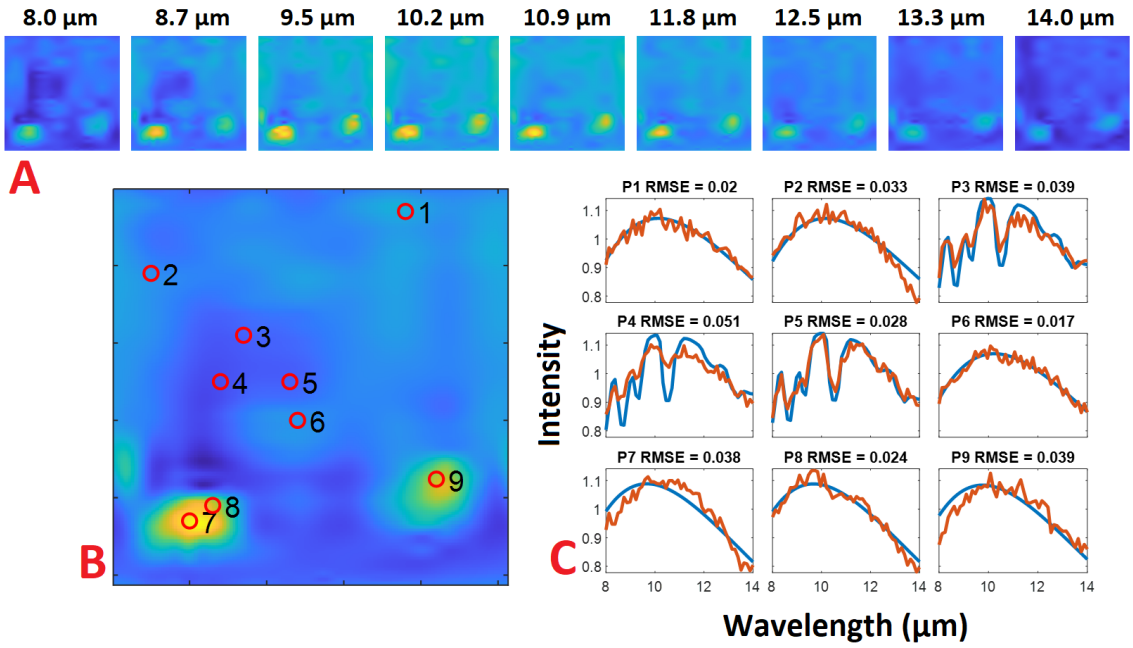

Figure 30 Reconstruction of Scene A using parameters of $SET_{spectral}$. (A) Selected reconstructed slices. (B) Integrated reconstructed slices. (C) Original (blue) and reconstructed (red) spectra from the points marked in B.

From this point of view, it is not possible to obtain accurate spatial and spectral information at the same time. Nevertheless, considering that our primary goal is to localize and identify a chemical

compound in a scene, we should prioritize the faithful spectra. Here comes into play, once again, the invaluable feature of our system that arises from the combination of CASSI extensions described in Chapters 5 and 6 – i.e., acquiring a zeroth-order image of a scene.

We know that the ZO image has correct spatial information as it is an integral combination of all encoded slices of the measured datacube. Hence, by scaling each spectrum according to the corresponding pixel intensity in the ZO, we preserve the spectra while achieving great spatial resolution. It can be written as:

$$\mathrm{DC}(i,\mathrm{j},\lambda) = DC_{Norm}(i,j,\lambda)\,.\,ZO(i,j) \qquad \textit{Equation 22}$$

where $DC_{Norm}(i,j,\lambda)$ is the reconstructed datacube normalized on its mean value, $ZO(i,j)$ is the zeroth-order image, and $\mathrm{DC}(i,\mathrm{j},\lambda)$ is the resulting datacube.

Performing $SET_{spectral}$ reconstruction and then applying Equation 22 to the reconstructed datacube, we obtain results in Figure 31, which clearly outperform the reconstructions $\mathrm{RecA}_{\mathrm{spatial}}$ and $\mathrm{RecA}_{\mathrm{spectral}}$ – see $\mathrm{RecA}_{\mathrm{comb}}$ in Table 5 for the evaluation metrics.

For a better idea about the spectral error distribution in the reconstructed scene, we present spectral error maps for $\mathrm{RecA}_{\mathrm{spatial}}$ and $\mathrm{RecA}_{\mathrm{spectral}}$ in Figure 32, where each pixel is the calculated spectral error sam from Equation 15. $\mathrm{RecA}_{\mathrm{spatial}}$ in Figure 32A has much higher error values in the area of chemical absorption compared to $\mathrm{RecA}_{\mathrm{spectral}}$ in Figure 32B, which is in accordance with observations of spectra P3-P5 in Figure 29C and Figure 30C. This should serve as a demonstration that points P1-P9 were selected quasi-randomly beforehand. It is worth noting that the spectral error

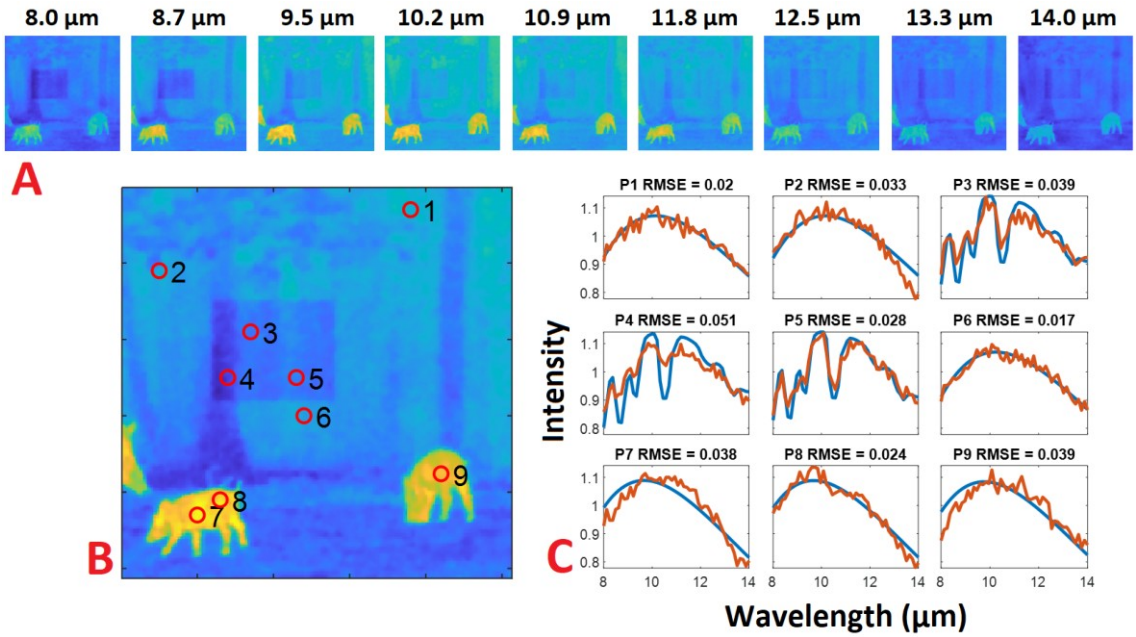

Figure 31 The best achievable results for Scene A using a combination of $SET_{spectral}$ and post-reconstruction utilization of zeroth-order image according to Equation 22. The reconstruction is denoted as $RecA_{comb}$. (A) Selected reconstructed slices. (B) Integrated reconstructed slices. (C) Original (blue) and reconstructed (red) spectra from the points marked in B.

map for reconstruction $RecA_{comb}$ would look the same as for $RecA_{spectral}$ since it features the same spectra.

Figure 32 also visualizes the need for characterization of the spectral fidelity by SAM_{chem} metric as everywhere except the chemical absorption area is a low spectral error for $RecA_{spatial}$, which yields lower overall SAM compared to $RecA_{spectral}$.

Table 5 The best achievable results in terms of spatial information ($RecA_{spatial}$ using $SET_{spatial}$), spectral information ($RecA_{spectral}$ using $SET_{spectral}$), and a combination of $SET_{spectral}$ and post-reconstruction utilization of ZO image according to Equation 22 ($RecA_{comb}$) for Scene A

	PSNR	SSIM	SAM	SAM_{chem}
$RecA_{spatial}$	34.84	0.90	1.19	3.55
$RecA_{spectral}$	29.86	0.84	1.35	2.24
$RecA_{comb}$	34.63	0.90	1.35	2.24

In order to evaluate how the reconstruction quality would change in dependence on different conditions, we tested the reconstruction parameters on different scenes – see Figure 33, Table 6 for results of a scene with a more complicated area of chemical absorption (Scene B, Figure 28B), and Figure 34, Table 7 for results of a scene with a more complicated area of chemical absorption and more complex spatial features (Scene C, Figure 28C).

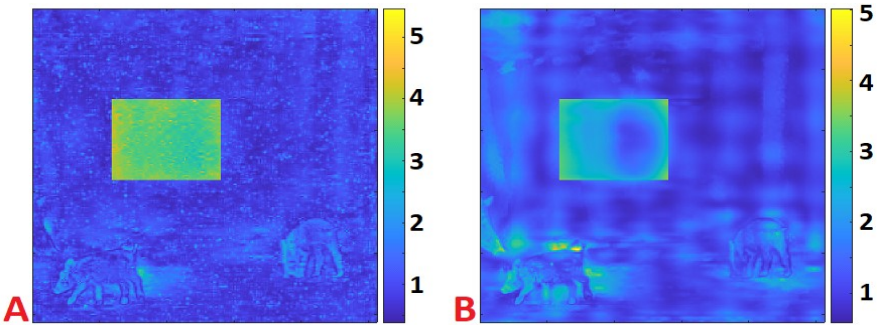

Figure 32 Spectral error map of reconstruction for (A) $RecA_{spatial}$ from Figure 29 and (B) $RecA_{comb}$ from Figure 31.

Points with higher intensity in the original scene lead to worse quality of reconstructed spectra in these points – see points P7 and P8 compared to points P1 and P2 in Figure 31, Figure 33, and Figure 34. The fine details in the reconstructed slices deteriorated, which is caused mainly due to the noise. Nevertheless, the degree of retrieved details is more than satisfactory in terms of imaging a scene and locating a chemical substance. Joint reconstructions yielded the same SSIM as reconstructions using $SET_{spatial}$, as well as the same SAM and SAM_{chem} scores as $SET_{spectral}$. PSNR slightly decreased for Scene A and improved for Scene B and Scene C.

Table 6 The best achievable results in terms of spatial information ($RecB_{spatial}$ using $SET_{spatial}$), spectral information ($RecB_{spectral}$ using $SET_{spectral}$), and a combination of $SET_{spectral}$ and post-reconstruction utilization of ZO image according to Equation 22 ($RecB_{comb}$) for Scene B

	PSNR	SSIM	SAM	SAM_{chem}
$RecB_{spatial}$	33.87	0.90	1.43	3.55
$RecB_{spectral}$	29.03	0.84	1.41	2.22
$RecB_{comb}$	34.41	0.90	1.41	2.22

Table 7 The best achievable results in terms of spatial information ($RecC_{spatial}$ using $SET_{spatial}$), spectral information ($RecC_{spectral}$ using $SET_{spectral}$), and a combination of $SET_{spectral}$ and post-reconstruction utilization of ZO image according to Equation 22 ($RecC_{comb}$) for Scene C

	PSNR	SSIM	SAM	SAM_{chem}
$RecC_{spatial}$	32.24	0.85	1.44	3.65
$RecC_{spectral}$	27.35	0.72	1.56	2.28
$RecC_{comb}$	32.56	0.85	1.56	2.28

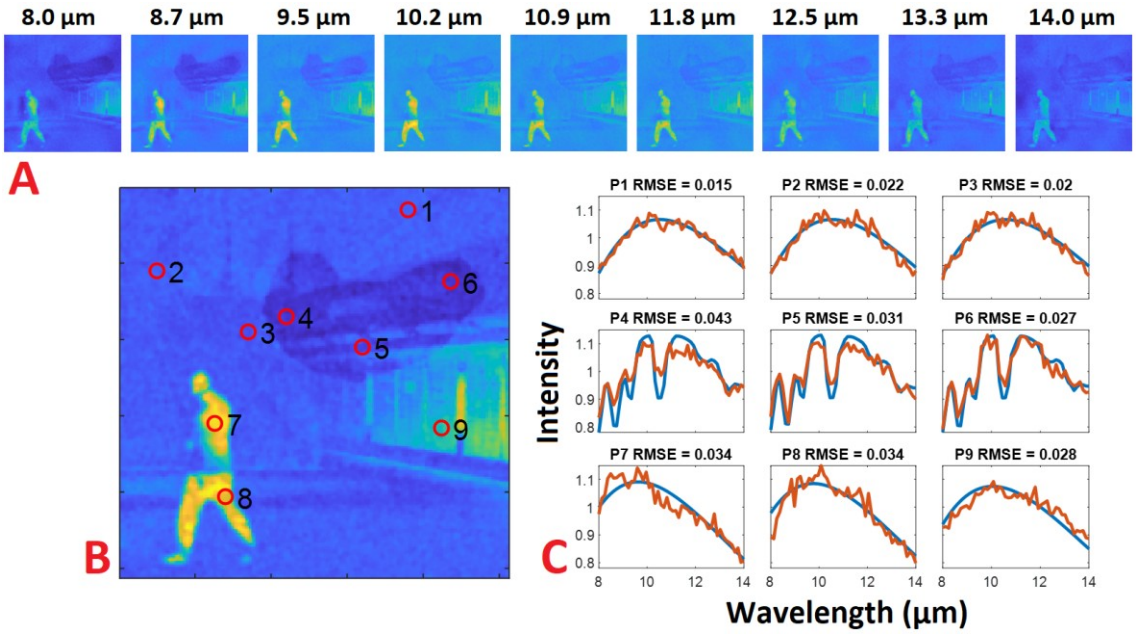

Figure 33 The best achievable results for scene B using a combination of $SET_{spectral}$ and post-reconstruction utilization of zeroth-order image according to Equation 22. The reconstruction is denoted as $RecB_{comb}$. (A) Selected reconstructed slices. (B) Integrated reconstructed slices. (C) Original (blue) and reconstructed (red) spectra from the points marked in B.

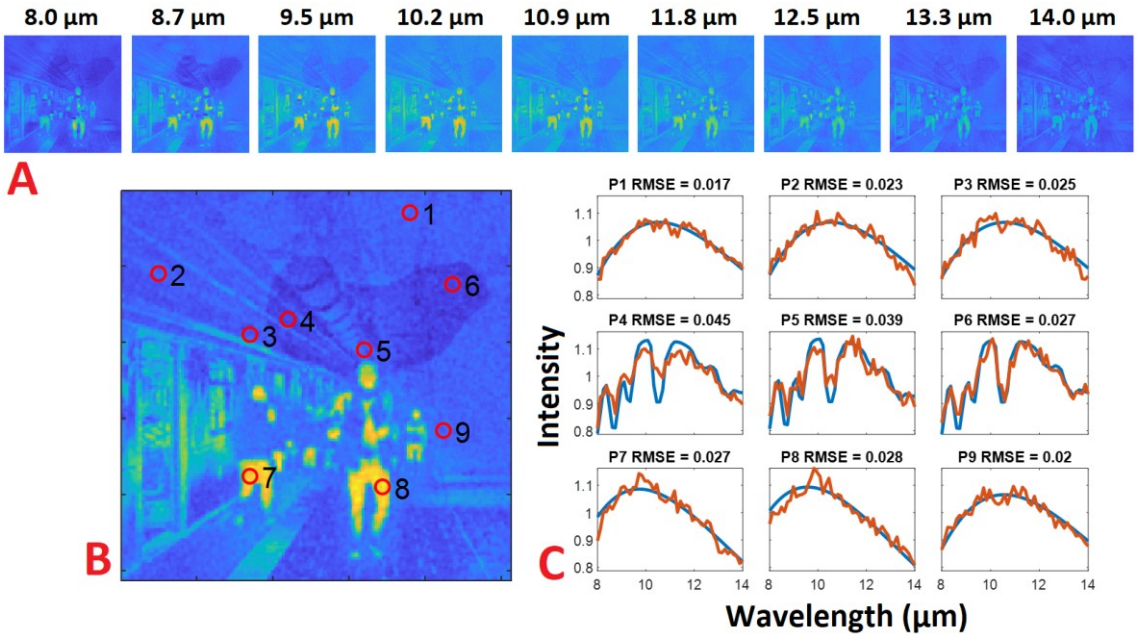

Figure 34 The best achievable results for scene C using a combination of $SET_{spectral}$ and post-reconstruction utilization of zeroth-order image according to Equation 22. The reconstruction is denoted as $RecC_{comb}$. (A) Selected reconstructed slices. (B) Integrated reconstructed slices. (C) Original (blue) and reconstructed (red) spectra from the points marked in B.

7.7 Robustness against noise for the combined retrieval

A good indication of the performance of the system would be to assess the highest amount of noise at which it can still achieve reliable reconstruction. For this purpose, reconstructions of data with varying noise were performed. The regularization parameter tau was set to 0.1 as it consistently provided the best results for all the scenes. Moreover, in a real scenario, tuning this parameter without prior knowledge of the scene would not be possible. The results are summarized in Table 8.

Table 8 Results for Scene C achieved by using a combination of $SET_{spectral}$ and post-reconstruction utilization of ZO image according to Equation 22 with fixed regularization parameter tau for different amounts of noise

SNR (dB)	PSNR (dB)	SSIM	SAM (°)	SAM_{chem} (°)
34.8	32.65	0.87	1.44	2.26
29.8	32.34	0.86	1.55	2.33
25	30.89	0.79	1.95	2.58
20	28.23	0.67	2.69	3.34
18	26.30	0.57	3.36	3.82
15	23.38	0.43	4.45	5.23

Figure 35 shows that the dependency of reconstruction quality on noise is not linear for both PSNR and SAM_{chem}. From this point of view, an inflection point from which the quality deteriorates rather quickly is near 25 dB of SNR. Hence, the reconstructed data with the noise of 25 dB and 20 dB

are presented in Figure 36A and Figure 36B, respectively. These values translate to 6.4% and 11.3%, respectively. It is still possible to distinguish the spatial information in Figure 36B, even though it is visibly noisier. However, evaluating the spectral information would be a problem as the spectra are severely disrupted. SAM and $SAM_{chem,}$ in this case, surged from 1.95 and 2.58 to 2.69 and 3.34, which is approximately 38% and 30% increase, respectively. We can draw a conclusion that in order to obtain a reliable reconstruction, the maximal SNR in the detected image needs to reach 25 dB.

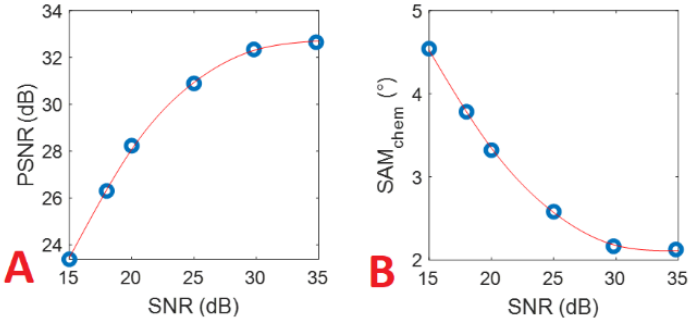

Figure 35 Dependency between SNR of the data and reconstructed (A) PSNR and (B) SAM_{chem}. Plotted data from Table 8, fitted by spline.

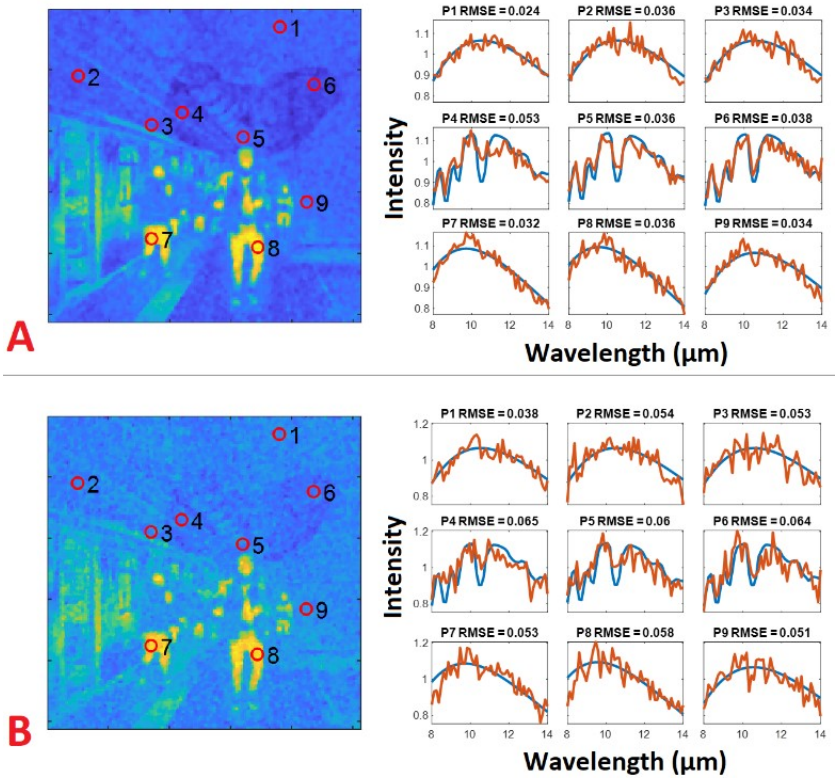

Figure 36 Results for Scene C achieved by using a combination of $SET_{spectral}$ and post-reconstruction utilization of ZO image according to Equation 22 with fixed regularization parameter `tau` *for noise of (A) 25 dB and (B) 20 dB.*

8 Conclusions

Throughout this thesis, we have seen how the compressed sensing (CS) technique CASSI relying on a single snapshot, is a unique aspirant in the field of hyperspectral imaging (HSI). It has enormous potential for obtaining chemical information remotely. Yet, the method possesses a lot of room for improvement. There are two main reasons for this. First, it dismantles the primary disadvantages of HSI, which are halting more widespread use of HS cameras – these are cost and complexity. As was shown in this thesis, CASSI can be constructed as a relatively simple optical setup. Second, CASSI allows capturing HS information at a rate not comparable to any other mean. The increasing research interest in snapshot HSI in recent years is a hint at the potential of these devices, but because of manufacturing limitations, it has not seen wider adoption in commercial use. One exception is the professional astronomical community, which highly benefits from the vast light throughput of a telescope when conducting HSI [71].

The main contribution of this work is that it determined the limiting factors of CASSI – namely, very high compression of the measured data, which is even more pronounced for datacubes with many spectral slices. We proposed and realized extensions needed to overcome the obstacles and then provided a conceptual study of CASSI operating in the LWIR spectral region, which can perform detection and localization of a chemical substance in noisy conditions.

We developed a differential CASSI (D-CASSI) using two complementary binary random masks and, thus, two imaging paths to multiply the measured information and consequently lift the limits of the compression. On top of that, this system combines a diffraction grating and a prism as a dispersive element allowing for concentric mounting. It is designed so that it can capture both the first- and zeroth-order diffraction of the grating on the same detector. Utilizing the zeroth order improves the spatial quality of the reconstructed data dramatically. It is worth noting that thanks to masks' complementarity, we were able to develop a completely new approach to data processing and reconstruction, which utilizes a random mask consisting of {-1,+1} pixels. The synergy of the CASSI extensions brought into existence gives rise to post-reconstruction processing that has a huge positive impact on reconstruction fidelity.

It can be concluded that the modified CASSI system makes it, indeed, possible to perform HSI on a broad spectral range in the IR spectral region in order to localize a chemical substance if the resulting SNR on the detector is at least 25 dB. Note that the made extensions retained the simplicity of the optical system and the main advantage of CASSI, which is a single-snapshot operation regime.

Comment on the papers

The publications listed below are a part of this thesis. They constitute a substantial portion of Chapters 4-6 and contain information about the experimental design of the presented optical system, altogether with additional results. An interested reader is referred to the following pages, where they are enclosed.

I. HLUBUČEK, J., ŽÍDEK, K.: Evaluation of using coded aperture imaging in the mid- and far-infrared region. In: *5th International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing, and Imaging (CoSeRa)*, Eurasip, 2018. 19.
II. HLUBUČEK, J., et al. Improving Compression Ratio in CASSI. In: *Computational Optical Sensing and Imaging*. Optical Society of America, 2021. CTh2F.3.
III. HLUBUČEK, J., et al. Enhancement of CASSI by a zero-order image employing a single detector. *Applied Optics*. 2021, **60**(5), 1463-1469.
IV. HLUBUČEK, J., et al. Differential coded aperture single-snapshot spectral imaging. *Optics Letters*. 2022, **47**(9), 2342-2345.

The work in Chapter 7 is yet to be published.

My contribution to the papers

In papers **I-IV,** I performed the predominant part of the experimental work and all the data analysis. I wrote papers **I-IV** with supervision by Karel Žídek.

On behalf of the co-authors, the above-mentioned declaration was confirmed by

Doc. RNDr. Karel Žídek, Ph.D.

References

[1] MANLEY, M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews*. 2014, **43**(24), 8200–8214.

[2] TÜRKER-KAYA, S., HUCK, Ch. W. A Review of Mid-Infrared and Near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. *Molecules*. 2017, **22**(1), 168.

[3] GABRIELI, A., WRIGHT, R., LUCEY, P.G., PORTER, J.N., GARBEIL, H., PILGER E. a WOOD, M. Characterization and initial field test of an 8–14 μm thermal infrared hyperspectral imager for measuring SO2 in volcanic plumes. *Bulletin of Volcanology*. 2016, **78**(10).

[4] ROSI, F., MILIANI, C., BRAUN, R., HARIG, R., SALI, D., BRUNETTI, B.G. a SGAMELLOTTI, A. Noninvasive Analysis of Paintings by Mid-infrared Hyperspectral Imaging. *Angewandte Chemie International Edition*. 2013, **52**(20), 5258-5261.

[5] CALVINI, R., ULRICI, A. a AMIGO, J.M. Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*. 2015, **146**, 503-511.

[6] DONOHO, D.L. Compressed sensing. *IEEE Transactions on Information Theory*. 2006, **52**(4), 1289-1306.

[7] CANDES, E.J. a WAKIN, M.B. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*. 2008, **25**(2), 21-30.

[8] BRADY, D.J. *Optical Imaging and Spectroscopy*. Hoboken, NJ, USA: John Wiley & Sons, 2009.

[9] WAGADARIKAR, A., JOHN, R., WILLETT, R. a BRADY, D.J. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*. 2008, 47(10).

[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*. 2014, **31**(1), 105-115.

[11] WANG, L., XIONG, Z., GAO, D., SHI, G. a WU, F. Dual-camera design for coded aperture snapshot spectral imaging. *Applied Optics*. 2015, **54**(4).

[12] GENDRIN, C., ROGGO, Y., COLLET, C. Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review. *Journal of Pharmaceutical and Biomedical Analysis*. 2008, **48**(3), 533–553.

[13] STUART, B. H., *Infrared Spectroscopy: Fundamentals and Applications*. Chichester, UK: John Wiley & Sons, 2004. Analytical Techniques in the Sciences.

[14] HUCK, Ch. W., OZAKI, Y., VERENA, A., HUCK-PEZZEI, V. A. Critical Review Upon the Role and Potential of Fluorescence and Near-Infrared Imaging and Absorption Spectroscopy in Cancer Related Cells, Serum, Saliva, Urine and Tissue Analysis. *Current Medicinal Chemistry*. 2016, **23**(27), 3052–3077.

[15] GOWEN, A. A., et al. Recent applications of chemical imaging to pharmaceutical process monitoring and quality control. *European journal of pharmaceutics and biopharmaceutics*. 2008, **69**(1), 10–22.

[16] KATRAŠNIK, J., BÜRMEN, M., PERNUŠ, F., LIKAR, B. Spectral characterization and calibration of AOTF spectrometers and hyper-spectral imaging systems. *Chemometrics and Intelligent Laboratory Systems*. 2010, **101**(1), 23-29.

[17] AMIGO, J. M. Practical issues of hyperspectral imaging analysis of solid dosage forms. Analytical and Bioanalytical Chemistry. 2010, **398**(1), 93-109.

[18] HAMMOND, S. V., CLARKE, F. C. *The handbook of vibrational spectroscopy, sampling techniques, microscopy*. 2002.

[19] MANOLAKIS, D., et al. Longwave infrared hyperspectral imaging: Principles, progress, and challenges. *IEEE Geoscience and Remote Sensing Magazine*, 2019, **7**(2), 72-100.

[20] SINCLAIR, M. B., et al. Hyperspectral confocal microscope. *Applied optics*. 2006, **45**(24), 6283-6291.

[21] SINCLAIR, M. B., et al. Design, construction, characterization, and application of a hyperspectral microarray scanner. *Applied optics*. 2004, **43**(10), 2079-2088.

[22] KASILI, P. M., VO- DINH, T. Hyperspectral imaging system using acousto- optic tunable filter for flow cytometry applications. *Cytometry Part A: The Journal of the International Society for Analytical Cytology*. 2006, **69**(8), 835-841.

[23] TAYLOR, H. L., BANKS, S. C., MCCOY, J. F. Deconvolution with the ℓ 1 norm. *Geophysics*. 1979, **44**(1), 39-52.
[24] HAYES, B. Computing Science: The Best Bits. *American scientist*. 2009, **97**(4), 276-280. https://www.jstor.org/stable/27859352.
[25] WANG, P., LIANG, J. a WANG, L. V. Single-shot ultrafast imaging attaining 70 trillion frames per second. *Nature Communications*. 2020, **11**(1).
[26] GAMEZ, G. Compressed sensing in spectroscopy for chemical analysis. *Journal of Analytical Atomic Spectrometry*. 2016, **31**(11), 2165-2174.
[27] ŽÍDEK, K., DENK, O., HLUBUČEK, J. a VÁCLAVÍK, J. Compact and robust hyperspectral camera based on compressed sensing. In: *Optics and Measurement International Conference 2016*. SPIE, 2016. 165-171.
[28] ISO Update. 2022. *What is Quality - ISO Update*. [online] Available at: https://isoupdate.com/resources/what-is-quality [Accessed 12 October 2022].
[29] MOORTHY, A. K., WANG, Z., BOVIK, A. C. Visual perception and quality assessment. *Optical and Digital Image Processing: Fundamentals and Applications*. Weinheim: Wiley, 2011, 419-439.
[30] AVIRIS - Airborne Visible / Infrared Imaging Spectrometer. 2022. *AVIRIS Data Portal - JPL - NASA*. [online] Available at: https://aviris.jpl.nasa.gov/dataportal [Accessed 12 October 2022].
[31] CHANG, N. B., VANNAH, B., YANG, Y.J. Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in Lake Erie. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2014, **7**(6), 2426-2442.
[32] GAO, L., LIANG, J., LI, Ch. a WANG, L. V. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*. 2014, **516**(7529), 74-77.
[33] JACQUINOT, P. The luminosity of spectrometers with prisms, gratings, or Fabry-Perot etalons. *Journal of the Optical Society of America*. 1954, **44**(10), 761-765.
[34] HAGEN, N. A., et al. Snapshot advantage: a review of the light collection improvement for parallel high-dimensional measurement systems. *Optical Engineering*. 2012, **51**(11), 111702.
[35] KITTLE, D., et al. Multiframe image estimation for coded aperture snapshot spectral imagers. *Applied optics*. 2010, 49.36: 6824-6833.
[36] WU, Y., et al. Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Optics letters*. 2011, **36(**14), 2692-2694.
[37] ARGUELLO, H, ARCE, G. R. Code aperture optimization for spectrally agile compressive imaging. *Journal of the Optical Society of America A*. 2011, **28(**11), 2400-2413.
[38] ARGUELLO, H., ARCE, G. R. Rank minimization code aperture design for spectrally selective compressive imaging. *IEEE transactions on image processing*. 2012, **22(**3), 941-954.
[39] ARGUELLO, H., ARCE, G. R. Code aperture design for compressive spectral imaging. In: *2010 18th European Signal Processing Conference*. IEEE, 2010. 1434-1438.
[40] ARGUELLO, H., et al. Higher-order computational model for coded aperture spectral imaging. *Applied optics*. 2013, **52(**10), D12-D21.
[41] YUAN, X., TSAI, T. H., ZHU, R., LLULL, P., BRADY, D. J., CARIN, L. Compressive Hyperspectral Imaging With Side Information. *IEEE Journal of Selected Topics in Signal Processing*. 2015, **9**(6), 964-976.
[42] WANG, L., XIONG, Z., SHI, G., ZENG, W. a WU, F. Compressive hyperspectral imaging with complementary RGB measurements. *Visual Communications and Image Processing (VCIP)*. IEEE, 2016. 1-4.
[43] RUEDA, H., ARGUELLO, H., ARCE, G. R. Dual-ARM VIS/NIR compressive spectral imager. In: *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015. 2572-2576.
[44] WANG, L., et al. High-speed hyperspectral video acquisition with a dual-camera architecture. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015. 4942-4950.
[45] WAGADARIKAR, A. A., et al. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics express*. 2009, **17(**8), 6368-6388.
[46] FENG, Y. Z., SUN, D. W., Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review. *Critical Reviews in Food Science and Nutrition*. 2012, **52**(11), 1039-1058.
[47] ROXBURY, D., et al. Hyperspectral microscopy of near-infrared fluorescence enables 17-chirality carbon nanotube imaging. *Scientific reports*. 2015, **5**(1), 1-6.
[48] CARRASCO, O., et al. Hyperspectral imaging applied to medical diagnoses and food safety. In: *Geo-Spatial and Temporal Image and Data Exploitation III*. SPIE, 2003. 215-221.

[49] CALVINI, R., ULRICI, A., AMIGO, J. M. Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*, 2015. **146**, 503-511.
[50] FENG, Y. Z., SUN, D. W. Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, 2012. **52**(11) 1039-1058.
[51] EDELMAN, G.J., GASTON, E., VAN LEEUWEN, T.G., CULLEN, P.J. a AALDERS, M.C.G. Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic Science International*. 2012, **223**(1-3), 28-39.
[52] SCHULER, R. L., KISH, P. E., PLESE, C. A. Preliminary observations on the ability of hyperspectral imaging to provide detection and visualization of bloodstain patterns on black fabrics. *Journal of forensic sciences*, 2012. **57(**6), 1562-1569.
[53] PEREIRA, J. F. Q., et al. Detection and identification of Cannabis sativa L. using near infrared hyperspectral imaging and machine learning methods. A feasibility study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020. **237,** 118385.
[54] BALAS, C., et al. Hyperspectral imaging and spectral classification for pigment identification and mapping in paintings by El Greco and his workshop. *Multimedia Tools and Applications*, 2018. **77**(8), 9737-9751.
[55] MAHESH, S., JAYAS, D.S., PALIWAL, J. a WHITE, N.D.G. Hyperspectral imaging to classify and monitor quality of agricultural materials. *Journal of Stored Products Research*. 2015, **61**, 17-26.
[56] YANG, S., et al. Mid-infrared compressive hyperspectral imaging. *Remote Sensing*, 2021. **13(**4), 741.
[57] RUSSELL, T. A., et al. Compressive hyperspectral sensor for LWIR gas detection. In: *Compressive Sensing*. SPIE, 2012. 55-67.
[58] DUPUIS, J. R., KIRBY, M., COSOFRET, B. R. Longwave infrared compressive hyperspectral imager. In: *Next-Generation Spectroscopic Technologies VIII*. SPIE, 2015. 232-242.
[59] CRAIG, Stephanie M., et al. Compressive sensing hyperspectral imager in the LWIR for chemical plume detection. In: *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXVIII*. SPIE, 2022. 186-199.
[60] BIOUCAS-DIAS, J. M., FIGUEIREDO, M. A. T. A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration. *IEEE Transactions on Image Processing*. 2007, **16**(12), 2992-3004.
[61] HLUBUČEK, J., et al. Differential coded aperture single-snapshot spectral imaging. *Optics Letters*. 2022, **47**(9), 2342-2345.
[62] HLUBUČEK, J., et al. Enhancement of CASSI by a zero-order image employing a single detector. *Applied Optics*. 2021, **60(**5), 1463-1469.
[63] HLUBUČEK, J., ŽÍDEK, K.: Evaluation of using coded aperture imaging in the mid- and far-infrared region. In: *5th International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing, and Imaging (CoSeRa)*, Eurasip, 2018. 19.
[64] SOLDEVILA, F., et al. Use of balanced detection in single-pixel imaging. In: *Computational Optical Sensing and Imaging*. Optica Publishing Group, 2016. CW5D.5.
[65] HLUBUČEK, J., LUKEŠ, J., ŽÍDEK, K. Improving Compression Ratio in CASSI. In: *Computational Optical Sensing and Imaging*. Optical Society of America, 2021. CTh2F.3.
[66] CANDÈS, E. J., ROMBERG, J., TAO, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 2006. **52(**2), 489-509.
[67] Euregiohunt. 2022. *Calonox View*. [online] Available at: https://cdn.webshopapp.com/shops/215633/files/360391341/leica-calonox-view.jpg [Accessed 12 October 2022].
[68] National Institute of Standards and Technology. 2022. *NIST Chemistry WebBook*. [online] Available at: https://webbook.nist.gov/chemistry/ [Accessed 12 October 2022].
[69] REES, W. G. *Physical principles of remote sensing*. Cambridge university press, 2013.
[70] Leica geosystems. 2022. *LiDAR Based Security Solutions - Leica BLK247*. [online] Available at: https://shop.leica-geosystems.com/sites/default/files/styles/gallery_medium/public/2020-11/BLK247_Subway_IR_view.jpg [Accessed 12 October 2022].
[71] HAGEN, N. A., et al. Snapshot advantage: a review of the light collection improvement for parallel high-dimensional measurement systems. *Optical Engineering*, 2012. **51(**11), 111702.

Optics Letters

Differential coded aperture single-snapshot spectral imaging

J. Hlubuček,[1,2,*] J. Lukeš,[1,2] J. Václavík,[1,2] and K. Žídek[1]

[1]Research Centre for Special Optics and Optoelectronic Systems (TOPTEC), Institute of Plasma Physics of the CAS, Prague, Czech Republic
[2]Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Liberec, Czech Republic
*Corresponding author: hlubucek@ipp.cas.cz

Received 3 February 2022; revised 19 March 2022; accepted 6 April 2022; posted 6 April 2022; published 28 April 2022

We propose a novel, to the best of our knowledge, concept of the differential coded aperture snapshot spectral imaging (D-CASSI) technique exploiting the benefits of using a {–1,+1} random mask, which is demonstrated by a broadband single-snapshot hyperspectral camera using compressed sensing. To double the information, we encode the image with two complementary random masks, which proved to be superior to two independent patterns. We utilize dispersed and non-dispersed encoded images captured in parallel on a single detector. We explored several different approaches to processing the measured data, which demonstrates significant improvement in retrieving complex hyperspectral scenes. The experiments were completed by simulations in order to quantify the reconstruction fidelity. The concept of differential CASSI could be easily implemented also by multi-snapshot CASSI without any need for optical system modification.

https://doi.org/10.1364/OL.454729

Hyperspectral imaging (HSI) denotes imaging where a spectrum is recorded for each pixel of the image. It is a very useful technique for a broad range of samples—for instance, in the infrared (IR) spectral region, light makes it possible to remotely sense the chemical composition owing to the specific absorption fingerprints of each chemical compound. Since the acquired dataset is a 3D datacube consisting of many stacked 2D images, HSI inevitably collects a large amount of data. Processing the datacube is very demanding for computation power, acquisition times are usually very lengthy, and the HSI requires a high intensity of light. Moreover, in the IR region, there is a need for special optical materials and IR array detectors.

A possible solution to this problem is using a compressed sensing method called coded aperture snapshot spectral imaging (CASSI), which makes it possible to compress a 3D hyperspectral scene in a single instant on a 2D detector and then retrieve the 3D information back thanks to a reconstruction algorithm, such as TwIST [1]. The core of the method lies in encoding a measured scene with a binary random mask pattern, which is then spectrally sheared and captured on a detector. However, since the basic CASSI method relies on a single snapshot, the data compression ratio is immense, making the reconstruction of complex datacubes very problematic. Therefore, an extension of this method is required.

In recent years there have been efforts to decrease the compression ratio by numerous means [2–9]. Promising enhancements are (i) acquiring multiple snapshots of the same scene using different random mask patterns [2–4] or (ii) capturing a non-diffracted image of a scene and using this knowledge in the reconstruction [5–9]. However, for (i) the CASSI method requires some advanced modulators due to the need for changing the random mask pattern, and for (ii) a second camera is often needed.

A way to avoid using a second detector could be utilizing a grating in the CASSI system and consequently capturing a zero-order (ZO) of diffraction, i.e., a non-diffracted image of the scene, next to the first-order—a spectrally sheared image of the scene—on the same detector [10]. Nevertheless, this approach itself does not provide sufficient reconstruction fidelity for real-life HSI in a broad spectral range.

In this Letter, we demonstrate a novel approach to obtaining more information about the measured scene in the CASSI technique. The scene is imaged by a double-lens and subsequently encoded via two random binary masks. A diffraction grating provides us with both the first-order diffraction image (standard CASSI information) and the ZO diffraction image, i.e., the spectrally integrated image. We show that the doubled information is a promising way to improve the reconstruction quality without making the optical setup more complex. Moreover, by smart design of the random mask patterns, we are able to improve the quality even further.

The imaging of a scene is done with two lenses cut into a rectangular shape (size 10 mm × 50 mm, f = 100 mm), which were glued together along their long side. The double-lens combined with a field lens projects the measured scene into two identical images encoded by two different random masks (64 px × 64 px). The encoded images propagate through the system depicted in Fig. 1(a) and are captured on the detector above each other. Here, M denotes random mask, L refers to the lenses, P and G are the prism and the grating, respectively, and D represents the doublets. A detector with resolution 2056 px × 2464 px was used. However, a detector with approximately 145 px × 275 px would be sufficient regarding the mask size and resolution. It is worth noting that our system is also able to capture a non-diffracted image of the scene (ZO) on the same detector, which provides us with more information about the measured scene without the need to split an incoming light as in the standard extension of the CASSI method [5–9].

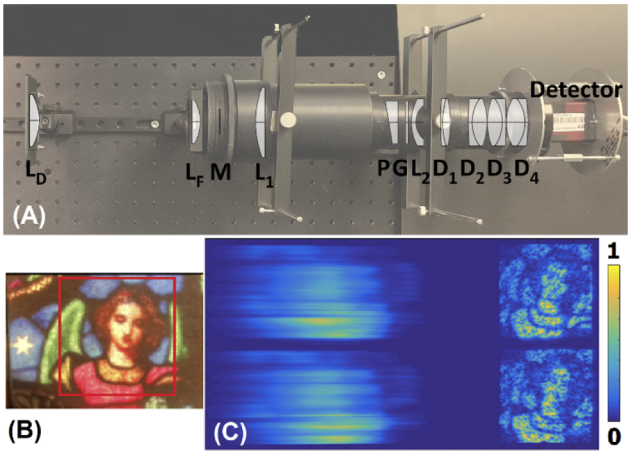

Fig. 1. (a) Scheme of the optical setup. M = mask, L = lens, P = prism, G = grating, D = doublet (see Supplement 1 for more details). (b) Photo of the measured scene. The square marks the imaged area. (c) Double mask CASSI detector image; first-order of diffraction on the left, zero-order of diffraction on the right. Note that the same color bar applies to all figures.

The two random masks (patterns) encoding the scene might be, in principle, entirely independent and random. However, our simulations prove that it is beneficial to use two complementary masks—see Supplement 1 for detailed information. The complementary masks approach was used previously for a color-coded mask [11]. For a single mask in a standard CASSI system, 50% of the information is lost at the pixels, where the random mask binary information is 0. Therefore, using two random masks that are complementary to each other, i.e., the positions of ones in the first mask are zeros in the second mask and vice versa, we are guaranteed to acquire the information from all the pixels of the scene. It decreases the compression ratio but without the need for the second detector as in Wang *et al.* [5,6,8], Rueda *et al.* [7], and Yuan *et al.* [9].

In the optical setup, we put stress on a simple construction of the CASSI camera using a minimal number of optical elements. The resulting device is relatively compact and uses a concentric mounting. This approach is retained due to the vision of using an analogous device in the IR regime, where the optical element fabrication and alignment are significantly more challenging than in the VIS region. As a result, our CASSI setup suffers from optical aberrations, which need to be overcome.

Owing to the fact that we have two complementary spectrally sheared images, we can approach the doubled information in several ways. The basic one would be to simply consider the two images separately, as is done in multi-frame CASSI extensions [2,3]. We denote this approach as *Double*. Another approach, abbreviated as *Diff*, is to calculate a difference of the detected images, which simulates a measurement with a random mask pattern consisting of ±1, i.e., a differential image of the two masks. Note that from the compressed sensing (CS) theory, there is a qualitative difference between the {+1,–1} Bernoulli matrices and the {+1,0} Bernoulli matrices in their compressed sensing performance. It has been proven that CS algorithms work better for mask {+1,–1} [12].

Finally, we also used the approach labeled as *Diffsum*, where we calculated with two images, where one is a difference and the other one is a sum of the two traces. While we seemingly gain no benefits from using *Diffsum*, the difference and the sum of the two images might be beneficial. It provides us with more information about the image intensity magnitude and improves the reconstructions, as we discuss later.

Table 1. Different Approaches to Data Processing

Approach	*Matrix Notation*
Single	D = [U]
Double	D = [U; L]
Diff	D = [U-L]
Diffsum	D = [U – L; U + L]

Table 1 shows an overview of the labels. Here, U denotes the image corresponding to the upper first-order image, L represents the lower first-order image, D is a detector image, and the final detector image D' = [D ZO], where ZO is a sum of the upper and lower ZO images. Detector image D' is fed to the reconstruction algorithm, whose core is TwIST, which transforms it into a datacube with 123 spectral channels. For a detailed description of the data processing, see Supplement 1.

We measured various scenes by using our CASSI-based camera and carried out the hyperspectral datacube reconstructions to test the camera performance.

In Fig. 2, there is a reconstructed scene of an opaque cross evenly illuminated by a white light. We present selected spectral slices for different approaches—compare the rows. There are also two spectral slices of the *Single* and *Diffsum* approaches, magnified for better comparison. It can be seen that for such a simple scene, the approaches provide us with a similar quality of image reconstruction. However, the intensity distribution in reconstructed spectral slices of *Single* is less homogenous. In addition, all approaches except *Diffsum* leave residual intensity in the region below 500 nm, where the incoming light was cut off by filter OG-515 (see the corresponding spectral slices or Fig. S3 in Supplement 1). The zero spectral intensity in this spectral region is, therefore, a useful measure of the spectral reconstruction quality.

While the scene in Fig. 2 consists of a single spectral shape modulated in intensity, Fig. 3 depicts the reconstruction of the scene with four color filters illuminated by a broadband light source. Each quadrant, therefore, featured an entirely different spectrum. *Single*, *Double*, and *Diffsum* were able to accurately recreate the original filters in corresponding quadrants, while *Diff* struggles to reconstruct the green and blue filters in the bottom two quadrants (see spectral slices 552 nm and 588 nm). This is caused mainly by the aberrations in our system, which are, moreover, slightly different for the upper and lower images. Therefore, the image intensity within one line of a random mask leaks into the neighboring line and distorts the reconstruction. The summed image, which provides a guideline about the actual local intensity on the detector, is then able to compensate for this problem in the *Diffsum* approach. It is worth noting that *Single* or *Double* reconstruction cannot reconstruct well the onset of the yellow filter spectrum (see Fig. S4 in Supplement 1). Due to the strong signal from the red filter, we can observe that the spectrum tends to follow the red-filter spectrum in these cases, which is even more prominent with *Diff*. By contrast, the *Diffsum* reconstruction can distinguish between the two aforementioned filters better (see the enlarged spectral slices in Fig. 3). While the benefit of the Diffsum approach for the image quality is not so

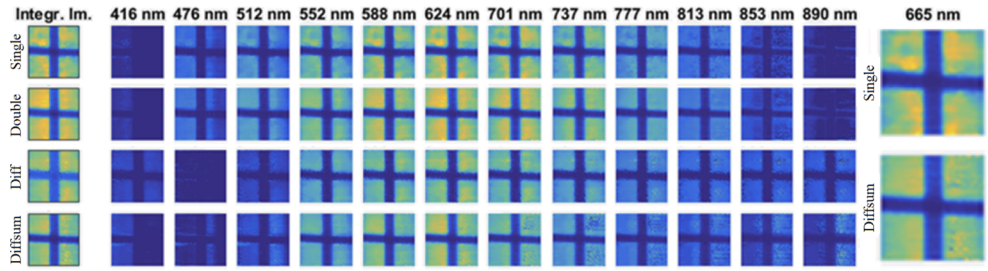

Fig. 2. Reconstructed spectrally integrated image and individual spectral slices of an opaque cross illuminated with broadband light by using four processing approaches (see Table 1). Two spectral slices of *Single* and *Diffsum* were enlarged for comparison.

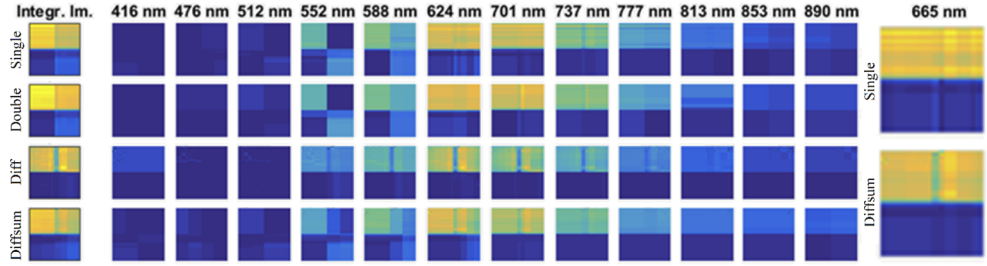

Fig. 3. Reconstructed spectrally integrated image and individual spectral slices of four color filters illuminated with broadband light by using four processing approaches (see Table 1). Two spectral slices of *Single* and *Diffsum* were enlarged for comparison.

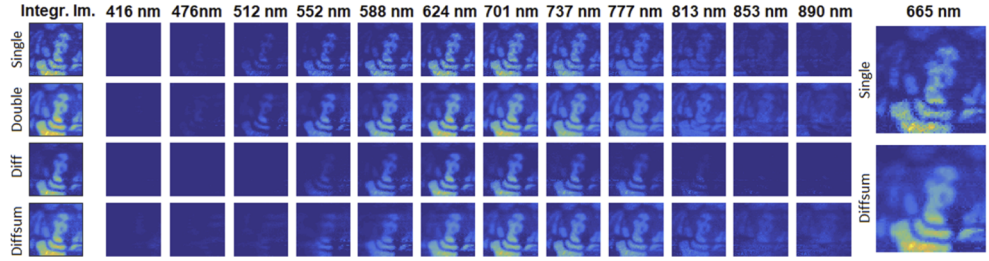

Fig. 4. Reconstructed spectrally integrated image and individual spectral slices of stained glass foil illuminated with broadband light by using four processing approaches (see Table 1). Two spectral slices of *Single* and *Diffsum* were enlarged for comparison.

prominent here, the ability to discern different spectral features is highly improved.

Finally, in Fig. 4, we depict the reconstruction of a complex scene with many varying spectral regions. It is possible to notice the improvement in the spatial quality of the reconstructed images when we extend into the double mask approaches. Nevertheless, even for the *Diffsum* case, we attain only qualitative agreement between the reconstructed and measured spectral shapes of the individual regions.

So far, we have discussed the results qualitatively. Hereafter, we will focus on quantification of the benefits connected to using two complementary masks. It was not possible to quantify the reconstruction quality of the experimental data by residuals from the detector image because we observed that we were able to obtain low residual metrics even for a reconstruction that clearly did not match the original scene. Hence, we created a set of artificial datacubes and detector images faithfully simulating the real detected images by a careful analysis of the aberrations present in our system (see Supplement 1 for details). We also created a few synthetic detector images using CAVE database data [13] as a template.

Comparison between the original and reconstructed datacubes was carried out by the peak signal-to-noise ratio (PSNR), spectral angle mapper (SAM) method for finding the spectral match, and the structural similarity (SSIM) index for measuring the spatial similarity. The SAM and SSIM values can be found in Supplement 1 (Table S3).

Firstly, we evaluated the effect of using two complementary random masks. We compared it to the situation where we use two random masks not related to each other. We consistently attained a higher reconstruction fidelity for the complementary masks (see the results in Table S2 in Supplement 1). We ascribe it to the fact that we obtain a non-dispersed image without any pixels missing for the complementary masks, which in turn improves the reconstruction quality.

Secondly, we applied the simulations on the used complementary masks, where we used the four approaches listed in Table 1. Table 2 provides an overview of the results achieved during the reconstruction for each case characterized by the PSNR. We point out that PSNR was calculated by scaling the whole datacube by a single factor, i.e., not by slice-by-slice comparison. In accordance with the real data reconstructions, it proves that the double mask approach *Diffsum* surpasses *Single*, while it

Table 2. Reconstruction Results for Different Scenes

Approach	*Scene A (Cross)* PSNR	*Scene B (Filters)* PSNR	*Scene C (Feathers)* PSNR
Single	17.78	21.78	21.29
Double	18.79	21.90	21.66
Diff	16.26	21.35	19.90
Diffsum	19.67	22.11	21.85

turned out that *Diff* was providing the worst results for some scenes. We ascribe this decrease in the reconstruction quality to the huge uncertainty in magnitude, which arises when two similar datasets, which are shifted with respect to each other, are subtracted. However, in the case of *Diffsum* the information about the magnitude is still present in sum of the snapshots. This combines the differential characters of the random masks while it retains the information about the image intensity scaling.

For the experimental data, the difference between the modes is more prominent for the complex scenes, while the simple scenes feature a similar image quality. Yet, the fidelity of the reconstructed spectra is improved even for the simple scenes.

In conclusion, we demonstrated a simple optical setup for a single-snapshot double-image CASSI system in a broad range of 400–900 nm with 123 spectral slices. Our experimental results, confirmed by a set of simulations, show that capturing two images of the same scene encoded by different random masks is superior to the standard approach, i.e., we gain better reconstruction quality. Furthermore, owing to the uniqueness of our system, where we use both ZO and first-order images, we are able to set the initial estimate of the datacube very close to the measured datacube, which in turn decreases the number of iterations needed for the reconstruction.

We also performed a comparison of artificial data reconstructions between complementary and non-complementary masks, which confirms that using two complementary masks provides us with more information and, therefore, better reconstruction quality. Hence, our system works like a differential CASSI (D-CASSI) method, where we are able to utilize a random mask consisting of {–1,+1} pixels.

From the selected approaches to the measured data, the best one and, at the same time, the most robust is *Diffsum*, which works well, particularly with the aberrated imaging system. Note that the optical setup was not optimized for the double mask approach and, therefore, the reconstruction quality could be further improved by limiting the aberrations in the system.

In summary, with a simple adjustment of the system without increasing its complexity, one can obtain more information about the measured scene and improve the compression ratio and the reconstructed image quality. Moreover, CASSI systems utilizing spatial light modulators (SLMs), such as digital micromirror devices (DMDs), for generating random mask patterns could benefit from our proposed approach as it can be straightforwardly implemented simply by generating complementary patterns one after another without any need for optical system modification. However, capturing two or possibly multiple images of the same scene encoded by different random masks simultaneously is a way to avoid using SLMs and retain the main advantage of CASSI—the single snapshot.

Funding. Technická Univerzita v Liberci (SGS-2019-3054, SGS-2021-3022); Ministerstvo Školství, Mládeže a Tělovýchovy (CZ.02.1.01/0.0/0.0/16_026/0008390).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. J. M. Bioucas-Dias and M. A. T. Figueiredo, IEEE Trans. on Image Process. **16**, 2992 (2007).
2. D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, Appl. Opt. **49**, 6824 (2010).
3. Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather, Opt. Lett. **36**, 2692 (2011).
4. H. Arguello and G. R. Arce, J. Opt. Soc. Am. A **28**, 2400 (2011).
5. L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, Appl. Opt. **54**, 848 (2015).
6. L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu, in *Visual Communications and Image Processing (VCIP)* (2016), pp. 1–4.
7. H. Rueda, H. Arguello, and G. R. Arce, in *IEEE International Conference on Image Processing (ICIP)* (2015), pp. 2572–2576.
8. L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 4942–4950.
9. X. Yuan, T. Tsai, R. Zhu, P. Llull, D. J. Brady, and L. Carin, IEEE J. Sel. Top. Signal Process **9**, 964 (2015).
10. J. Hlubuček, J. Lukeš, J. Václavík, and K. Žídek, Appl. Opt. **60**, 1463 (2021).
11. H. Arguello and G. R. Arce, IEEE Trans. on Image Process. **23**, 1896 (2014).
12. F. Soldevila, P. Clemente, E. Tajahuerce, N. Uribe-Patarroyo, P. Andrés, and J. Lancis, in *Imaging and Applied Optics 2016* (Optica Publishing Group, 2016), paper CW5D.5.
13. Columbia Vision and Graphics Center. (2022) Multispectral image database. [Online]. Available: http://www.cs.columbia.edu/CAVE/databases/multispectral/

Research Article Vol. 60, No. 5 / 10 February 2021 / Applied Optics 1463

applied optics

Enhancement of CASSI by a zero-order image employing a single detector

J. Hlubuček,* J. Lukeš, J. Václavík, and K. Žídek

Regional Center for Special Optics and Optoelectronic Systems (TOPTEC), Institute of Plasma Physics, Czech Academy of Sciences v.v.i., Sobotecká 1660, 511 00 Turnov, Czech Republic
**Corresponding author: hlubucek@ipp.cas.cz*

Received 12 November 2020; revised 18 January 2021; accepted 18 January 2021; posted 20 January 2021 (Doc. ID 414402); published 10 February 2021

Coded aperture snapshot spectral imaging (CASSI) makes it possible to recover 3D hyperspectral data from a single 2D image. However, the reconstruction problem is severely underdetermined, and efforts to improve the compression ratio typically make the imaging system more complex and cause a significant loss of incoming light intensity. In this paper, we propose a novel approach to CASSI that enables capturing both a spectrally sheared and an integrated image of a scene with a single camera. We performed hyperspectral imaging of three different testing scenes in the spectral range of 500–900 nm. We demonstrate the prominent effect of using the nondiffracted image on the reconstruction of data from our camera. The use of the spectrally integrated image improves the reconstruction quality, and we observed an approximately fivefold reduction in reconstruction time.

https://doi.org/10.1364/AO.414402

1. INTRODUCTION

Hyperspectral imaging (HSI) instrumentation is essential for many applications ranging from scientific research, such as volcanology [1] or imaging the chiralities of single nanotubes [2], to practical problems, including food analysis and safety inspection [3,4], medical imaging [5], quality control [6], forensic sciences [7,8], or art conservation [9].

Besides standard methods, commonly used to acquire a hyperspectral (HS) data cube, such as whiskbroom, pushbroom, and plane scanning, a range of new techniques has been developed with the vision to create a single-snapshot HSI that can be operated with a high frame rate and that does not require any movable part [10]. One of the methods is coded aperture snapshot spectral imaging (CASSI), based on compressed sensing [11,12].

CASSI can outperform the standard techniques, mainly in the length of the acquisition time, since it captures the whole data cube in one instance, i.e., a snapshot, eliminating the need for scanning. This makes the system highly robust. At the same time, the single-frame CASSI system has certain limitations, including image quality, compression ratio, and the time needed for the HS data cube reconstruction, since the reconstruction problem is severely underdetermined.

It is possible to improve the reconstruction quality of CASSI, for instance, by optimizing a coded aperture [13,14], utilizing multiple camera shots [15–17], or using a higher-order image reconstruction [18]. On the other hand, refining the method often brings in certain limitations. Multiframe CASSI loses the advantages of using a single snapshot, while more complex models for the detector description slow down the reconstruction process. Another promising way to boost the performance of CASSI is to capture a nondiffracted image that aids in the reconstruction. However, this approach normally requires splitting an incoming beam and employing two cameras [19–23], which makes the CASSI system inconveniently complex and causes a loss in the light intensity, which can reach as much as 50% [23].

Another limitation of the CASSI method consists in the size of the measurable spectral range. The spectral reconstruction can be highly improved by identifying key spectral features in the spectrum for specific applications [24]. This is, however, not our case, as we aim at a reconstruction of an arbitrary spectral shape, including spectrally flat broadband sources. Acquisition of a broader bandwidth decreases the compression ratio, which lowers the quality of the retrieved HS information. Therefore, the above-mentioned upgrades of CASSI typically aim at increasing the compression ratio along with capturing a narrow spectral range. At the same time, the CASSI reconstruction assumes an ideal image for each wavelength, which brings in the necessity to highly reduce optical aberrations of the CASSI system in the case of spectrally broad light. This leads to complex optical systems limited in their spectral range. Hence, there is a trade-off between the ability to carry out broadband HSI and the complexity of the setup. This is even more prominent in the infrared (IR) spectral range, where the construction of complex systems is costly and their precise alignment is a challenging task.

In this paper, we present a robust concentric HS camera based predominantly on off-the-shelf optics, which can be used for CASSI HSI. In contrast to previous reports, we aim at obtaining HSI in a broad spectral range between 500 and 900 nm covered by 123 spectral frames. In combination with the simplicity of the camera, the broad spectral range leads to the presence of aberrations in the system. This camera serves as a model system for the perspective of IR CASSI imaging, where the acquisition of a broad spectral range is needed to capture and distinguish between different chemical agents in the IR region.

However, the uniqueness of our HS camera lies in the design of the dispersive elements, which are able to attain both a nondiffracted image and first-order (FO) diffraction with a single detector. We demonstrate that by using a zero-order (ZO) image of a diffraction grating, we can highly improve the reconstruction quality of the system in spite of the aberrations present. Moreover, owing to the camera construction used, we utilize the light intensity, which is otherwise dumped in other grating-based CASSI systems [23].

By providing measurements of three testing scenes, we show that the use of ZO diffraction is indispensable for the aberrated system in order to attain spatial quality of HS data cube reconstruction. This is particularly prominent for scenes of spectrally broad light. We compare the use of the ZO in the calculation of an initial guess in the iterative reconstruction, as well as in the reconstruction itself. The presented concept can serve as an efficient approach to improving reconstruction in CASSI systems suffering from aberrations and low compression ratio.

2. EXPERIMENTAL SETUP

The HS camera used, depicted in Fig. 1(A), was built based on off-the-shelf optics except for elements L2 and P, which were manufactured at our facilities. Its main features are a high numerical aperture (NA $\sim$ 0.35) and a telecentric object (mask) space. In the scheme, L denotes plano–convex lenses, D denotes doublets, M is a random mask, P is a prism, and G is a grating. A detailed description of the system with a list of all its elements can be found in [25]. A total of six optical elements available from optics catalogs and a custom-made lens (L2) and prism (P) were used for the construction. A combination of the transmission grating (G, Thorlabs, 300 lines/mm) and the custom-made prism (P, SF11 optical glass) allows for a concentric construction of the camera, which is beneficial for calibration, and it also enables simple mechanical housing in a single tube. Mask M was a binary pattern, which was prepared via photolithography on a BK7 substrate with a thin chromium layer. It has 64 pixels $\times$ 64 pixels and a side length of 13.55 mm. The resulting image, which consists of both FO and ZO diffraction, was detected by using a Manta G-507 camera (Sony IMX264, resolution 2464 $\times$ 2056).

Due to the different spectral response of the optical system for the FO and the ZO, we characterized the relative intensity between the FO and the ZO intensities, which we denote as η_{FZ}. The intensity ratio, affected dominantly by the grating response, is depicted in Fig. 1(B). The spectral efficiency of the FO versus the ZO was employed in the calculations to reliably reproduce the detector image in Eq. (2). The monochromatic light for spectral calibration was obtained using a monochromator (Chromex 250 IS) in combination with a broadband quartz tungsten-halogen lamp (Thorlabs).

For the sake of the testing experiments, the testing scenes described below were imaged on the mask M by a single thin lens combined with a cutoff filter OG-515, which restricted the measured spectral range below 500 nm, as we explain below.

3. DATA PROCESSING AND RECONSTRUCTION

HS data cube reconstruction requires a transfer of the captured detector image with a high resolution (2464 pixels $\times$ 2056 pixels) into an image of the FO and the ZO corresponding to the resolution of the random mask (64 pixels $\times$ 64 pixels). First, the detector image is cropped and resized to match the pixel size of the random mask. The cropping employs calibration with a diffused monochromatic light (Nd:YAG laser, 532 nm). The crude cropping is based on aim pointers. These are transparent pixels located in the proximity of the mask, which can be identified in the dark detector area. Owing to the narrow spectrum of the calibration laser, the image of the diffused laser light on the detector is an image of the random mask without any spectral shear. We determined the cropping range of both orders by searching for the best correlation between the image and the random mask.

Since the detector has a higher resolution than the random mask—one mask pixel corresponds to approximately eight pixels on the detector—it is necessary to resize the cropped image. For the sake of contrast improvement, we avoid the border pixels, which, in the sense of binary mask pattern, could be classified as "gray." For the ZO image, the border pixels are avoided in both directions, while for the FO one, the omission can be performed only in the direction of spectral shearing. The image, where the border pixels were nullified, is consequently rescaled into a 64 pixel $\times$ 186 pixel FO image and a 64 pixel $\times$ 64 pixel ZO image corresponding to the mask pixels.

The processed data are reconstructed using the TwIST algorithm [26] minimizing the expression,

$$f(D) = \frac{1}{2}\|I - \hat{W}D\|^2 + \lambda\Phi(D), \tag{1}$$

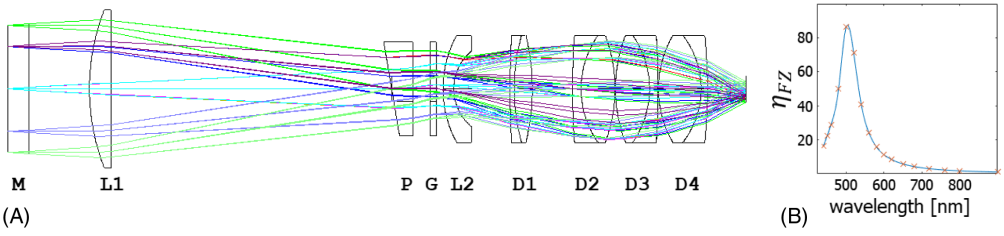

Fig. 1. (A) Scheme of our system described in the text; (B) spectral dependence of a relative intensity between FO image and ZO image η_{FZ}.

where I is the detector output; $\hat{W} = \hat{S}\hat{M}\hat{H}^{-1}$ is an operator describing the propagation of the incoming light through the system, including modulation by the random mask $\hat{M}$ and the spectral slice placement $\hat{S}$ in the FO and ZO images. $D(x, y, \lambda)$ is the HS data cube, where each spectral frame is transformed by the Haar wavelet transform $\hat{H}$. We use l_1-norm regularization $\Phi(D) = \sum |D|$, since the common scenes are sparse in the Haar wavelet basis. The regularization term is weighted by a coefficient λ, which can emphasize the sparsity of the reconstructed data cube.

The FO and the ZO can be included in the operator of spectral shearing $\hat{S}$ as

$$\hat{S} = \sum_{\lambda} [\eta_{FZ}(\lambda)\, \hat{T}_1(\lambda) + \hat{T}_0], \tag{2}$$

where $T_1(\lambda)$ is the wavelength-dependent translation of the image to the FO area and T_0 is the wavelength-independent translation of the image to the ZO area. The coefficient $\eta_{FZ}(\lambda)$ is the measured spectral efficiency of the FO versus the ZO depicted in Fig. 1(B).

The TwIST algorithm uses two operators: (i) to transform the data cube to the detector image, and (ii) to transform the detector image into the data cube. These operators correspond, in the compressed sensing theory, to the sensing operator $\hat{W}$ and to its transposition $\hat{W}^T$, respectively. Since a matrix representation of $\hat{W}$ and $\hat{W}^T$ would be very large and unsuitable for fast reconstruction, we evaluate them as functions.

Since TwIST is an iterative algorithm, an important factor is the initial guess of the HS data cube. Therefore, the ZO image can be implemented not only in the TwIST reconstruction itself, but also in the initial guess. It is highly favorable, in the sense of reconstruction time and quality, to make the initial guess as similar to the real data cube as possible.

An issue connected with the use of the ZO consists in the fact that the ZO image has blank pixels where the binary values of the random mask are equal to zero. We have overcome this by approximating these pixels by the mean value of their neighboring pixels. At the same time, the reconstruction needs to take into account that the ZO image is not evenly represented by all wavelengths.

The initial guess was created from the detector output I for each wavelength of a spectral slice as follows:

1) We extracted the 64 pixel × 64 pixel spectral slice $\Gamma(\lambda)$ of the HS datacube from the detector FO image, where the slice position corresponds to the selected wavelength λ. We multiplied the slice with the random mask: $\Gamma(\lambda) = \hat{M}\hat{T}_1^{-1}(\lambda)\, I$.
2) The spectral weight of the slice was calculated as a sum of all elements of the slice $\Gamma(\lambda)$: $w(\lambda) = \sum_{x,y} \Gamma(\lambda)$.
3) The ZO image Z extracted from the detector was used to correct the spectral slice after normalization by its mean value $\bar{Z}$: $\tilde{\Gamma}(\lambda) = \frac{Z}{\bar{Z}}(\Gamma(\lambda) + Z)$.
4) The initial guess $G(\lambda)$ was obtained by treating the $\tilde{\Gamma}(\lambda)$ slice with total variation denoising $\hat{N}$ corresponding to the Rudin–Osher–Fatemi denoising model and the denoised slice is multiplied by its spectral weight: $G(\lambda) = w(\lambda).\, \hat{N}\Gamma(\lambda)$.

The resulting data cube guess G is finally rescaled by the ratio between the original detector image intensity and the detector image intensity obtained by applying the operator $\hat{W}$ to the data cube guess.

4. RESULTS AND DISCUSSION

We carried out a set of experiments where we studied the HS data cube reconstruction from our broadband-aberrated HS camera based on the CASSI method. As we described in the previous two sections, besides the standard CASSI method, where the image is modulated by a mask M and spectrally dispersed, the construction of our device makes it possible to also capture the ZO diffraction. Hence, we exploit a part of the light intensity that would otherwise be lost in a standard system, and we use it to gain more information about the measured scene. The aim of the experiments was to reveal the effect of the information about the ZO image on the data cube reconstruction.

In order to gain a quantitative evaluation of the reconstruction quality, we created artificial data, faithfully simulating the real detected FO images as well as the ZO image by a careful analysis of the aberrations present in our system. Namely, we included the effect of wavelength-dependent: (i) vertical shift of spectral slices on the detector, and (ii) spectral slices' acutance. The effects were simulated by varying the size and position of each spectral slice jointly with a wavelength-dependent Gaussian filter. The scale of aberrations was extracted based on the acquired images of a monochromatic source illuminating a mask [see Fig. 2(A)], as we discuss in the next paragraphs. The simulated data were calculated to follow the camera resolution, and they were processed and reconstructed by using the exact same procedure as the experimental data.

The simulations allowed us to compare the reconstructed data cube with the ground truth and, therefore, to attain a quantitative measure of the reconstruction quality. We define the *difference* Δ between the ground truth and the reconstructed artificial data cube by least squares, where we optimize scaling factor c to minimize the *difference* value,

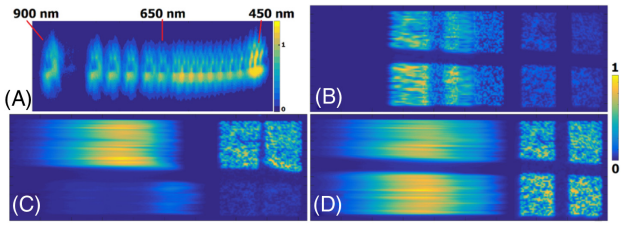

Fig. 2. (A) FO images of a fixed spot on the random mask illuminated by a set of monochromatic lights with wavelengths ranging from 440 to 900 nm (superimposed normalized images). Differences in the spot vertical position, scaling, and sharpness emerge from aberrations in the FO image, image resolution: 80 pixels × 1050 pixels; (B)–(D) scenes reconstructed in the article (normalized, color bar on the right), image resolution 600 pixels × 2260 pixels; (B) quasi-monochromatic laser sources illuminating dark cross; (C) spectrally broad light transmitted through four color filters; (D) spectrally broad light illuminating dark cross.

$$\Delta = \min_c \left\{ \frac{1}{n} \sum_{i=1}^{n} (c.y_i - \bar{y}_i)^2 \right\}, \quad (3)$$

where residuals for the ith point are calculated as a difference between the original data cube value y_i and reconstructed data cube value $\bar{y}_i$ and n is the number of data cube voxels.

The camera was constructed with a primary restriction on the number of elements and their off-the-shelf availability and, at the same time, we use the camera on a broad spectral range of 500–900 nm. Due to these restraints, the resulting detector image is aberrated.

We can visualize the aberrations present in the system [see Fig. 2(A)] by a superposition of detected images of a single spot on a mask illuminated by a set of quasi-monochromatic wavelengths. The detector was consequently illuminated in the spectral range of 440–900 nm, where the wavelength of the imaged spot decreases from left to right on the detector. The image for each wavelength was normalized before being added to the overall sum. Note that, compared to the other panels in Fig. 2, panel (A) is highly rescaled to demonstrate the aberrations.

For wavelengths around 450–500 nm [Fig. 2(A), on the right], the spot vertical position changes rapidly. This discrepancy is around one mask pixel, which makes the correct reconstruction impossible. Therefore, we used the OG-515 filter to block this problematic part of the spectrum. At the same time, you can see that the image acutance changes with wavelengths, and a sharp image is obtained only in the central part of the spectrum. This is another source of imperfections in the reconstruction.

Figures 2(B)–2(D) show a detector output of three different scenes: (i) an opaque cross illuminated simultaneously by a green laser and a red diode [Scene A, Fig. 2(B)]; (ii) four color filters illuminated by a broadband light source [Scene B, Fig. 2(C)]; (iii) an opaque cross illuminated by a broadband light source [Scene C, Fig. 2(D)]. On the left-hand side in the respective pictures, you can see the FO diffraction and on the right is the ZO diffraction. In Fig. 2(B) the intensity of the ZO diffraction is very weak, which is caused by using only two wavelengths and by the spectral effectivity of the FO versus the ZO. As you can see, the FO image of the green laser is basically an image of the random mask, since the laser spectral width is well below the spectral resolution of our system. On the contrary, the FO image of the red diode is a bit sheared due to the spectral width, FWHM being 18.5 nm.

It is worth noting that the spatial resolution of reconstructed scenes is restricted by the resolution of the mask M used, and not by the detector. While the photolithographic process used allows fabrication of a much finer binary mask, the resolution is limited by the aberrations present in our system. For example, in the spectral range where the light source used is the most intense, i.e., 520–720 nm, the variance in the vertical shift of different images is less than one-quarter of the mask pixel. This is still a feasible inaccuracy, as we do not include in our calculations the border pixels between the lines of the mask (discussed in Section 3). However, the use of a finer mask, i.e. smaller pixels, would inevitably lead to wrongly encoded information on the detector, where the information from one mask line would leak into the neighboring ones.

To study the influence of the ZO in the reconstruction, it can be implemented in two ways: First, to improve the initial guess of the reconstruction, and second, to be included in the operator $\hat{W}$ in the TwIST.

Figure 3 depicts the effect of four different modes of (not) using the ZO: panel (A) standard CASSI reconstruction with no ZO information; panel (B) ZO-assisted initial guess calculation followed by a standard CASSI reconstruction avoiding the ZO inclusion; panel (C) initial guess omitting the ZO information while using the ZO in the TwIST reconstruction; and panel (D) using the ZO in both the initial guess estimate and the data cube reconstruction. All calculations employed the scene with an opaque cross illuminated by a broadband light source. In each panel, we present eight selected spectral slices of the data cube, together with the total spectrum in the bottom right graph, which is a sum of all elements of each slice.

As one can see in Fig. 3(A), without the information about the ZO, it was not possible to retrieve the vertical line of the imaged cross. This problem was commonly encountered in the scenes where a broadband light was included. An initial guess promoting the vertical feature by using the ZO [panels (B) and (D)] serves sufficiently in this case to retrieve the data cube, irrespective of the mode of the reconstruction itself. On the contrary, an incorrect initial guess cannot be compensated by using the ZO information in the reconstruction routine [see panel (C)]. It is worth noting that, even for reconstructions ignoring the vertical features, we can attain a reconstruction with low residuals in the detector estimate, i.e., $\|\mathrm{I} - \hat{W}D\|^2$ from Eq. (1). Therefore, the residuals cannot be generally taken as a good measure to assess reconstruction quality in our system. This was further confirmed by the simulations.

Even though the reconstructed slices of the data cube in Figs. 3(B) and 3(D) have correct spatial information, the overall spectra are not accurate for the wavelengths below 500 nm, where there was no light intensity due to the use of the OG-515 cutoff filter. Here we observe a significant effect of the regularization weight λ. By putting stress on the sparsity of the reconstructed signal, i.e., higher λ in Eq. (1), we obtain a better agreement in the spectrum, but the spatial information is impaired. (This effect can be observed in Fig. 5, presented below.) Note that it is possible to improve the reconstruction by restricting it to the range of 500–900 nm. However, the main aim here was to evaluate the limitations of our system and the reconstruction of the imperfect data.

An evaluation of the reconstructions shows that the best results were achieved while using the ZO in both 1) and 2) simultaneously. Hence, we will hereafter show only the comparison between the case of not using the ZO (i.e., the standard CASSI approach) and using it both in the initial guess and reconstruction.

In Fig. 4, we can see the reconstructed slices of the scene illuminated by a green laser and a red light-emitting diode. Note that the selected wavelengths of slices shown in Fig. 4 are different than in the other figures. The wavelengths were selected to match the maximum spectral intensity of the two peaks. It is possible for the algorithm to distinguish the cross even without the use of ZO, compared to Fig. 3(A), because the cross is visible in the FO image; see Fig. 2(B) (left), compared to Fig. 2(D)

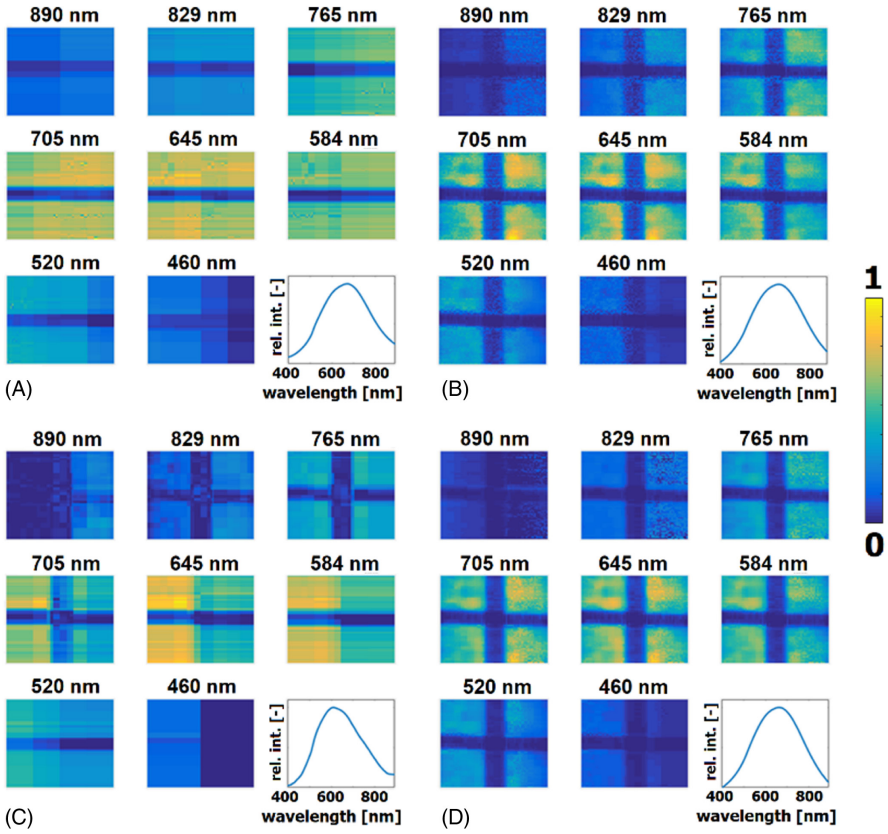

Fig. 3. Reconstruction of the scene from Fig. 2(D); each selected spectral slice is normalized to the maximum data cube value, color bar on the right. (A) Not using ZO; (B) using ZO in initial guess only; (C) using ZO in operator $\hat{W}$ only; (D) using ZO both in initial guess and operator $\hat{W}$.

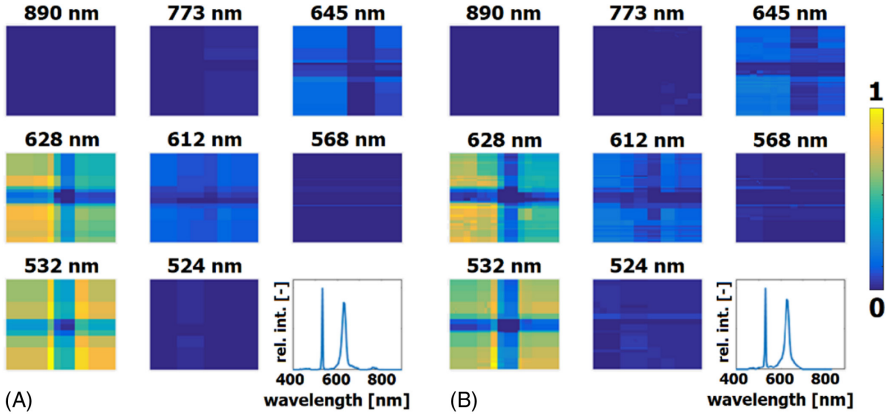

Fig. 4. Reconstruction of the scene from Fig. 2(B) (notice that the selected wavelengths are different than in the other figures); each selected spectral slice is normalized to the maximum data cube value, color bar on the right. (A) Not using ZO; (B) using ZO both in initial guess and operator $\hat{W}$.

(left). There are only minor differences between the reconstructions in Figs. 4(A) and 4(B), and the reconstructed spectrum has a good quality in both cases. Nevertheless, the reconstruction in Fig. 4(B) is slightly superior both in the sense of spatial reconstruction and spectrum quality. Therefore, the spectrally narrow features in the data cube can be well reproduced without the inclusion of the ZO.

Finally, we focused on the scene divided by four color filters illuminated by a broadband light source (see Fig. 5). The involvement of the broadband light causes the standard CASSI to face a problem with reconstructing vertical lines in the images, due to spectral shearing. This is highly improved by using the ZO in the reconstruction, as can be seen in the borderlines between the quadrants, i.e., the filters, which are visible in Fig. 5(B), while in Fig. 5(A) they are merged.

We extracted the spectrum of the light transmitted through each filter [see Figs. 6(B) and 6(C)], and we compared them with the spectra acquired by a fiber spectrometer (Ocean Optics, Flame), which were corrected for the grating efficiency [see Fig. 6(A)]. The colors of the lines used (yellow, red, blue, and

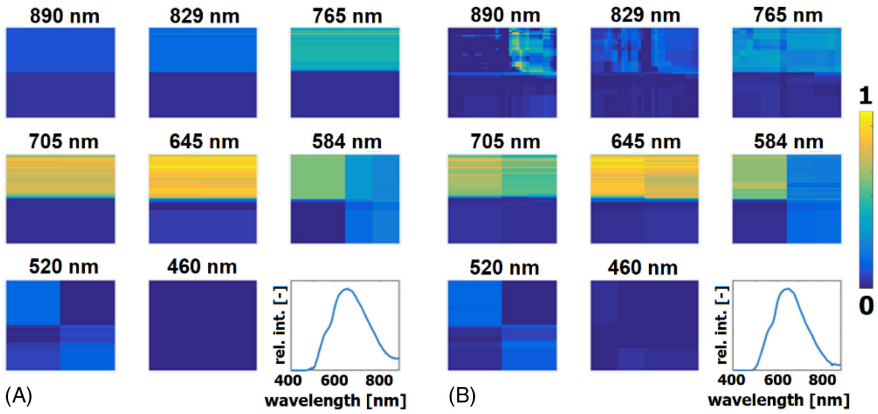

Fig. 5. Reconstruction of the scene from Fig. 2(C); each selected spectral slice is normalized to the maximum data cube value, color bar on the right. (A) Not using ZO; (B) using ZO both in initial guess and operator $\hat{W}$.

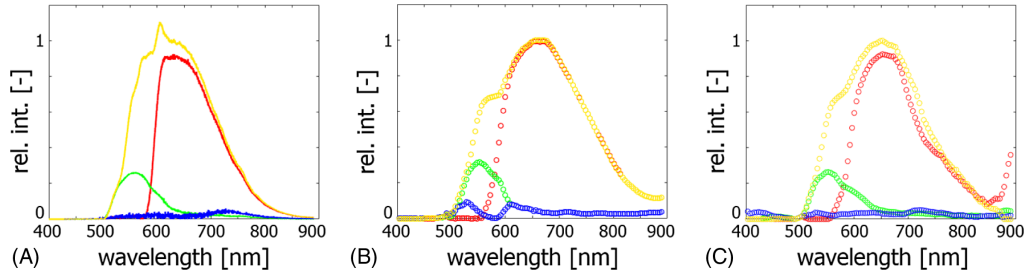

Fig. 6. Reconstructed normalized spectra of the individual filters of the scene from Fig. 2(C), where each line corresponds to a single filter located on the upper left (yellow), upper right (red), lower left (blue), and lower right (green). (A) Independently measured spectra; (B) reconstructed spectra not using ZO; (C) reconstructed spectra using ZO both in initial guess and operator $\hat{W}$.

green), correspond to the colors of the filters placed in the upper-left, upper-right, lower-left, and lower-right quadrant, respectively. Owing to the fact that we used a high λ value, we attained spectra that are cropped at 500 nm, in accordance with the OG515 filter used. We attained reasonable agreement between both the reconstructed and the reference spectra. Nevertheless, the reconstruction employing the ZO image reproduces very well—even the weak signal from the blue and green filter. The most problematic task is the reconstruction of the overlapping spectra of the red and yellow filters. Here, even the ZO-assisted reconstruction fails to fully reproduce the shape, in spite of reaching a better agreement.

We consistently observed that when the ZO is not used in the reconstruction, the resulting spectra are very dependent on the set parameters, and it is possible to obtain good results pictured in Fig. 6(B) only with a very narrow set of parameters, while the reconstruction with the ZO [Fig. 6(C)] is much more robust. On top of that, the use of the ZO, in the case of the scene with four color filters, greatly helps to shorten the time needed for the reconstruction (52 s without using the ZO versus 10 s using the ZO on a standard laptop), since it improves the initial guess and therefore it converges faster to the results obtained.

To quantify the effect of the ZO usage, we employed the calculations, where we simulated the aberrated detector image and its reconstruction under various conditions. For the sake of comparison between using the ZO and not using it, we evaluated the lowest attainable *difference* [see Eq. (3)], between the original and reconstructed data cube.

Table 1. Difference between the Original and Reconstructed Data Cube for Different Scenes

	Scene A	Scene B	Scene C
Not using the ZO image	2.1147e − 03	1.0002e − 03	9.7144e − 04
Using the ZO image	2.0556e − 03	9.2579e − 04	6.0386e − 04

In Table 1, you can see that the effect of the use of the ZO image depends on the properties of each detected scene. The effect for scenes illuminated with narrow spectral sources is only subtle (Scene A). In the case of broadband light, the influence could be of great importance (Scene C), especially for the scenes, where the spectra are dominated by a single light source. However, for specific scenes and parameter settings, the difference might be lessened (Scene B). Nevertheless, it is worth stressing that while the level of *difference* might be comparable for both the original and the ZO-assisted CASSI methods, the use of the ZO is much more robust with respect to change in reconstruction parameters.

5. CONCLUSIONS

We built a broadband HS single-snapshot camera with a limited number of optical elements based mainly on off-the-shelf optics. Our HS camera is capable of capturing a standard CASSI

snapshot of a scene together with a nondispersed ZO image on a single detector. Hence, we can attain more information about the HS data cube and use the incident light more efficiently.

We carried out HSI on a broad spectral range (500–900 nm) as well as simulations faithfully representing measured data in order to gain more control over the reconstruction algorithm. Due to the inevitable aberrations in the imaging system, we observed that the resulting image highly differed from the ideal case. Therefore, by using the standard CASSI approach, we attained a reliable reconstruction only for simple scenes with quasi-monochromatic light sources.

However, we have achieved a significant improvement in the reconstruction quality by including a ZO image in the CASSI reconstruction. We can employ the ZO image both in the initial guess and the iterative data cube reconstruction. Data show that capturing the ZO image and using it in the reconstruction is beneficial for reconstruction quality and time, which is decreased approximately fivefold. An important factor is that by using the ZO, we are able to estimate the data cube in the initial guess very closely to the measured scene.

We observe that the weight of the regularization term in the reconstruction algorithm has a profound effect on the spectral reconstruction quality, where high values of the weight promote correct spectra reconstruction, whereas low values improve the image spatial quality.

In spite of the improvement, the aberrations across the measured broad spectral range still lead to a severe problem with reconstruction quality. However, the results prove that using additional information about the detected scene can partly compensate for the image imperfections. This can be, in the future, utilized in the design of systems for the IR spectral range, where the reduced imaging system complexity can be of huge importance.

Funding. Ministerstvo Školství, Mládeže a Tělovýchovy (CZ.02.1.01/0.0/0.0/16_026/0008390); Grantová Agentura České Republiky (17-26284Y).

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. A. Gabrieli, R. Wright, P. G. Lucey, J. N. Porter, H. Garbeil, E. Pilger, and M. Wood, "Characterization and initial field test of an 8–14 μm thermal infrared hyperspectral imager for measuring SO_2 in volcanic plumes," Bull. Volcanol. **78**, 73 (2016).
2. D. Roxbury, P. Jena, R. Williams, B. Enyedi, P. Niethammer, S. Marcet, M. Verhaegen, S. Blais-Ouellette, and D. A. Heller, "Hyperspectral microscopy of near-infrared fluorescence enables 17-chirality carbon nanotube imaging," Sci. Rep. **5**, 14167 (2015).
3. R. Calvini, A. Ulrici, and J. M. Amigo, "Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging," Chemom. Intell. Lab. Syst. **146**, 503–511 (2015).
4. Y. Z. Feng and D. W. Sun, "Application of hyperspectral imaging in food safety inspection and control: a review," Crit. Rev. Food Sci. Nutr. **52**, 1039–1058 (2012).
5. O. Carrasco, R. B. Gomez, A. Chainani, and W. E. Roper, "Hyperspectral imaging applied to medical diagnoses and food safety," Proc. SPIE **5097**, 215–221 (2003).
6. S. Mahesh, D. S. Jayas, J. Paliwal, and N. D. G. White, "Hyperspectral imaging to classify and monitor quality of agricultural materials," J. Stored Prod. Res. **61**, 17–26 (2015).
7. R. L. Schuler, P. E. Kish, and C. A. Plese, "Preliminary observations on the ability of hyperspectral imaging to provide detection and visualization of bloodstain patterns on black fabrics," J. Forensic Sci. **57**, 1562–1569 (2012).
8. G. J. Edelman, E. Gaston, T. G. Leeuwen, P. J. Cullen, and M. C. G. Aalders, "Hyperspectral imaging for non-contact analysis of forensic traces," Forensic Sci. Int. **223**, 28–39 (2012).
9. C. Balas, G. Epitropou, A. Tsapras, and N. Hadjinicolaou, "Hyperspectral imaging and spectral classification for pigment identification and mapping in paintings by El Greco and his workshop," Multimedia Tools Appl. **77**, 9737–9751 (2018).
10. L. Gao and L. V. Wang, "A review of snapshot multidimensional optical imaging: measuring photon tags in parallel," Phys. Rep. **616**, 1–37 (2016).
11. D. J. Brady, *Optical Imaging and Spectroscopy* (Wiley, 2009).
12. G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: an introduction," IEEE Signal Process. Mag. **31**(1), 105–115 (2014).
13. H. Arguello and G. R. Arce, "Rank minimization code aperture design for spectrally selective compressive imaging," IEEE Trans. Image Process. **22**, 941–954 (2013).
14. H. Arguello and G. Arce, "Code aperture design for compressive spectral imaging," in *European Signal Processing Conference* (2010), pp. 137–140.
15. D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," Appl. Opt. **49**, 6824–6833 (2010).
16. Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather, "Development of a digital-micromirror-device-based multishot snapshot spectral imaging system," Opt. Lett. **36**, 2692–2694 (2011).
17. H. Arguello and G. R. Arce, "Code aperture optimization for spectrally agile compressive imaging," J. Opt. Soc. Am. A **28**, 2400–2413 (2011).
18. H. Arguello, H. Rueda, Y. Wu, D. W. Prather, and G. R. Arce, "Higher-order computational model for coded aperture spectral imaging," Appl. Opt. **52**, D12–D21 (2013).
19. L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," Appl. Opt. **54**, 848–858 (2015).
20. L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu, "Compressive hyperspectral imaging with complementary RGB measurements," in *Visual Communications and Image Processing (VCIP)* (2016), pp. 1–4.
21. H. Rueda, H. Arguello, and G. R. Arce, "Dual-ARM VIS/NIR compressive spectral imager," in *IEEE International Conference on Image Processing (ICIP)* (2015), pp. 2572–2576.
22. L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 4942–4950.
23. X. Yuan, T. Tsai, R. Zhu, P. Llull, D. J. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," IEEE J. Sel. Top. Signal Process. **9**, 964–976 (2015).
24. Z. He, N. Williamson, C. D. Smith, M. Gragston, and Z. Zhang, "Compressed single-shot hyperspectral imaging for combustion diagnostics," Appl. Opt. **59**, 5226–5233 (2020).
25. K. Zidek, O. Denk, J. Hlubucek, and J. Vaclavik, "Compact and robust hyperspectral camera based on compressed sensing," Proc. SPIE **10151**, 101510N (2016).
26. J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**, 2992–3004 (2007).

OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP) © OSA 2021

Improving Compression Ratio in CASSI

J. Hlubuček[1,2], J. Lukeš[1,2], K. Žídek[1]
[1]*Regional Center for Special Optics and Optoelectronic Systems (TOPTEC), Institute of Plasma Physics, Czech Academy of Sciences v.v.i., Za Slovankou 1782/3, 182 00 Prague 8, Czech Republic*
[2]*Technical University of Liberec, Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Studentská 1402/2, 461 17 Liberec, Czech Republic*
hlubucek@ipp.cas.cz

Abstract: We propose a new approach to improve the compression ratio in CASSI while maintaining the simplicity of the system and single-snapshot acquisition. We evaluate simulations of the modified method, which confirm the reconstruction quality improvement.

1. Introduction

Coded aperture snapshot spectral imaging (CASSI) is a method combining hyperspectral imaging and compressed sensing. CASSI makes it possible to recover 3D hyperspectral data from a single 2D image, the so-called snapshot. It is worth noting that the data compression is immense. Therefore, the standard CASSI method trades off between the quality of recovered data and the measured spectral range as well as the spectral resolution. A way to improve the reconstruction quality is to acquire more data about the measured scene. However, improving CASSI often brings in some other limitations, e.g., higher complexity of the system – using two detectors [1], slowing down the reconstruction/measurement process [2], etc. A recent enhancement of CASSI has been done, which consists in capturing a non-diffracted image of a scene next to the spectrally sheared one on the same detector [3].

Here we present a possibility of doubling the CASSI information by the use of two imaging lenses while taking into account the improvements made in [3]. Simultaneously obtaining two measurements of the same scene allows for an improvement of reconstruction quality while keeping the CASSI's main advantage – using a single snapshot.

2. Experimental methods

Simulations and reconstructions of data were evaluated by Matlab. Two different scenes were selected for reconstruction – a scene with an opaque cross – Fig. 1A (Scene A), and a scene with four different color filters (red, yellow, green, blue) in each quadrant of the scene – Fig. 1B (Scene B), both illuminated by a broadband light source. In the CASSI system, the 3D datacube – a scene multiplied by a broadband light spectrum, is first encoded by a random binary mask pattern for every wavelength. While it propagates through the optical system, it is spectrally sheared by a dispersive element and consequently captured by a detector. The resulting signal on the detector (see Fig. 1) was created by shifting every encoded spectral slice of the datacube by one pixel-column to the right compared to the previous one.

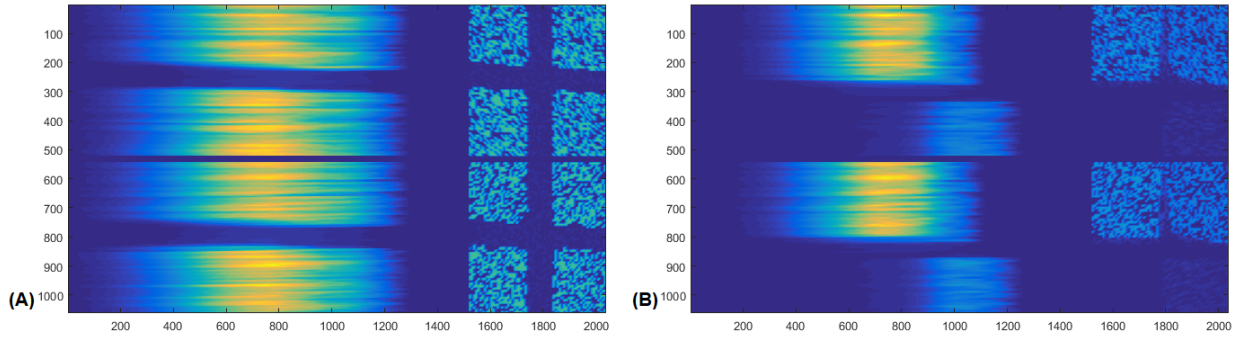

Fig. 1. Two types of scenes used in the simulations. (A) Scene A, (B) Scene B.

In Fig. 1, you can see two spectrally sheared images encoded by different random masks and two spectrally integrated images for each scene. It simulates the case of using two imaging lenses and capturing a non-diffracted image of the scene. It is worth noting that the data were created with higher resolution (521x1517 px spectrally sheared images, 521x521 px spectrally integrated images) than it was used during the reconstruction (64x186 px spectrally sheared images, 64x64 px spectrally integrated images) in order to be able to simulate the real case scenario. Thus, the reconstructed datacube has dimensions 64x64x123. Moreover, we added aberrations of the

system from [3] to the simulations, enabling us to carry out the simulations more faithfully. The datacubes were retrieved from the aberrated detector image by the TwIST (Two-Step Iterative Shrinkage/Thresholding) algorithm [4] for data reconstruction. which minimizes the following expression:

$$f(x) = \frac{1}{2}\|y - Wx\|^2 + \lambda\Phi(x)\,, \tag{1}$$

where W is the transfer matrix which includes all the processes during measurement of 3D datacube x such as imaging, encoding by a random mask or spectral shearing. An image on the detector is y, Φ is a regularizer, and λ is a regularization parameter that puts stress on the datacube sparsity.

3. Results

Evaluating reconstructions of simulated data shows that capturing two images of the same scene encoded by different random masks is superior to the standard approach. In Fig. 2 is a comparison of the original and reconstructed datacubes of Scene B. Selected spectral slices are shown as well as the overall spectrum for the standard method – see Fig. 2B, and the double lens method – see Fig. 2C.

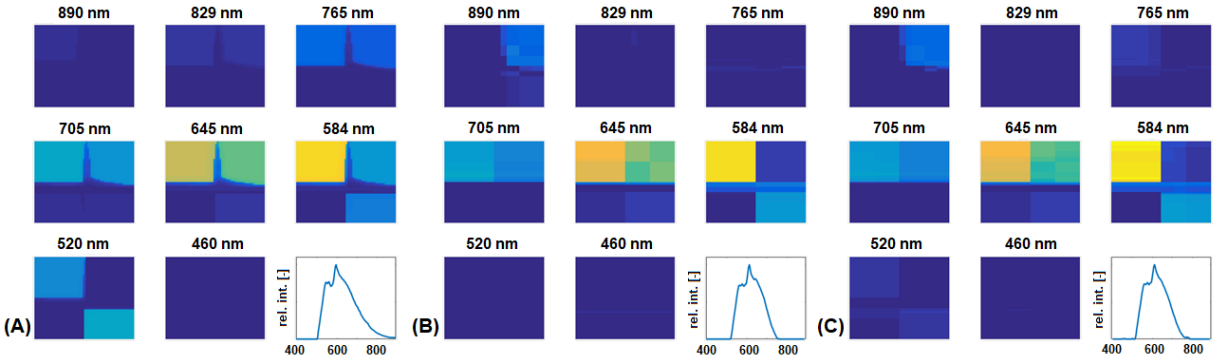

Fig. 2. Selected spectral slices and overall spectrum of (A) the original datacube (Scene B), (B) the reconstructed datacube – single lens, (C) the reconstructed datacube – double lens.

Visually there is not a big difference between the mentioned approaches. However, the improvement of the double lens method can be seen in Table 1, which shows the difference between the original and reconstructed datacube for both scenes. The difference was calculated using least squares related to the number of datacube voxels.

Table 1. Difference between the original and reconstructed datacube

	Scene A	Scene B
Single lens	5.81e-04	1.09e-03
Double lens	4.76e-04	9.82e-04

4. Conclusion

Using two imaging lenses and two random binary masks in the CASSI method is a straightforward way to obtain more information about the measured scene and, consequently, improve the reconstruction quality.

5. Acknowledgment

This work was supported by the Student Grant Scheme at the Technical University of Liberec through project nr. SGS-2019-3054; and project nr. SGS-2021-3022; and the Ministry of Education, Youth and Sports ("Partnership for Excellence in Superprecise Optics," Reg. No. CZ.02.1.01/0.0/0.0/16_026/0008390).

6. References

[1] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," Applied Optics **54**, 848–858 (2015).
[2] H. Arguello, H. Rueda, Y. Wu, D. W. Prather, and G. R. Arce, "Higher-order computational model for coded aperture spectral imaging," Applied Optics **52**, D12-D21 (2013).
[3] J. Hlubuček, J. Lukeš, J. Václavík, and K. Žídek, "Enhancement of CASSI by a zero-order image employing a single detector," Applied Optics **60**, 1463-1469 (2021).
[4] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**, 2992–3004 (2007).

Evaluation of using coded aperture imaging in the mid- and far-infrared region

Jiri Hlubucek, Karel Zidek
Regional Centre for Special Optics and Optoelectronic Systems (TOPTEC)
IPP, Academy of Sciences of the Czech Republic
Za Slovankou 1782/3, 182 00 Prague 8, Czech Republic
hlubucek@ipp.cas.cz

***Abstract*— Hyperspectral imaging in the infrared spectral region makes it possible to identify chemical compounds and, at the same time, locate the compound. We provide simulations of using coded aperture snapshot spectral imaging (CASSI) to reconstruct the hyperspectral information from a single snapshot. We study the effect of using a complex scene, scene illumination by a black-body radiation, and effect of adding a noise to the synthetic datasets. Our results show, that the use of CASSI method with a simple binary mask leads to partially satisfactory results for the realistic scenes in the sense of determining the chemical compound but not for retrieving the quality of reconstructed scene.**

Keywords—hyperspectral imaging; infrared spectrum; coded aperture snapshot spectral imaging; compressed sensing

I. INTRODUCTION

Hyperspectral imaging (HSI) denotes all methods where, in addition to an image, we obtain a spectrum of light at each point of the image. HSI in the infrared (IR) spectral region is of great importance, as it can provide us with large amount of information about the scene of interest that cannot be obtained in any other way. An example can be the remote sensing of chemical compounds. For this reason, HSI in the IR has been very lively topic in the recent decades.

A number of studies focuses on HSI in the near-IR spectral range, which is accessible for a commonly used optics and Ge- or InGaAs-based detectors. However, the so-called mid- and far IR ($\lambda > 2.5$ μm) region is not widely utilized due to the need to use uncommon optical materials and array detectors.

A possible solution to this problem can be utilization of compressed sensing (CS) methods. CS refers to a signal processing technique that uses the principle that many natural signals are sparse, i.e. can be described by only few components in a certain basis. Use of CS for sparse signals makes it possible to reconstruct the signals from far fewer measurements than the Shannon-Nyquist theorem requires. In other words, it is possible to reconstruct signals by finding solutions to underdetermined linear systems. For more details we refer reader to [1].

In this article we focus on the use of the so-called CASSI (Coded Aperture Snapshot Spectral Imaging) method in the IR HSI, where sharp spectral absorption peaks superimposed on the black body radiation represent a specific type of scenes. Moreover, the scene has to be recorded by using an IR detector with a high level of a dark noise. The main goal of this article is to evaluate feasibility of using the standard CASSI method for the IR HSI, which would allow a simpler and less expensive construction of HSI devices. An overview of the HSI is supplemented by samples of reconstructions of artificial data (hyperspectral scenes) where we simulate the presence of chemical compounds on parts of the image and subsequently we reconstruct the hyperspectral scene.

II. EXPERIMENTAL METHODS

Simulations and reconstructions of data were evaluated by Matlab. Two types of scenes were selected for reconstruction – a simple and a complex scene. The simple one was a scene with constant intensity in every pixel, while the more complex one was an image from infrared camera. Several different variations were simulated for every type of scene by using different sizes of the scene in pixels (32x32, 64x64, 128x128, ...), number of spectral slices (117, 235, 470, ...), and concentration of the chemical substance. Nevertheless, all simulations presented in this article were carried out by using 128x128 mask and scene image, resolved in 470 spectral slices, detected on a 128x597 detector.

As a chemical substance we chose isopropyl alcohol, if not stated otherwise, we employed the path-concentration of 1000 ppm m. The IR spectrum for isopropyl alcohol was obtained from The National Institute of Standards and Technology (NIST), data was compiled by: P.M. Chu, F.R. Guenther, G.C. Rhoderick, and W.J. Lafferty with resolution of 0.4820 cm^{-1} and parameters IFS66V (Bruker); 3-Term B-H Apodization.

For simulation we focused on one of the standard methods of compressed scanning, the so-called CASSI (Coded Aperture Snapshot Spectral Imaging) method. We refer reader for the detailed description of the CASSI method to a number of available articles [1, 2].

For image restoration during the reconstruction we used the TwIST (Two-Step Iterative Shrinkage/Thresholding) algorithm to minimize the following expression:

$$f(x) = \frac{1}{2}\|y - Kx\|^2 + \lambda\Phi(x) \qquad (1)$$

where K is the linear direct operator describing projection of a hyperspectral datacube x onto a single detector snapshot y, Φ is a regularizer, λ is a regularization parameter. We employed as a regularizing term a sum of total variations in each spectral

image. [3] Operator K is calculated, for the sake of simplicity, so, that each spectral slice (image for every wavelength) is shifted by one column in the resulting detected image y.

III. HYPERSPECTRAL IMAGING

A. Overview

There are three basic configurations of how one can obtain hyperspectral information (3D datacube). (1) *Whisker-broom* – the sample is scanned point by point, and for each such point one spectrum is recorded (1D detector, 2D scanning). (2) *Push-broom* – the detector acquires the spectral information for the entire line of pixels of the image simultaneously (2D detector, 1D scanning). The light passes through a slit and it is spectrally sheared on the detector, thus making is possible to record the spectral information along the entire line depending on the location from which the light comes. In this way a two-dimensional array is obtained which has one spectral dimension and one spatial dimension. For another spatial dimension of the datacube, we need to scan the sample in a direction perpendicular to the imaging line. (3) *Staring configuration* – (2D detector, 1D scanning) this type of configuration does not require any movement (or spatial scanning) of the sample or a slit, so it is also referred as "staring configuration". The incoming light is recorded on the detector as a two-dimensional spatial array for each wavelength. This is done by means of filters (band-pass filters [4] or adjustable acousto-optic filters [5]) which can be placed on a revolving disc or change the passing wavelength respectively. [1]

Whisker-broom and push-broom scans have excellent spatial and spectral resolution, however, the necessity to mechanically scan an image implies that the acquisition times are long. Typically, the times are in the order of tens of minutes to hours, depending on the size of the scanned area, the wavelength range and the number of scans per pixel [6]. For processes that are not stable in time is favorable to use the staring configuration since it is possible to record a complete datacube in a matter of seconds or minutes, depending on the number of scanned wavelength intervals.

Selection of the suitable method depends highly on the concrete field of application, since HSI is being used in a wide variety of fields, e.g. medical imaging [7], quality control and food analysis [8, 9], forensic sciences [10, 11], art conservation [12], etc.

B. Compressed sensing

In conventional signal processing, we are limited by the so-called Shannon-Nyquist theorem, which imposes that for the correct reconstruction of the signal, the sampling frequency must be at least twice as high as the highest frequency present in the signal. This is very inconvenient for capturing rapid processes or for the IR region, where we are significantly limited by the structural elements of IR cameras and their high purchase prices.

However, Shannon-Nyquist theorem can be bypassed by compressed sensing (CS), which is based on two assumptions – (i) sparsity of a signal and (ii) signal measurement by using a set of incoherent (often random) projections of the signal.

CS is often employed in imaging since common images count to the sparse datasets in the Fourier or wavelet space. For example, a conventional camera captures the scene pixel by pixel, creating a huge amount of RAW data. However, the image can be compressed to few percent of the original size without apparent loss of the image quality by using the strongest Fourier transform coefficients (JPEG compression). The problem is that we are not able to compress the scene until we capture it because we do not know *a-priori* which Fourier components will carry the important information about the image.

The so-called CASSI (Coded Aperture Snapshot Spectral Imaging) method employing the CS theory makes it possible to encode the whole hyperspectral scene (3D dataset) in a single instant (2D snapshot) using a random mask. The random mask (random pattern) serves as an incoherent measurement projection. By employing a spectral shearing (prism or grating) the random mask is shifted to different positions for different wavelength, thus enabling subsequent HSI reconstruction.

Variations of the CASSI technique are used also for shearing the temporal information (e.g. CACTI), thus making it possible to capture events taking place in the order of tens of ps (CUP technique) [13].

C. Using compressed sensing in IR hyperspectral imaging

Absorption of mid-IR light changes the fundamental vibrational and rotational states of the chemical bonds. When the molecule interacts with IR light, chemical bonds begin to vibrate more energetically, and thus affect absorption at certain wavelengths in the spectrum that are characteristic for each chemical bond.

The ability to absorb near-IR is relatively small and depends on the harmonic and anharmonic movement of molecules, which is due to electronic transitions. [14] Therefore, this paper is focusing on mid- and far-IR region.

There are not many articles on application of CS in mid- and far-IR HSI [15-17]. This could be attributed to the problematic connected with the need of special optic elements and detectors in IR region.

IV. RESULTS AND DISCUSSION

The CASSI technique is exploiting a coded aperture and dispersive element(s) to modulate the optical field from the scene, which is captured in one instance on the detector into the two-dimensional snapshot. We used a random binary mask (see Fig. 1, right panel) as a coded aperture to encode a scene. The random mask is blocking approximately ½ of the incoming light and the columns in the mask, owing to their randomness, are incoherent. We created a HS datacube $H(i,j,\lambda)$ by using the same scene $S(i,j)$ in all spectral images multiplied by a radiation spectrum of the light illuminating the scene $R(\lambda)$:

$$H(i,j,\lambda) = S(i,j).R(\lambda) \quad (2)$$

To simulate the presence of a chemical substance the central part of the scene was “contaminated” with isopropyl alcohol which caused distinguishable difference in intensity at specific slices of the datacube (see Fig. 1, middle panel). In other words, we multiplied the datacube $H(i,j,\lambda)$ by an absorption spectrum of the studied compound $A(i,j,\lambda)$ for each scene pixel i,j:

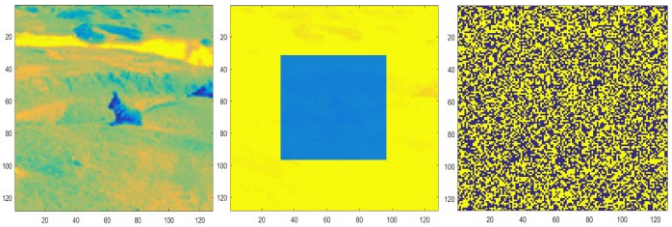

Fig. 1. Complex scene (left), slice of the datacube (complex scene 128x128 pixels) with applied chemical substance (middle), random mask 128x128 pixels (right)

$$H'(i,j,\lambda) = H(i,j,\lambda).A(i,j,\lambda) \quad (3)$$

$H'(i,j,\lambda)$ is a datacube that is coming to a HS camera where it is encoded by the random mask pattern $M(i,j)$ for every wavelength. The encoded image E can be expressed as:

$$E(i,j,\lambda) = H'(i,j,\lambda).M(i,j) \quad (4)$$

We used two types of test scenes S. One was a plain scene (S is constant), which equals for capturing an IR image obtaining a constant temperature through the whole scene, i.e. same intensity in each pixel. The second one was an arbitrarily chosen image from an IR camera (Fig. 1, left pannel).

Spectrum of the light illuminating the scene R was either left constant for initial experiments or set according to a black body radiation attribute (Planck's law) which is wavelength dependent. We set temperature to be 300 K for a plain scene or differ from 283 to 323 K for an IR image.

A. Detector signal

The detected IR light in a CASSI-type camera is transmitted through the IR optics and then is refracted by a dispersive element to different positions on the detector depending on the wavelength. Every spectral slice of the datacube (scene for each wavelength) was shifted on the detector by one pixel-column to the right compared to the previous slice, i.e. slices were overlaying each other, which led to the total signal on detector (Fig. 2, lower panel) $D(k,l)$:

$$D(k,l) = \sum_{\lambda} E(k,l+\lambda,\lambda) + n(k,l) \quad (5)$$

where the term $n(k,l)$ enabled us to add a certain noise level to the detected image.

In Fig. 2 (upper left panel) we can see the example of one spectral slice of the datacube *H'*. For the sake of clarity, we selected among many slices the wavelength, at which the absorption was the most significant. The resulting detector image for this scene is provided in Fig. 2 (lower panel).

B. Data reconstruction

Data reconstruction was evaluated by TwIST algorithm, which is an improved version of a standard IST algorithm [17]. For each slice of original datacube we obtain one slice of reconstructed datacube – see Fig. 2 (upper right panel) for an example of a reconstructed spectral slice. We can subsequently also recover the absorption spectrum of the chemical compound from the reconstructed datacube as a sum of the central area, where the chemical compound was in the original image. We obtain a good agreement between the original and reconstructed spectrum (see Fig. 3). The relative intensity and position of the

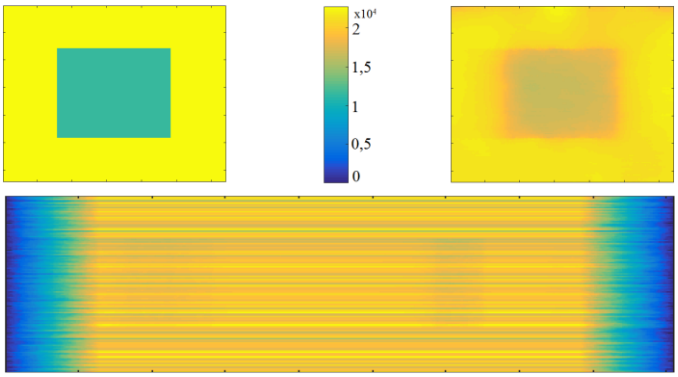

Fig. 2. Reconstruction of the simple scene. Slice of original datacube (upper left), detected image at the detector (lower), reconstructed slice of the datacube (upper right).

peaks to each other is particularly important to successfully determine the chemical compound and its volume.

The reconstructed data are satisfying in terms of recognition of the chemical compound and, most importantly, its localization. However, it is not possible to retrieve details of the original scene back. As you can see in Fig. 3 (small panels), there is not a significant difference between simple and complex scene in the reconstructed slices. This indicates that subtle changes in a scene due to absorption from a minor concentration of a chemical compound are likely to be suppressed by the reconstruction algorithm.

When we included the black body radiation (Planck's law) into the simulation, the quality of reconstructed spectrum is only slightly degraded (see Fig. 4), nevertheless the quality of the reconstructed datacube slices is notably worse in case of the complex scene – Fig. 4 (small panel). In this case, it would be very difficult to correctly localize the chemical compound. This is likely caused by a significant complexity of the IR spectra, which are sparse in terms of the image information, however, contain complex spectral information.

Finally, it is also worth noting, that with an addition of up to 5% noise the data reconstruction is still reasonable for simple

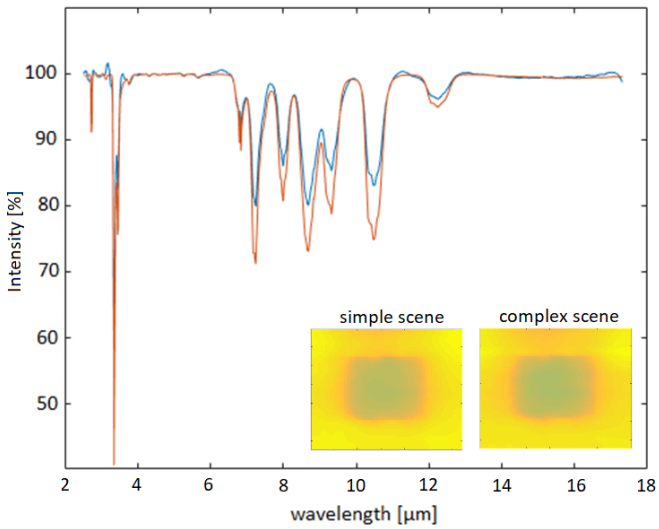

Fig. 3. Original (red) and reconstructed (blue) spectrum of constantly irradiated complex scene, reconstructed slices of datacube (small panel).

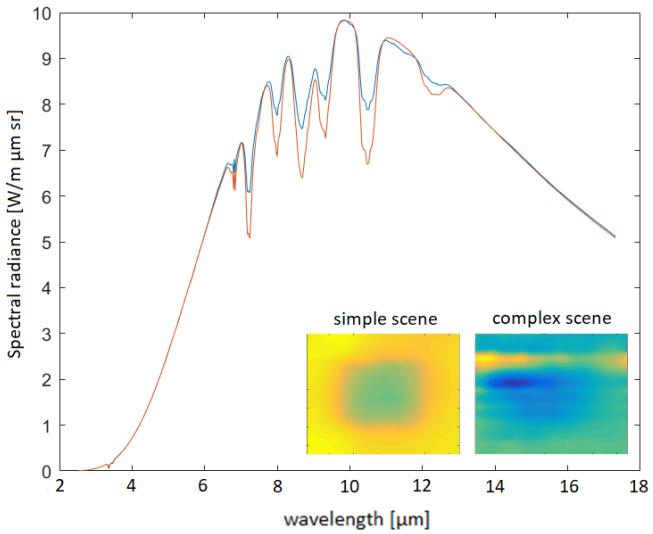

Fig. 4. Original (red) and reconstructed (blue) spectrum of the complex scene with black body radiation, reconstructed slices of datacube (small panel)

scene, however in the case of the complex scene, even 1% of noise level has a significant effect on the reconstruction.

V. Conclusion

In this article we provided an overview of IR hyperspectral imaging, with particular attention to compressed sensing. We also summarize results of our testing calculations, which evaluate the feasibility of using CASSI technique for hyperspectral imaging of IR absorption spectra of chemical compounds.

The central goal is to provide the possibility to rapidly capture spill of chemical substances, enabling both their localization and identification.

We came to conclusion that for complex scenes we are able to determine the type of chemical compound, nevertheless we do not achieve sufficient reconstruction quality. The CASSI method using binary masks cannot be therefore directly applied in this case. Further research will be focused on using, for example, several measurements of different random masks, rotation of spectral sweeping, improved mask design, improving of the reconstruction algorithm, etc.

Acknowledgment

We gratefully acknowledge the financial support of the Grant Agency of the Czech Republic (project 17-26284Y) and the Ministry of Education, Youth and Sports of the Czech Republic (Project NPU LO1206).

References

[1] Kitlle, D.S. et al. *Compressive Coded Aperture Spectral Imaging: An Introduction.* IEEE Signal Processing Magazine. 2014, 31(1), 105-115.

[2] Wagadarikar, A.; John, R.; Willett, R.; Brady, D. *Single disperser design for coded aperture snapshot spectral imaging,* Applied Optics. 2008, 47, B44-B51.

[3] Bioucas-Dias, J.M.; Figueiredo, M.A.T. *A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration.* IEEE Transactions on Image Processing. 2007, 16(12), 2992-3004.

[4] Junkermann, W.; Burger, J.M. *A New Portable Instrument for Continuous Measurement of Formaldehyde in Ambient Air*. Journal of Atmospheric and Oceanic Technology. 2006, 23(1), 38-45.

[5] Katrasnik, J.; Bürmen, M.; Pernus, F.; Likar, B. *Spectral characterization and calibration of AOTF spectrometers and hyper-spectral imaging systems*. Chemometrics and Intelligent Laboratory Systems. 2010, 101(1), 23-29.

[6] Hammond, S.V.; Clarke, F.C. *The handbook of vibrational spectroscopy, sampling techniques, microscopy*. In: Chalmers JM, Griffiths PR (eds), Wiley, Hoboken, 2002, 1405–1431.

[7] Carrasco, O. et al. *Hyperspectral imaging applied to medical diagnoses and food safety*, Proc. SPIE 5097, 2003, 215–221.

[8] Mahesh, S.; Jayas, D.S.; Paliwal, J.; White, N.D.G. *Hyperspectral imaging to classify and monitor quality of agricultural materials*, Journal of Stored Products Research, Volume 61, 2015, 17-26.

[9] Feng, Y.Z.; Sun, D.W. *Application of hyperspectral imaging in food safety inspection and control: a review*, Crit. Rev. Food Sci. Nutr. 2012, 52(11), 1039–1058.

[10] Schuler, R.L.; Kish, P.E.; Plese, C.A. *Preliminary observations on the ability of hyperspectral imaging to provide detection and visualization of bloodstain patterns on black fabrics*, J. Forensic Sci. 2012, 57(6), 1562–1569.

[11] Edelman, G.J. et al. *Hyperspectral imaging for non-contact analysis of forensic traces*, Forensic Sci. Int. 2012, 223(1–3), 28–39.

[12] Balas, C.; Epitropou, G.; Tsapras, A. et al. *Hyperspectral imaging and spectral classification for pigment identification and mapping in paintings by El Greco and his workshop*. Multimedia Tools and Applications. 2018, 77(8), 9737-9751.

[13] Gao, L.; Liang, J.; Li, C.; Wang, L.V. *Single-shot compressed ultrafast photography at one hundred billion frames per second*, Nature 516, 2014, 74–77.

[14] Stuart, B. *Infrared Spectroscopy: Fundamentals and Applications*; Wiley Press: London, UK, 2004, 29–36.

[15] Fernandez, C.; Guenther, B.D.; Gehm, M.E.; Brady, D.J.; Sullivan, M.E. *Longwave infrared (LWIR) coded aperture dispersive spectrometer*, Opt. Express, 2007, 15(9), 5742-5753.

[16] Russell, T.A.; McMackin, L.; Bridge, B.; Baraniuk, R. *Compressive hyperspectral sensor for LWIR gas detection*, Proc. SPIE 8365,2012, 83650C.

[17] Dupuis, J.R.; Kirby, M.; Cosofret, B.R. *Longwave infrared compressive hyperspectral imager*, Proc. SPIE 9482, Next-Generation Spectroscopic Technologies VIII, 2015, 94820Z.