Graph and Rank Regularized Matrix Recovery for Snapshot Spectral Image Demosaicing

Grigorios Tsagkatakis, Maarten Bloemen, Bert Geelen, Murali Jayapala, and Panagiotis Tsakalides

***Abstract*—Snapshot spectral imaging (SSI) is a cutting-edge technology for enabling the efficient acquisition of the spatio-spectral content of dynamic scenes using miniaturized platforms. To achieve this goal, SSI architectures associate each spatial pixel with a specific spectral band, thus introducing a critical trade-off between spatial and spectral resolutions. In this paper, we propose a computational approach for the recovery of high spatial and spectral resolution content from a single exposure or a small number of exposures. We formulate the problem in a novel framework of spectral measurement matrix completion and we develop an efficient low-rank and graph regularized method for SSI demosaicing. Furthermore, we extend state-of-the-art approaches by considering more realistic sampling paradigms that incorporate information related to the spectral profile associated with each pixel. In addition to reconstruction quality, we also investigate the impact of recovery on subsequent analysis tasks, such as classification using state-of-the-art convolutional neural networks. We experimentally validate the merits of the proposed recovery scheme using synthetically generated data from indoor and satellite observations and real data obtained with an Interuniversity MicroElectronics Center (IMEC) visible range SSI camera.**

***Index Terms*—Spectral sensing, sparse and low rank models, computational photography.**

I. INTRODUCTION

THE objective in Hyperspectral Imaging (HSI) is the acquisition of the spectral profile at each spatial location in the field-of-view of a camera, aiming at critical insights on a number of remote sensing applications, from climate change monitoring and medical diagnosis, to precision agriculture and food safety [1]. While the potential of HSI is clear, acquiring imagery with high *spectral*, *spatial*, and *temporal* resolution is a formidable task, due to the challenges associated with obtaining 4D measurements (two spatial, one spectral and one temporal), using 1D or 2D detector arrays. The discrepancy between the requested and the available dimensionality of detectors has sparked different philosophies with respect to HSI imaging architectures. Depending on the acquisition protocol, state-of-the-art HSI architectures can be broadly classified into spatial, spectral, and frame scanning approaches [2], [3]. Regardless of the approach, a shared characteristic among all methods is the need for repetitive scanning of the scene and the acquisition of a large number of exposures (frames) in order to capture the full spatio-spectral resolution observations at high temporal sampling rates. As a result, for specific architectures, different operating points are selected, achieving an "optimal" balance between different aspects of resolution [4].

To address the need for high resolution imaging, a new breed of HSI imaging architectures under the name of *Snapshot Spectral Imaging* (SSI) have been proposed, involving systems which sample the full 3D hyperspectral cube on every exposure [5]–[7]. Unlike traditional architectures, SSI approaches do not require moving parts like rotating mirrors, thus significantly reducing the complexity of the systems. Furthermore, operating on a single exposure basis means that they do not suffer from motion artifacts like blurring or ghosting allowing for the efficient imaging of dynamic scenes. As such, SSI technologies have a great application potential, since they can be employed in moving platforms like Earth Observation satellites and unmanned aerial vehicles, as well as in stationary setups like microscopy or even consumer photography, where they can achieve high frame rate imaging of dynamic scenes at significantly lower weight and volume, compared to traditional architectures.

One very successful SSI paradigm relies on the use of Spectrally Resolved Detector Arrays (SRDA) where each pixel is associated with a specific spectral region [4]. SRDA architectures allow the acquisition of a sub-sampled hyperspectral cube from a single exposure, making them ideal for scenarios where the camera or the scene are characterized by motion and dynamics. These architectures accomplish their objective by associating a single spectral band with each detector element. As a result, to achieve a high frame rate spectral imaging, SRDA architectures must sacrifice spatial resolution since only a small subset of spectral information is available for each spatial pixel, as illustrated in Fig. 1.

In this work, we propose a novel *Computational Spectral Imaging* system, which can exploit spatial, spectral and temporal correlations for the estimation of the full resolution HSI observations from a small number -even a single- of snapshot. To achieve this goal, we exploit two critical aspects of the ac-

---

Manuscript received May 29, 2018; revised October 1, 2018 and December 4, 2018; accepted December 4, 2018. Date of publication December 20, 2018; date of current version May 7, 2019. This work was supported in part by the PHySIS project (Contract 640174) within the H2020 Framework Program of the European Commission. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. James Theiler. *(Corresponding author: Grigorios Tsagkatakis.)*

G. Tsagkatakis is with the Institute of Computer Science, Foundation for Research and Technology—Hellas, 73100 Crete, Greece (e-mail: greg@ics.forth.gr).

M. Bloemen, B. Geelen, and M. Jayapala are with the Interuniversity MicroElectronics Center, B-3001 Leuven, Belgium (e-mail: maarten.bloemen@imec.be; bert.geelen@imec.be; murali.jayapala@imec.be).

P. Tsakalides is with the Institute of Computer Science, Foundation for Research and Technology—Hellas 73100, Crete, Greece, and also with the Computer Science Department, University of Crete, 73100 Crete, Greece (e-mail: tsakalid@ics.forth.gr).

Digital Object Identifier 10.1109/TCI.2018.2888989

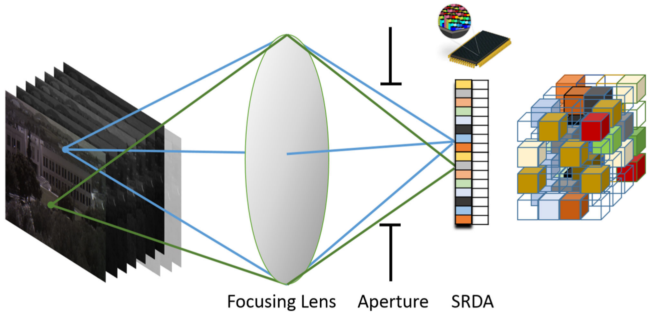


Fig. 1. An SRDA-based SSI camera consists of a lens focusing the image on the focal plane array such that each detector element captures light from a specific spectral region. The corresponding spatio-spectral observations are severely undersampled, inevitably leading to spatial resolution reduction.

quisition process. First, we leverage knowledge regarding the spectral characteristics associated with each filter/pixel, which can be easily obtained through spectral calibration. Second, we implicitly capitalize on the spatio-spectral correlations among the observations by formulating the spectral cube reconstruction as a graph and rank regularized nuclear norm minimization problem. To demonstrate the merits of our proposed approach, we provide extensive experimental results on three diverse datasets, namely, an indoor scenario using the Columbia CAVE collection, observations from the Sentinel 2 satellite, and measurements from two actual SSI cameras.

The rest of the paper is organized as follows: Section II provides an overview of the state-of-the-art in both spectral imaging architectures and associated recovery mechanisms. In Section III, we present an analysis of the ideal and the real sampling process involved in SSI, while in Section IV we introduce our novel graph and rank regularized matrix recovery technique. Section V describes the experimental setup while Sections VI, VII and VIII showcase the results on synthetic data under various conditions, comparisons to state-of-the-art methods, and evaluation on real data, respectively. Conclusions and future directions are discussed in Section X.

II. STATE-OF-THE-ART SPECTRAL DEMOSAICING

Demosaicing RGB color images is a well studied problem in the imaging community [8], [9], taking into account the presence of Bayer-structured filters for sampling the three primary colors. Within this framework, similar to the typical RGB Bayer filter pattern which samples at a 1:2:1 ratio, in the SRDA case the highest sampled spectral band is first interpolated and a guided filtering is subsequently applied on the rest of the bands for demosaicing [10]. Extensions of this work using residual interpolation have also been proposed [11], [12]. Advancing this process to a larger number of spectral bands and SRDA architectures, where no preferential spectral sampling is considered, requires overcoming formidable new challenges introduced by the reduction in sampling frequency.

Focusing on the case of SSI recovery, various approaches for the estimation of the missing spectral information have been proposed. In [13], a weighted bilinear interpolation method was developed involving a two-step approach where for each band a sparse raw image containing only the observed measurements is recovered by a convolution with a low-pass filter, appropriately normalized to account for the missing measurements. An extension of this method which considers the spectral correlation through the interpolation of sparse channel differences was proposed by Brauers and Aach in [13]. An extension, termed Iterative Spectral Difference (ItSD), was later presented by Mizutani *et al.* [14]. A recent work on generic SRDA image demosaicing was proposed by Mihoubi *et al.* [15] where the authors introduce PPID, a multispectral demosaicing approach using a pseudo-panchromatic image. The results reported suggest that the PPID method is able to achieve the best performance in diverse conditions, surpassing state-of-the-art methods. We should also note that the proposed methods considers the utilization of appropriately generated graphs for imposing smoothness on the reconstruction. The notion of graphs has also been considered for the problem of RGB image demosaicing in [16], were the authors impose a smoothness with respect to contours through a graph based minimization where graph weights are produced based on stencils. In our work, the graph is generated using the raw pixel values and the regularization also involves a low rank term.

A disruptive new concept in state-of-the-art signal processing refers to the sparse representations framework, where one seeks a sparsity-promoting decomposition of the input by exploiting available *training observations*. Particular applications of the sparse representations framework in spectral imaging involve denoising [17] and super-resolution [18], [19] among others. Recently, Degraux *et al.* presented a 3D inpainting method for demosaicing SSI measurements, where reconstruction is achieved by exploiting appropriate sparsifying dictionaries, wavelets for the spatial dimension and DCT for the spectral [20]. A closely related problem is the recovery of HSI from a small number of compressive (incoherent) measurements, based on the framework of Compressed Sensing where in [21], a low rank and sparse matrix formulation was proposed for the recovery from noisy compressed measurements.

In addition to the sparse representations framework, the concept of low rank representations has been applied for numerous HSI signal processing tasks, exploiting the sparsity in terms of singular values of appropriately formulated matrices. Low rank modeling has been employed in problems including restoration from noisy measurements [22], registration [23], simultaneous denoising and unimixing [24], as well as supervised learning problems such as classification [25], [26]. The low rank assumption is also critical in the context of Matrix Completion (MC) [27], [28] where recovery of severely sub-sampled matrices is possible provided a sufficient number of observations is available and the matrix is characterized by a lower rank compared to its ambient dimensions. The potential of MC for the recovery of missing hyperspectral measurements was demonstrated in [29] for the case of randomly missing entries. In this case, recovery of low-rank unfolded matrices was shown to offer a superior reconstruction performance when compared to tensor recovery. The low-rank nature of appropriately formulated matrices was later combined with a non-local approach for encoding spatial self-similarities [30], [31], while the notion of low rank has also been explored in higher dimensional structures for tasks such as light-field photography [32].

III. SNAPSHOT SPECTRAL IMAGE RECOVERY

Prevalent hyperspectral imaging architectures fall under two major categories, namely spatial and spectral scanning [1]–[3]. Spatial scanning methods can either utilize push-broom type sensors to obtain the spectral profile of a single spatial line, or 2D spatial scanning whisk-broom architectures to scan the scene along both rows and columns in order to collect the entire spectral profile of each pixel. Frame or spectral scanning designs are more recent approaches which employ 2D arrays of detectors and frequency selective tunable filters in order to capture full spatial resolution single-band images at each exposure. A shortcoming shared by the aforementioned methods is the repetition requirement for constructing the complete 3D hyperspectral datacube, resulting in slow acquisition time and motion artifacts.

Snapshot (or Simultaneous) Spectral Imaging (SSI) systems acquire the complete spatio-spectral cube from a single or a few exposures, i.e., during a single or a few integration periods. To achieve this goal, early approaches utilized apparatuses like lenslets, coherent fiber bundles, and mirror slicers, in order to manage the direct collection of hyperspectral images from traditional sensors during a single exposure [4], [5]. More recent approaches employ state-of-the-art signal processing and novel optical components to achieve this task [6].

A key design parameter in SSI is the type of spectral filters used for selecting the appropriate wavelengths. Unlike traditional approaches such as linear variable filters or liquid crystal tunable filters, Fabry-Pérot based SSI architectures, such as Interuniversity MicroElectronics Center (IMEC)'s snapshot imagers [6], [7], enable the design and fabrication of ranges of optimized filters at various wavelengths with flexible layouts and customizable spectral bandwidths. The integrated filter-detector approach reduces the number of discrete and bulky optical components, resulting in cost effective and extremely compact imaging systems. Since this integrated production method can utilize mass manufacturing facilities, the cost and time of production can be substantially reduced for volume applications. The result is a compact and fast hyperspectral imager made with low-cost CMOS process technology.

A straightforward approach for populating the full hyperspectral cube from a single snapshot involves grouping the appropriate number of pixels with a corresponding reduction in spatial resolution. For the 16 band case for example, groups of 4×4 pixels are collapsed into to a single "super-pixel," incurring a $\times 16$ or 93.75% reduction in spatial resolution, as shown in Fig. 2.

In this work, we develop a computational method for demosaicing SSI observations, where the problem is casted as a missing measurement recovery problem. The key steps of the recovery processes involve: (i) the generation of appropriate (subsampled) matrices from the acquired observations, called spectral measurements matrices, and (ii) the recovery of missing entries by exploiting the underlying structure. We explore the full range from single frame to multiple frames recovery, effectively covering the entire recovery spectrum shown in Fig. 2. To that end, we consider scenarios involving the acquisition of a limited number of spectral bands per spatial location. Examples of such scenarios include:

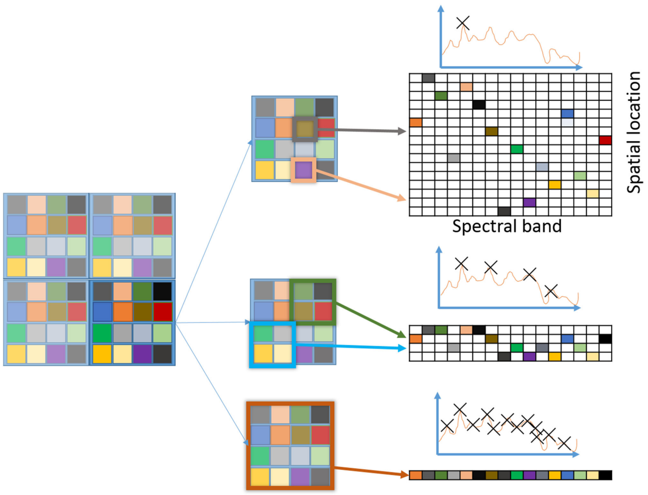


Fig. 2. A frame acquired by an SSI architecture employs a spectral filter pattern to associate a single spectral band to each spatial location - pixel (left), leading to significant sub-sampling (top right). A typical demosaicing approach involves binning together groups of pixels, known as "super-pixels," to produce lower spatial resolution observation with more (middle) or all (bottom right) spectral bands.

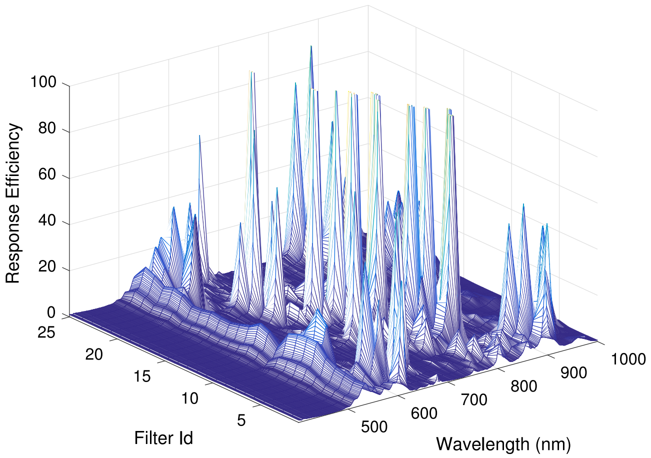


Fig. 3. Spectral response of the 25 spectral filters of the 5×5 IMEC SSI camera in the 400–1000 nm range.

- Naturally moving sensors, such as in the case of video acquisition by portable platforms, as well as sensors on-board orbiting satellites.
- MEMS-based sensors via the introduction of Digital Micromechanical Mirror devices for temporal multiplexing of acquisitions.
- Post-acquisition binning of multiple pixels, however, at a much higher resolution compared to the naive approach, suffering a dramatically smaller reduction in spatial resolution.

A. Spectral Measurements Matrix Sampling

In an ideal scenario, the Fabry-Pérot filters allow light from a very limited spectral range to propagate, blocking the light outside this range. However, both technical challenges and physical laws cause a behavior which diverges from the ideal. Fig. 3 illustrates a real spectral response profile, where one can easily observe that significant non-linearities occur during the sampling process. A particular phenomenon that is highlighted is that Fabry-Pérot filters introduce additional harmonics which must be removed through the introduction of appropriate cut-off filters.

Formally, a SSI camera acquires a two-dimensional image of $m_1 \times m_2$ pixels during each exposure, where m_1 and m_2 indicate the number of pixels in the horizontal and the vertical dimensions, respectively. Each one of the $m_1 \times m_2$ pixels is associated with a different spectral band. This matrix/image corresponds to a subsampling of the full resolution tensor $\mathcal{M} \in \mathbb{R}^{m_1 \times m_2 \times n}$, consisting of n spectral bands, obtained through the introduction of the pixel-level spectral filters. To process and recover the missing observations, we invoke an unfolding operator $\mathcal{T} : \mathbb{R}^{m_1 \times m_2 \times n} \to \mathbb{R}^{m \times n}$ which maps the full resolution hyperspectral tensor to the *spectral measurements matrix* $\mathbf{M}$, by collapsing the two spatial dimensions m_1 and m_2 into a single pixel index m. The unfolding operator $\mathcal{T}$, as well as the folding operator $\mathcal{T}^*$, serve to transform the familiar hyperspectral cube into a matrix which is the appropriate format for the proposed recovery mechanism, while information related to the spatial structure, which is lost during the folding process, is reintroduced via the proposed graph based recovery. A visualization of the proposed SSI recovery architecture is shown in Fig. 4.

Assuming the hyperspectral tensor is mapped to the spectral measurements matrix via $\mathcal{T}$, the process of sampling/filtering can be mathematically expressed though the sampling operator $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^k$ which is responsible for reporting only k out of the $m \times n$ observations. The sampling operation can also be defined through a set of sampling matrices $\mathcal{A}(\cdot) \equiv \{\mathbf{A}_i \in \mathbb{R}^{m \times n} | i = 1, \ldots, k\}$, in which the sampling process can be expressed in a more intuitive form as

$$\mathcal{A}(\mathbf{M}) = \begin{bmatrix} \mathrm{vec}(\mathbf{A}_1) \\ \mathrm{vec}(\mathbf{A}_2) \\ \vdots \\ \mathrm{vec}(\mathbf{A}_k) \end{bmatrix} \mathrm{vec}(\mathbf{M}) \tag{1}$$

where the vec$(\cdot)$ operator indicates the rearrangement of matrices in vectors [33]. Given an observation model $\mathbf{Y} = \mathcal{A}(\mathbf{M})$, each acquired observation $\mathbf{y}_k$, can thus are expressed as

$$\mathbf{y}_k = \langle \mathbf{A}_k, \mathbf{M} \rangle = \mathrm{trace}(\mathbf{A}_k^T \mathbf{M}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{j,i}^k m_{i,j} \tag{2}$$

We explore three cases of sampling operators $\mathcal{A}$ which can effectively encode knowledge regarding the acquisition process.

1) Binary Sampling: An important case of the sampling operator involves a *binary* sub-sampling operator given by

$$\mathcal{A}(\mathbf{M}) = \begin{cases} m_{i,j} & \text{if } (i,j) \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where Ω is the set of observed measurements. The sampling operation can also be defined through the application of a sequence of matrices as shown in (2), where each element is $\mathbf{A}_k[i,j] = 1$ only if $k = i$. When the sampling operator corresponds to an entry-wise binary subsampling, the recovery problem is known as Matrix Completion (MC) [27], [28].

2) Random Projections: The binary sub-sampling operator in (3) captures a large number of real-life applications, including ratings in recommendation engines and sensing in wireless sensor networks. However, it many cases, much more information can be encoded in each individual measurement. The case of *Random Projections* based sampling, an extension of the binary operator, involves sampling through sensing matrices where each element is independently drawn from a random distribution [34]. Gaussian Random Projections correspond to the case where $\mathcal{A}$ is a Gaussian measurement ensemble, i.e., each row $\mathcal{A}_i, 1 \le i \le k$ encoded as in (1) contains elements drawn i.i.d from $\mathcal{N}(0, 1/m)$ entries, where k is the number of measurements. While the case of Gaussian Random Projections offers concrete theoretical guarantees, there is a limited number of scenarios which can support such sampling patterns. In this work, we experimentally explore the case of sampling matrices whose elements are drawn from a uniform distribution, i.e., $\mathbf{A}_k[i,j] \sim \mathcal{U}[0,1]$ subject to the normalization constrain $\|\mathbf{A}_k\|_F = 1$ [33].

3) Spectral Filter Profile Sampling: In addition to the cases of random sampling operators, we investigate the more relevant case of an input signal modulated by the spectral profile associated with a given spectral filter (and corresponding spatial location). Fig. 3 showcases the complex spectral response associated with different filters. One can easily observe the significant deviation of the responses from the ideal "delta" function, both in terms of the large amount of energy which is spilled into neighboring bands and the appearance of additional harmonics.

During the calibration process, the calibration matrix $\mathbf{C} \in \mathbb{R}^{n \times p}$ can be constructed by measuring the response of the filter associated with band n given a source emitting light at wavelength p. Given the calibration matrix $\mathbf{C}$, each measurement obtained via the sampling operator, can be defined in terms of individual matrices such that $\mathbf{A}_k[i,j] \equiv \mathbf{C}[i,j]$. Each of the k acquired samples from spatial location i can thus be expressed as

$$y_{i,k} = \sum_j c_{i,j} m_{i,j} \tag{4}$$

IV. SPECTRAL MEASUREMENTS MATRIX RECOVERY

We model the SSI demosaicing problem as the regularized recovery of the spectral measurements matrix $\mathbf{X}$ from a limited set of measurements $\mathcal{A}(\mathbf{M}) \in \mathbb{R}^k$, expressed as

$$\begin{aligned} \min_{\mathbf{X}} \quad & \mathcal{R}(\mathbf{X}) \\ \text{subject to} \quad & \mathcal{A}(\mathbf{M}) = \mathcal{A}(\mathbf{X}) \end{aligned} \tag{5}$$

where $\mathcal{R}$ is a regularization term and $\mathcal{A}$ is the sampling operator.

In this work, we explore how appropriately selected regularization and sampling terms can lead to high quality reconstruction. On the one hand, we extend the state-of-the-art approach introduced in Matrix Completion by investigating different forms of sampling operations, beyond the case of

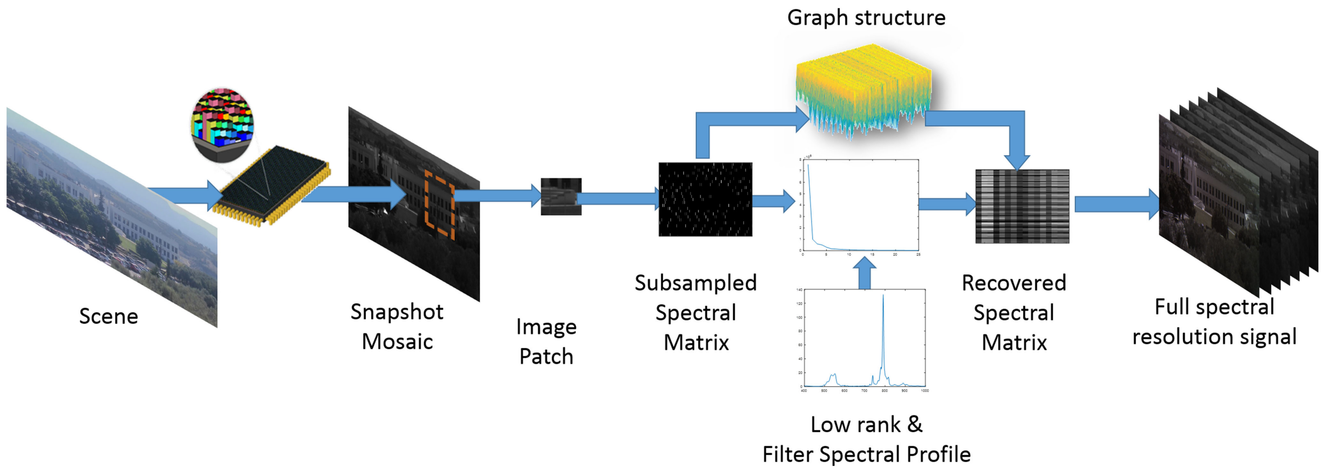


Fig. 4. Overview of acquisition and recovery process. The full spectral profile at a given location undergoes a modulation with the spectral profile of the filter associated with each pixel, producing a two-dimensional (2-D) spectral snapshot of the scene. Spectral mosaic patches are extracted and transformed to undersampled measurement matrices with a recovered spectral matrix in order to produce the full spectral content of the scene.

binary sampling. On the other hand, we propose the introduction of two types of regularization, one encoding the low rank structure of the measurements matrix and the other enforcing a spatial smoothness that typically characterizes image data.

A. Rank Regularization

According to the low rank matrix recovery framework, provided a sufficient number of randomly selected entries is available, the missing entries of an undersampled matrix $\mathbf{M}$ can be perfectly recovered by solving a constrained rank minimization problem given by

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \text{rank}(\mathbf{X}) \\ \text{subject to} \quad & \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{M}) \end{aligned} \tag{6}$$

Unfortunately, direct rank minimization is an NP-hard problem and therefore cannot be applied in realistic settings. However, a relaxation of the above problem where the rank constraint is replaced with the computationally tractable nuclear norm, which represents the convex envelope of the rank, was shown that is able to achieve the same reconstruction [27]. According to Matrix Completion (MC) [28], a significantly undersampled matrix whose only $k \ll (m \times n)$ elements are available, can be "perfectly" recovered provided the matrix is characterized by a sufficiently small rank and enough randomly selected entries of the matrix are observed.

Formally, one can recover an accurate approximation $\mathbf{X}$ of the matrix $\mathbf{M}$ from a small number of entries by solving the minimization problem

$$\underset{\mathbf{X}}{\text{minimize}} \ \|\mathbf{X}\|_* + \|\mathcal{A}(\mathbf{M}) - \mathcal{A}(\mathbf{X})\|_F \tag{7}$$

To allow the recovery of a sub-sampled matrix, the sampling process must preserve a sufficient amount of signal characteristics, such as pair-wise distances. For the recovery of low-rank matrices, this requirement is formaly expressed through the *rank-restricted isometry property* [34], which for a matrix $\mathbf{M}$ of rank at most r is defined as the smallest δ_r such that

$$(1 - \delta_r)\|\mathbf{M}\|_F^2 \leq \|\mathcal{A}(\mathbf{M})\|_F^2 \leq (1 + \delta_r)\|\mathbf{M}\|_F^2 \tag{8}$$

A consequence of this bound is that for all matrices of rank at most r, the images of these matrices after projection are unique. As a result, the number of observations that are required for the perfect reconstruction of a matrix is a function of the rank and the sampling process. For Matrix Completion, where the sampling operator is binary, and assuming the incoherence of the sampling pattern, a notion directly related to the matrix RIP and the uniformity of the sampling pattern, is minimum and equal to 1, then the recovery is possible from

$$k \geq r(m + n)log^2 m \tag{9}$$

measurements, where $m > n$. For the Random Projections based measurements sensing matrix, the matrix RIP property holds with high probability provided the elements of the sampling operator are i.i.d. Gaussian observations in which case, a total of

$$k \geq r(n + m) \tag{10}$$

observations have to be acquired for the recovery of a rank r matrix of dimensions $n \times m$.

In order to acquire some insight of the impact of these bound on the problem at hand, we consider the case of recovering measurements matrices of spatial dimensions 5×5 pixels along 16 and 25 spectral bands, which are mapped to 25×16 and 25×25 matrices, respectively. We assume that the rank of these matrices is 2, a reasonable assumption based on our experiments. According to this bound, for the case of binary measurements, at least 0.9634 observations per pixel are necessary for a 16 band spectral cube and 0.6644 observations per pixel for a 25 band spectral cube. For the Random Projections case, the bound becomes 0.1450 for 16 bands and 0.1 for 25 bands. Hence, in theory for this setup, recovery is possible even from a single explore, i.e. one observed band per pixel.

B. Graph Regularization

In addition to the low rank regularization, in this work we also consider an additional regularization term responsible for encoding structural information in the form of a graph. The motivation behind this choice is that for a large number of problems related to high dimensional observations like hyperspectral image sequences, the transformation from a *3D* tensor to a *2D* matrix fails to capture correlations that reside in high dimensional spaces. To address this issue, we assume that different spatially-constrained regions are related through an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$ composed of a set of $\mathcal{V}$ vertices, connected through a set of edges $\mathcal{E}$, each one associated with a non-negative weight encoded in the weight matrix $\mathbf{W}$.

In our setup, vertices correspond to regions of limited spatial resolution, 5×5 pixels, while the edges encode the correlations between different spatial locations. Weights, and the associated matrices derived based on them, play a key role in the encoding of correlations and can take a number of forms, depending on the employed distance metric. A prevalent example is the case where the weight matrix $\mathbf{W}$ is constructed such that each element encodes a Gaussian weighted distance among measurements

$$w_{ij} = \exp\left(-\frac{\|\mathbf{M}_i - \mathbf{M}_j\|_2^2}{\sigma^2}\right) \tag{11}$$

where $\mathbf{M}_i$ and $\mathbf{M}_j$ are columns of the measurements matrix $\mathbf{M}$ corresponding to spatial patches, σ is a weight normalization factor and the l_2 norm corresponds to the Euclidean distance in the spectral-induced space. One can observe that in this case, the distance is measured in terms of the spectral dimension, while one could also encode information related to spatio-spectral features.

The (unnormalized) Laplacian matrix $\mathbf{L}_0$ is defined as $\mathbf{L}_0 = \mathbf{D} - \mathbf{W}$, where the degree matrix $\mathbf{D}$ is defined as the diagonal matrix whose elements are given by $d_{ii} = \sum_{l=1}^{j} w_{ij}$, while the normalized graph Laplacian is defined as

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-\frac{1}{2}}. \tag{12}$$

The graph Laplacian matrix is an approximation to the Laplace-Beltrami operator and is responsible for encoding the underlying manifold structure of the data such to spatial proximity. In order to enforce the graph smoothness constraint, we introduce the associated regularization term given by

$$\|\nabla_{\mathcal{G}}\mathbf{X}\|_F = tr(\mathbf{X}\mathbf{L}\mathbf{X}^T). \tag{13}$$

The regularization in terms of the Laplacian matrix can also be interpreted as a regularization imposed on the eigenvalues of the Graph Fourier Transform representation [35].

The motivation for the graph regularization in the particular problem of snapshot spectral image recovery is the need for imposing smoothness of the final reconstruction. More specifically, while the patch based low rank estimation can produce accurate results, it may introduce artifacts between adjacent spatial locations.

C. Graph and Rank Regularized Matrix Recovery (GRMR)

Given a set of observations $\mathbf{M}$, the proposed Graph and low Rank regularized Matrix Recovery (GRMR) is formulated in an unconstrained minimization form as

$$\underset{\mathbf{X}}{\text{minimize}}\ \|\mathbf{X}\|_* + \|\mathcal{A}(\mathbf{X}) - \mathbf{y}\|_2^2 + \beta tr(\mathbf{X}\mathbf{L}\mathbf{X}^T) \tag{14}$$

where $\mathbf{y} = \mathcal{A}(\mathbf{M})$. The GRMR objective function in (14) involves seeking a matrix $\mathbf{X}$ which (i) is low rank, (ii) is consistent with respect to the observations, and (iii) preserves the graph structure associated with the observations. In this work, we consider the graph which encodes the similarities between different region patches, which can be found from an initial estimation, e.g., by first applying a linear interpolation or estimating the distances in low spatial resolution.

In this work, we adopt the proximal splitting method, by defining two functions $f(\mathbf{X})$ and $g(\mathbf{X})$, such that the minimization in (14) is equivalently expressed as

$$\underset{\mathbf{X}}{\text{minimize}}\ f(\mathbf{X}) + g(\mathbf{X}) \tag{15}$$

where

$$f(\mathbf{X}) = \|\mathbf{y} - \mathcal{A}(\mathbf{X})\|_2^2 + \beta tr(\mathbf{X}\mathbf{L}\mathbf{X}^T) \tag{16}$$

and

$$g(\mathbf{X}) = \|\mathbf{X}\|_* \tag{17}$$

The solution to the minimization in (15) can be obtained by the proximal gradient method [36], an iterative approach where each iteration is given by

$$\mathbf{X}^{t+1} = \mathbf{prox}_{\lambda^t g}(\mathbf{X}^t - \lambda^t \nabla f(\mathbf{X}^t)) \tag{18}$$

For the GRMR objective, the gradient of $f(\mathbf{X}^t)$ in (16) is given by the sum of the gradient with respect to the l_2 norm and the trace

$$\nabla f(\mathbf{X}^t) = \mathcal{A}^*(\mathcal{A}(\mathbf{X}^t) - \mathbf{y}) + \beta \mathbf{X}^t \mathbf{L}^T \tag{19}$$

where the adjoint of the sampling operator is defined as

$$\mathcal{A}^*(\mathbf{w}) = \sum_{i=1}^{K} w_i \mathcal{A}_i. \tag{20}$$

The proximal operator of the nuclear norm in (17) is given by a thresholding operator applied on the singular values of matrix $\mathbf{X}$. Two versions of thresholding operators, a soft and a hard, can be used. The soft-thresholding operator is given by

$$\mathcal{D}_\tau(\mathbf{X}) = \mathbf{U}\mathcal{D}_\tau(\mathbf{\Sigma})\mathbf{V}^T, \tilde{\ }\mathcal{D}_\tau(\mathbf{\Sigma}) = \text{diag}(\{\sigma_i - \tau\}_+) \tag{21}$$

where the $\{_+\}$ designates that only the non-negative elements are preserved while the rest are set to zero. While the soft-thresholding has been extensively utilized, the identification of the optimal threshold is not trivial, while in many cases, different thresholds need to be selected depending on the particular signal characteristics (distribution of singular values). In this work, we employ the hard-thresholding operator

$$\mathcal{H}_\tau(\mathbf{X}) = \mathbf{U}\mathcal{H}_\tau(\mathbf{\Sigma})\mathbf{V}^T,\ \mathcal{H}_\tau(\mathbf{\Sigma}) = \begin{cases} \sigma_i \quad \text{if} \\ ;\ i \leq r \\ 0 \quad \text{otherwise} \end{cases} \tag{22}$$

such that only the r top singular values are preserved, giving rise to pre-determined rank matrices.

Given the proximal operator in (18), we propose a proximal gradient based optimization of the graph and rank regularized matrix recovery. The proposed GRMR scheme is an iterative process, where iteration t is described by the following operations

$$\begin{cases} \mathbf{Z^t} = \mathbf{X^t} - \mu_\mathbf{t}\mathcal{A}^*(\mathcal{A}(\mathbf{X^t}) + \mathbf{y}) - \beta\mathbf{X^t}\mathbf{L^T} \\ \mathbf{X}^{t+1} = \mathcal{H}_\tau(\mathbf{Z}^t) \end{cases} \tag{23}$$

A key parameter of the method involves the value of the stepsize μ_t, where both fixed and adaptive values have been considered, such that the algorithm convergences into a global minimum. In this work, the gradient involves both the nuclear norm and the graph smoothness. For the case of the graph smoothness, we select a fixed step size β which is defined by the weight associated with the trace in (14), while for the case of nuclear norm, we consider the adaptive thershold μ_t proposed by the normalized Iterative Hard Thresholding framework [37] For each iteration t, the stepsize associated with the rank constraint is given by

$$\mu_t = \frac{\|\mathbf{P}_U^t\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\mathbf{X}^t)\|}{\|\mathcal{A}(\mathbf{P}_U^t\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\mathbf{X}^t))\|} \tag{24}$$

where the projection operation $\mathbf{P}_U^t = \mathbf{U}_j\mathbf{U}_j^T$ where $\mathbf{U}_j$ is the matrix of the topmost j left singular vectors.

Regarding the convergence of the method, $\mathbf{X}^t$ will converge to the set of optimal solutions, given that f has a minimum as the sum of two convex functions [38]. From a computational complexity point of view, most of the processing time is allocated to the SVD for the truncation of the singular values, whose complexity is in the order of $\mathcal{O}(r^2m)$, where $m < n$. A key benefit of the proposed method is that while state-of-the-art methods have to be repeated for each exposure of the same spatial region, the proposed scheme follows a single execution pipeline for any number of observations per pixel. The outline of the algorithm is shown in Algorithm 1.

We note that graph regularization was investigated in [39] where the objective to identify a low rank regression matrix which could accurately model the input signals, i.e., estimate a low rank matrix $\mathbf{Z}$ such that $\mathbf{M} = \mathbf{MZ}$, while imposing graph smoothness containts in PCA analysis has also been explored [40]. In both cases, the objective is to obtain a descriptive representation, in contrast to the proposed framework where one seeks to estimate missing or sub-sampled observations. The novelty of the graph regularized low rank matrix recovery proposed in this work from a theoretical standpoint, lays in the fact that we employ a different optimization scheme (proximal gradient versus Alternating Direction Method of Multipliers), as well as the fact that we broaden the potential of the proposed method by considering three different sampling strategies, instead of the classical binary sampling based Matrix Completion, e.g., [22], [37].

Algorithm 1: Graph and Rank Regularized Matrix Recovery (GRMR).

```
Input: The sampled measurements matrix $\mathbf{M}$,
    the regularization parameter $\beta$ and rank estimate $k$,
    the maximum number of iterations $limit$ or variation
less than $\delta$ between successive iterations.
Output: The estimated spectral measurements matrix $\mathbf{X}$.
initialization
    Initialize $\mathbf{X}^{(0)}$ from interpolation, $t = 0$
    Calculate the Degree and Weight matrix from $\mathbf{X}^{(0)}$ and
    estimate the Laplacian matrix $\mathbf{L}$ according to Eq. (12)
while $t < limit$ or $\|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_2 \leq \delta$ do
    Estimate update parameter $\mu_t$ according to Eq. (24)
    Estimate intermediate variable $\mathbf{Z}^{(t)}$
        $\mathbf{Z}^{(t)} = \mathbf{X}^{(t)} - \mu_t(\mathcal{A}^*(\mathcal{A}(\mathbf{X}^{(t)}) - \mathbf{y})) - \beta\mathbf{X}^{(t)}\mathbf{L}^T$
    Obtain the updated estimate of matrix $\mathbf{X}^{(k+1)}$:
        $(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}) = \mathrm{SVD}(\mathbf{Z}^{(t)})$
        $\mathbf{X}^{(t+1)} = \mathbf{U}\mathcal{H}_k(\mathbf{\Sigma})\mathbf{V}^T$
    set $t \longleftarrow t + 1$
end while
```

V. Experimental Setup

To validate the performance of the proposed demosaicing approach, we conducted a thorough set of experiments and evaluated the performance on three datasets under different conditions and sensing scenarios. The CAVE multispectral image database contains benchmark datasets of 512×512 pixel image sequences acquired in the 400—700 nm range with 10 nm steps [41]. In order to provide a representative analysis, in our experiments, we assume that the hyperspectral imagery in CAVE is acquired using a SSI camera system, equipped with either a 4×4 or a 5×5 spectral filter pattern, which suggests that for each case we randomly selected either 16 or 25 spectral bands out of the 30 and report the average performance over three independent trials.

In the experiments, we assume three types of sampling operators, idealistic (or binary), random and realistic (or filter based). In the first case, each filter associated with each detector element corresponds to an ideal filter with perfect cut-off characteristics which allow light from a single wavelength (the central one) to be captured by the detector element. In the realistic case, we consider the real spectral profiles derived from a spectral calibration process, and more specifically, we employ either the spectral profiles from an IMEC equipped camera or the publicly available spectral calibration data from Sentinel 2.

In Section VI we considered a subset of four examples from the CAVE dataset, namely the $chart\ and\ stuffed\ toy$, $beads$,

$flowers$, and $stuffed\,toys$ sequences and report the average performance, while in Section VI-C where an in-depth comparison with state-of-the-art is reported, we consider all 32 hyperspectral sequences. Furthermore, we investigated the recovery of spectral observations acquired by the Sentinel 2 HSI imager in the spectral range of 443–2190 nm, and consider a subset of 9 bands (B2-B8,B11, and B12) out of the 13, for which high (10 and 20 m) spatial resolution imagery is acquired. Last, we considered real snapshot spectral observations acquired by two IMEC-based SSI imagers, namely a 4×4 sensor observing the scene in the 470–620 nm range and a 5×5 sensor recording in the 600–840 nm range. Each mosaic frame has a $2K \times 1K$ pixels resolution. Reconstruction quality is measured in terms of PSNR (reported in dBs) and the Spectral Angle Map (SAM) metric, while the observed values were scaled to 8 bits. In the experiments, the threshold in (22) is set to 2, while $\sigma = 10$ for the weights in (11).[1]

VI. RECOVERY ON THE CAVE DATASET

In the first set of experiments, we assume that hyperspectral images from the CAVE dataset correspond to the ground-truth, and explore the recovery capabilities of the proposed and state-of-the-art method in when simulating a snapshot spectral image architecture. We embrark in this simulation process in order to provide quantitative, in addition to qualitative, performance metrics.

Since the CAVE dataset contains a single static image for each scene, a process for simulating sequences of multiple distinct frames for each scene is necessary. To achieve this goal, we simulate sequences of spectral snapshot frames where each pixel in every frame is associated with a different spectral band. This simulation protocol assumes that the camera is moved by one pixel in a specific (randomly selected) horizontal direction between consecutive acquisitions/frames. By spatially subsampling the full resolution spectral cube to create instances where the imager is observing the same spatial location over a limited number of distinct spectral bands, we were able to simulate the acquisition of multiple frames.

The particular simulation protocol is of interest due to its numerous practical applications in observation scenarios where the imager is under motion. For example, when the imaging sensor is spaceborne, the acquisition of multiple observations over the same location is naturally considered in a push-broom type of sensing. The results in this section indicate that the same capabilities can also be achieved using SSI camera architectures. The approach is also relevant to ground based video acquisition and the case of multiframe super-resolution.

A. Impact of the Sampling Process

We first turn our attention on the impact of the sampling operation on the recovery of the spectral measurements matrices, where the sampling operator $\mathcal{A}$ corresponds to either the ideal binary case or the realistic filter profile case. In this experiment, we consider three scenarios with respect to sampling and reconstruction, which correspond to (i) the case where both the sampling and the recovery processes employ the ideal operator, (ii) the case where both sampling and recovery adhere to the realistic operation, and (iii) the miss-match case where sampling is performed with the realistic operator while recovery assumes the ideal. Fig. 5 reports the reconstruction quality for a 16 spectral bands cube as a function of the number of iterations in the rank regularized spectral matrix recovery process, using 2 and 4 frames, i.e., observations per spatial location.

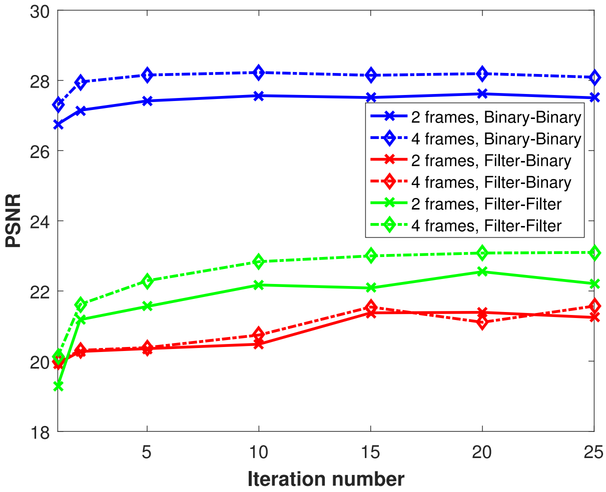


Fig. 5. Reconstruction quality with respect to iteration for the recovery of 16 bands from 2 and 4 frames (observations per pixel) under different sampling/reconstruction models.

Fig. 5 shows that more iterations lead to higher reconstruction quality, while a steady performance is achieved from a relatively small number of iterations. As expected, comparing the curves corresponding to 2 and 4 observations per pixel, more observations lead to better performance. Some more interesting results are that:

- Sampling and recovery with the ideal operator leads to the best performance for both 2 and 4 frames. Theoretically, this can be justified by the fact that random binary matrices achieve the optimal incoherence and are thus capable of fully preserving and exploiting the low rank nature of the measurements matrices.
- For both 2 and 4 frames, the combination of realistic sampling with ideal reconstruction leads to the worse performance. This is a key observation since this scenario corresponds to the assumption made by the state-of-the-art methods employing Matrix Completion for recovery [29], [31].
- When both sampling and recovery employ the realistic sampling operators, the performance is better than the case of miss-match, yet not as good as the case of ideal sampling, although for a moderate number of iterations, similar recovery capabilities are demonstrated.

These observations validate a core notion of this paper, that real-life conditions in signal acquisition can have a profound effect in reconstruction performance. Note that in this case,

---

[1] Code and data are [Online]. Available: https://github.com/gtsagkatakis/Snapshot_Spectral_Image_demosaicing

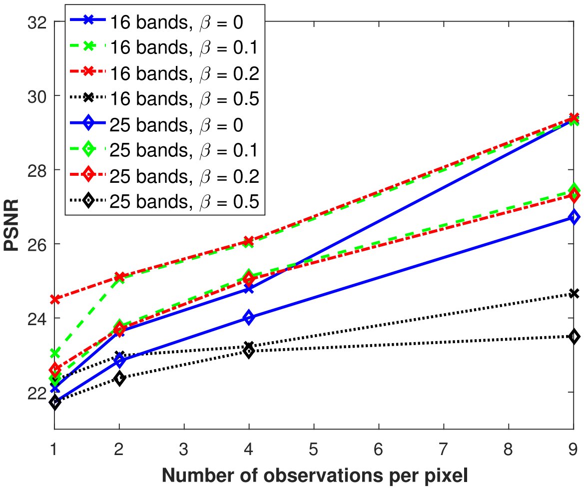


Fig. 6. Reconstruction quality for 16 and 25 band observations with respect to the number of observations per pixel, with different values of β in graph regularization.

TABLE I
IMPACT OF GRAPH WEIGHTING NORMALIZATION FACTOR σ IN RECONSTRUCTION QUALITY FOR CASES OF 16 AND 25 SPECTRAL BANDS

	0	0.001	0.01	0.1	1	10	100
16	25.8	25.8	25.8	25.8	25.9	26.5	26.5
25	24.8	24.9	24.9	25.1	25.4	25.5	25.4

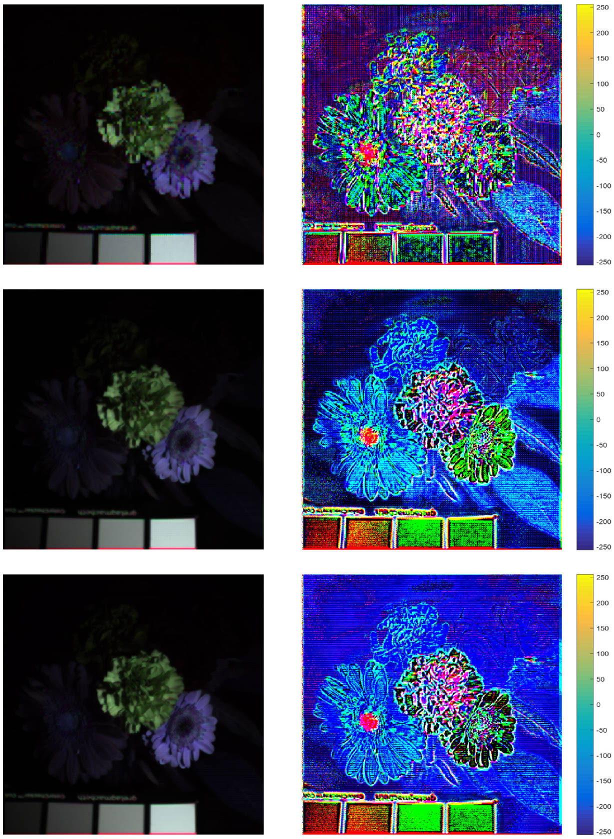


Fig. 7. Examples of 16 bands spectral cubes (left column) and difference images with ground truth (right column) from 1 (top row), 2 (middle row), and 4 (bottom row) frames.

graph regularization was not considered ($\beta = 0$) so that no bias is introduced to the results.

B. Impact of the Number of Observations and Graph Regularization

Fig. 6 shows the PSNR achieved by MC and GRMR as a function of the number of observations per pixel for a 4 × 4 (16-band) and a 5 × 5 (25-band) SSI architecture, respectively. In this set of experiments, we assumed realistic filter sampling and recovery. Furthermore, we explored the impact of graph-regularization during recovery, which is represented by the value of the β parameter in (14) corresponding to the case of the MC ($\beta = 0$) and the GRMR ($\beta = [0.1.0.2.0.5)$ approaches. Note that in this set of experiments, the true filter response is used as the sampling operator.

The experimental results shown in Fig. 6 confirm that increasing the number of observations leads to better reconstruction quality. Furthermore, the results indicate that higher spectral resolution architectures require a more observations for achieving the same reconstruction quality. Regarding the impact of graph regularization, the results demonstrate that significant gains are obtained when the spatial correlations are taken into account through graph modeling, especially in the limited observations regime. However, an optimal value of 0.2 must be selected in order to balance the smoothness with the preservation of high frequency components, while higher values lead to worse performance.

Last, we also investigated the impact of the value of σ and report the impact in terms of performance in Table I for 16 and 25 spectral band recovery cases, again using a realistic sampling operator from a single exposure. The results demonstrate that an optimal value of 10 leads to the best performance, however, selection of a different value has a minimal impact in term of reconstruction quality.

To further demonstrate the capabilities of GRMR in SSI recovery, Figs. 7 and 8 presents 3-color (bands 1, 6, 10 and bands 3, 7 and 12 respectively) full resolution imagse and the GRMR recovered image from 1, 2, and 4 observations for the cases of 16 and 25 bands. This example provides a visual illustration of the behavior observed in Fig. 6, where one can clearly see that by increasing the number of observations, higher quality reconstruction is possible, while even 4 observations are sufficient for producing an artifact-free reconstruction of 16 band observations. For the case when a single observation is available, while key structural content is present, artifacts related to high-frequency components, e.g., flower pedals, are manifested. The quality however quickly reaches a reasonable level even from 2 observations while with 4 observations, very high quality recovery is achieved.

C. Comparison With State-of-the-Art

In order to validate the merits of the proposed approach, we provide comparative results with state-of-the-art methods, including the Weighted Bilinear Interpolation (WB) [13], the Binary Tree-Based Edge-Sensing (BTES) [42], the Multispectral Demosaicing using Pseudo-Panchromatic Image (PPID) [15], and the Iterative Spectral Difference (ItSD)

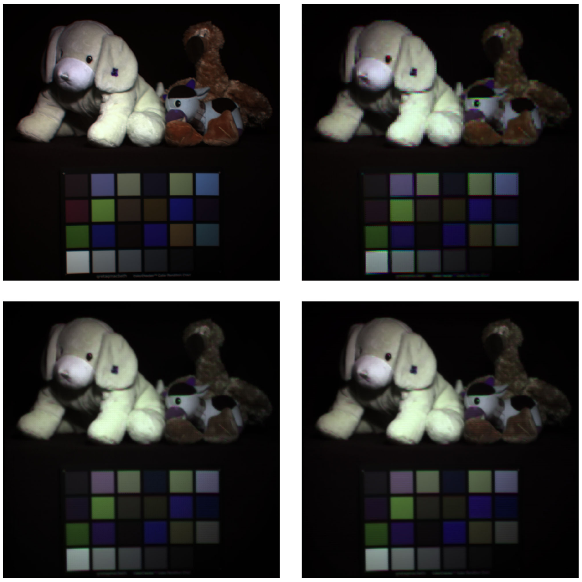


Fig. 8. Example of 25 bands spectral cube (top left) and recovery from 1 (top right), 2 (bottom left), and 4 (bottom right) frames.

TABLE II
PSNR AND SAM OF RECOVERED **16** BANDS FROM **2** FRAMES ON THE CAVE DATASET

		Sampling pattern		
		Idealistic	RPs	Filter profile
Method	GRMR	34.4 (0.11)	**28.9 (0.45)**	**30.6 (0.15)**
	BTES	32.5 (0.12)	26.2 (0.50)	25.8 (0.60)
	WB	33.6 (0.11)	26.6 (0.51)	26.1 (0.61)
	PPID	**35.1 (0.10)**	27.2 (0.50)	26.6 (0.61)
	ItSD	27.1 (0.49)	27.1 (0.49)	27.1 (0.49)

TABLE III
PSNR AND SAM OF RECOVERED **16** BANDS FROM **4** FRAMES ON THE CAVE DATASET

		Sampling pattern		
		Idealistic	RPs	Filter profile
Method	GRMR	**36.1 (0.10)**	**29.5 (0.44)**	**32.5 (0.30)**
	BTES	32.7 (0.11)	26.3 (0.50)	25.9 (0.60)
	WB	34.1 (0.09)	26.8 (0.50)	26.3 (0.61)
	PPID	35.7 (0.09)	27.3 (0.50)	26.8 (0.61)
	ItSD	27.1 (0.49)	27.0 (0.49)	27.1 (0.49)

TABLE IV
PSNR AND SAM OF RECOVERED **25** BANDS FROM **2** FRAMES ON THE CAVE DATASET

		Sampling pattern		
		Idealistic	RPs	Filter profile
Method	GRMR	32.8 (0.14)	**26.0 (0.50)**	**29.3 (0.38)**
	BTES	31.9 (0.14)	24.5 (0.56)	25.8 (0.59)
	WB	32.2 (0.13)	24.6 (0.56)	25.9 (0.59)
	PPID	**33.5 (0.12)**	24.9 (0.56)	26.5 (0.59)
	ItSD	24.9 (0.54)	24.9 (0.55)	25.0 (0.53)

TABLE V
PSNR AND SAM OF RECOVERED **25** BANDS FROM **4** FRAMES ON THE CAVE DATASET

		Sampling pattern		
		Idealistic	RPs	Filter profile
Method	GRMR	**34.6 (0.12)**	**26.4 (0.50)**	**31.0 (0.34)**
	BTES	32.4 (0.12)	24.6 (0.56)	26.0 (0.59)
	WB	32.8 (0.12)	24.7 (0.56)	26.2 (0.59)
	PPID	34.2 (0.11)	25.1 (0.56)	26.7 (0.59)
	ItSD	24.9 (0.54)	25.0 (0.55)	25.0 (0.53)

[14]. In general, WB is introduced as a baseline method in terms of performance, while BTES, PPID, and ItSD achieve very high performance under certain scenarios. Since state-of-the-art methods have been developed for the case of recovery from a single exposure, the results correspond the the error between the ground-truth measurement and the averaged reconstruction for each method. Note that in this set of experiments, we consider all 30 images available in the CAVE dataset.

The average PSNR and SAM of the reconstructed spectral cubes achieved by the various methods, averaged over the four CAVE examples, are presented in Table II and Table III for 2 and 4 frames (observations) of 16 band spectral cubes, while Table IV reports the recovery of 25 band cubes from 4 frames. For each case, we report the reconstruction quality with respect to the specific sampling operators, focusing on (i) the idealistic single-band observation binary, (ii) the Random Projections (RPs) based sampling, and (iii) the realistic Filter response profile based sampling cases.

A number of key observations can be made from the results presented in the tables:

- When a small number of frames, e.g., 2 out of the 16, is available, a high quality reconstruction is achieved by the PPID method, when the sampling operator corresponds to the idealistic binary case. For a moderate numbers of frames, e.g., 4 out of the 25, acquired by an idealistic sampling process, the proposed GRMR method achieves comparable performance with the PPID. The rest of the methods under investigation achieve lower reconstruction quality for this sampling case, typically WB, a baseline method reaching the same quality as the BTES, while ItSD typically offers poor performance.
- For the same number of observations, when a non-ideal sampling operator like Random Projections is considered, state-of-the-art methods suffer a major performance degradation. PPID of instance suffers an average of more than 8 dB reduction for all combinations of frames/spectral bands. Similarly, other methods achieve on average 24 dB reconstruction quality for the 16 bands case without any noticeable increase when more frames are acquired. On the contrary, the GRMR method offers a more reliable behavior, achieving higher quality while exploiting all available frames.
- For the realistic case where specific filter profiles are considered, the proposed GRMR approach is able to attain

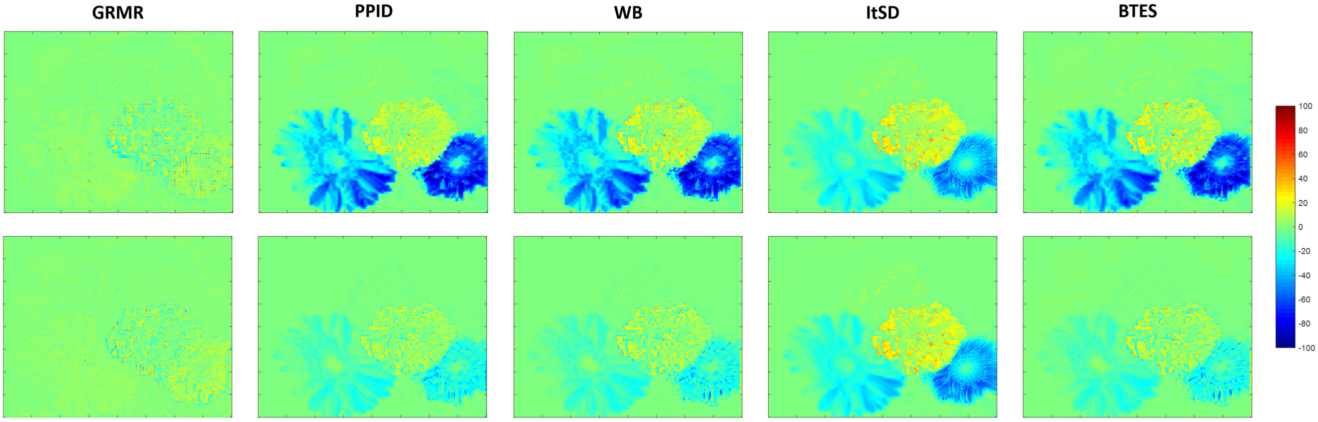


Fig. 9. Relative error for the recovery of 25 band cube from 1 (top row) and 4 (bottom row) exposures for the proposed and state-of-the-art methods.

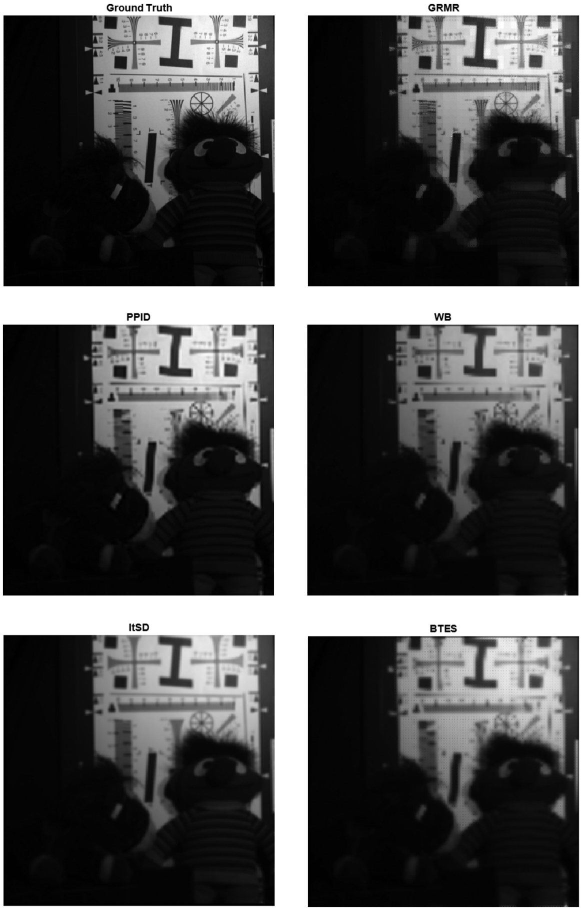


Fig. 10. Exemplary band reconstructions from two SSI exposures by the proposed and state-of-the-art methods, for 16 band spectral cubes.

a significantly better recovery, achieving a performance which is comparable to the case of idealistic sampling, again fully exploiting all available measurements. On the other hand, the other methods under investigation present a performance similar to the case of RPs, demonstrating the inability of handle cases which diverge from ideal.

To better understand the type of information that is lost by each method, Fig. 9 demonstrates images of pixel-differences for a specific band associated with the proposed and state-of-the-art methods for recovering a 25 spectral cube from 1 (top) and 4 (bottom) exposures. As far as the case of a single exposure is concerned, these results are in agreement with previously reported results which demonstrate that for realistic sampling operators, the GRMR is able to achieve the best performance, suffering the loss of limited high-frequency information. It is interesting to note that both PPID and WB also exhibit a very good performance, while ItSD and BTES appear to suffer the loss of both high and low frequency content. Increasing the number of exposures to 4, one can observe that methods can achieve better performance, however, the proposed GRMR leads to extremely small divergence compared to the ground truth, unlike the rest of the methods which appear to losses significantly high frequency content.

To qualitatively illustrate the performance of each method, Fig. 10 presents an indicative example of a recovered band by the various techniques. In this case, the entire cube consists of 16 bands and the acquisition assumes two exposures where a realistic filter based sampling operator is employed. Visually, one can observe that the proposed GRMR method achieves the best preservation of high frequency components (e.g., test target), followed by PPID, ItSD, WB, and BTES, in that order. An interesting observation is that for this scenario, WB, which serves as a benchmark method, achieves a fairly good quality reconstruction, while methods such as BTES are heavily affected by the high-frequency grid-like pattern associated with snapshot mosaics.

Last, we also report the computational complexity in term of processing time each method in Table VI. The results indicate that the proposed method is in general slower compared to other method like PPID or WB, yet significantly faster compared to BTES. However, an important benefit of the proposed method is that the computational time is scales much slower as a function of acquired frames, in contrast to competing method which require the re-calculation of the estimated frames and the subsequent averaging.

VII. RECOVERY OF SENTINEL 2 EARTH OBSERVATIONS

In this subsection, we consider an innovative scenario where full spectral and spatial resolution observations, acquired by SSI-equipped earth orbit satellites, are recovered by the

TABLE VI
COMPUTATIONAL TIME (IN SECONDS) FOR THE PROPOSED AND STATE-OF-THE-ART METHODS FOR 16 AND 25 BAND CASES AND DIFFERENT NUMBER OF ACQUIRED FRAMES

Bands	16			25		
Frames	1	2	4	1	2	4
GRMR	12.5	14.1	17.4	21.1	25.3	33.6
BTES	35.1	69.4	137.1	158.3	318.1	637
WB	0.1	0.2	0.4	0.2	0.4	0.8
PPID	4	8	15.7	6.5	13	26.1
ItSD	0.3	0.7	1.35	0.8	1.7	3.4

TABLE VII
RECONSTRUCTION ERROR FOR RECOVERY OF SENTINEL 2 OBSERVATIONS FROM A SINGLE EXPOSURE AND THREE EXPOSURES, ASSUMING AN IDEALISTIC AND A REALISTIC SAMPLING PROCESS

PSNR	Idealistic sampling		Realistic sampling	
Observations	1	3	1	3
GRMR	36.5	**41.2**	**29.4**	**30.4**
BTES	35.6	36.4	20.1	20.3
WB	36.7	38.3	20.2	20.4
PPID	**39.2**	40.2	20.7	20.8
ItSD	24.6	24.7	24.6	24.8

proposed and state-of-the-art methods. Given the strict size, volume, and weights restriction, SSI technology has the potential to revolutionize the space imaging field, since it offers unique advantages, especially when considering miniaturized satellites such as cubesats [43] and how they could serve as a platform for Earth observation.

To validate the merits of the proposed approach, we consider the recovery of multi-spectral observations from the ESA Sentinel 2 satellite. Sentinel 2, part of the EU Copernicus programme, consists of two identical satellites (2A and 2B), which acquire imagery in 13 spectral bands, from visible to short-wave infrared at 10, 20 or 60 meter spatial resolution. In this study, we consider the 9 high spatial resolution (10 and 20 m) bands (2–8, 11, 12) and simulate the acquisition by an SSI camera using a 3×3 snapshot mosaic pattern of images of size 600×600 pixels. We consider two sampling models, namely an ideal binary pattern which effectively captures only the central wavelength and a satellite-specific spectral filter profile where the real spectral profiles for the 9 bands of Sentinel 2 A are employed. Similar to the previous case of CAVE data, image sequences were simulated assume a horizontal motion of one pixel at a random direction.

The performance in terms of PSNR is given in Table VII for the proposed and the state-of-the-art methods regarding the recovery of full spectral cubes from a single and three exposures, i.e., 1 and 2 observations per pixel. The results in the table are in sync with the performance reported on the CAVE dataset. More specifically, for the case of idealistic sampling, we observe that for a single exposure/frame, the PPID method achieves the best performance, while the proposed GRMR is better when three exposures are available. For the realistic sampling case, all method suffer a reduction in quality, however, the proposed method is still able to produce high quality reconstruction.

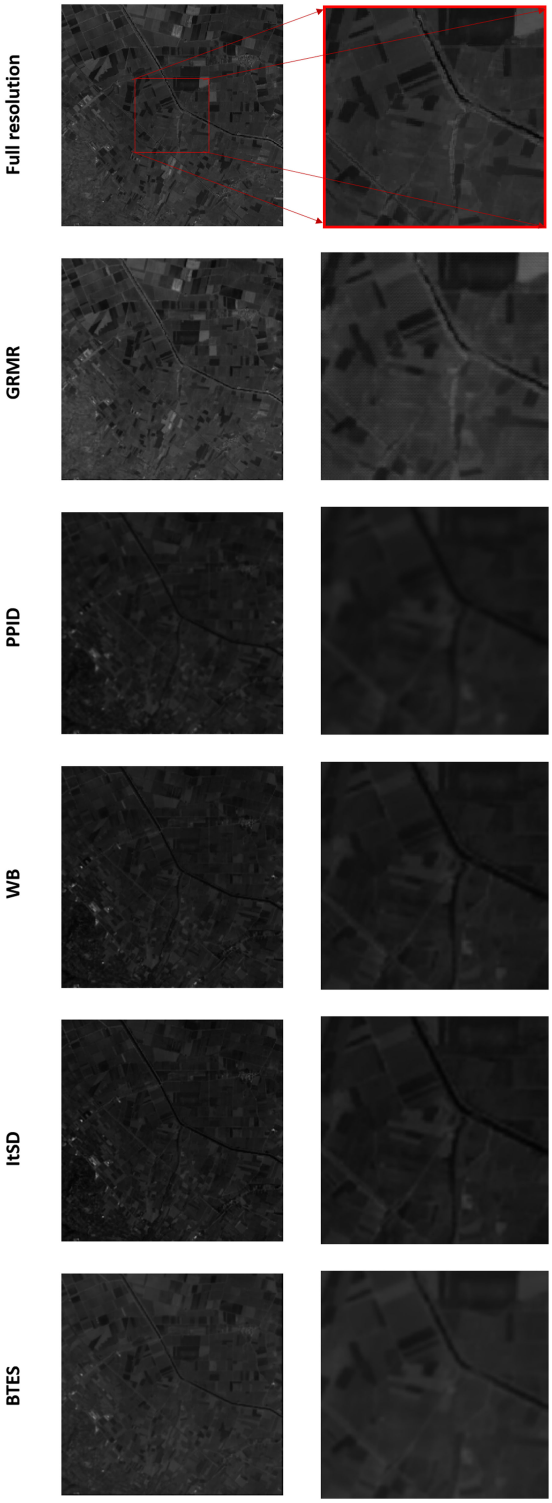


Fig. 11. Exemplary recovery of Sentinel 2 observations from 2 exposures using the real filter spectral profiles.

The significant performance gains achieved by the GRMR are also demonstrated in Fig. 11 for a moderate (left column) and a higher (right column) ground sampling distance. A visual comparison reveals that much higher content is recovered


(a) Mosaic (b) WB (c) PPID (d) GRMR

Fig. 12. Recovery of 10th band (620 nm) from a real 4 × 4 spectral snapshot of an outdoors scene.


(a) Mosaic (b) WB (c) PPID (d) GRMR

Fig. 13. Recovery of 1st band (608 nm) from a real 5 × 5 spectral snapshot of an indoors test scene.

by GRMR, while state-of-the-art methods like PPID and BTES suffer from very important artifacts, e.g., the profile of the continuous line (river), while benchmark methods, such as WB, completely fail and they introduce severe repeating patterns. The natural motion of spaceborne platforms can easily support the acquisition of more observations, suggesting an even higher performance boost for GRMR.

VIII. RECOVERY OF IMEC SSI OBSERVATIONS

In addition to the results shown in the previous section where the acquisition process was simulated, we herein provide results using snapshot spectral images acquired by two IMEC SSI technology based cameras. Since no ground truth is available for this case, we provide two illustrative examples of reconstruction in Figs. 12 and 13 for a 4 × 4 and a 5 × 5 SSI architecture, respectively. In this case, we simulated the acquisition of four exposures, thus reconstruction reaches half the spatial resolution of the detector on each axis, which is still much higher compared to the ×5 reduction for the 5 × 5.

Fig. 12 presents the spectral mosaic and recovery of the 620 nm profile of an outdoors scene using Weighted Interpolation, PPID, and the proposed GRMR method. Fig. 13 presents the acquired and reconstructed images of an indoors scene with targets of various resolutions, including the pattern shown in Fig. 14 which was photographed by a typical color camera. The recovered images correspond to the scenario where real spectral profiles are employed for the recovery of 16 (for Fig. 12) and 25 (for Fig. 13) band spectral cubes from 4 observations per pixel.

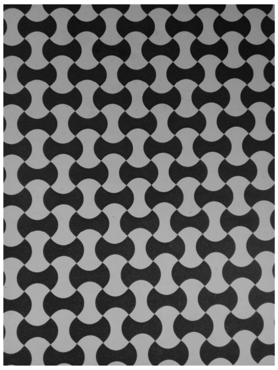


Fig. 14. Test pattern used for the indoors test scene.

One can easily observe that in both cases, the mosaic contains high frequency artifacts due to the incoherence of the sensing operation. Although the WB method provides a good baseline recovery, the low-pass nature of the operation leads to the loss of real high frequency information, encoding salience features like foliage. Furthermore, for the case of 25 bands recovery, the test recovered observations have significantly deteriorated the high frequency features of the target.

IX. IMPACT OF RECOVERY OF IMAGE ANALYSIS TASKS

While the estimation of full spatial and spectral resolution hyperspectral imagery from a limited number of snapshot mosaic acquisition in critical in its own right, so it the impact of the recovery on subsequent analysis tasks like scene classification. To validate the merits of the proposed scheme on this type

TABLE VIII
CNN ARCHITECTURE

Type	Size	Number
Convolution	3x3x2	64
Convolution	3x3x2	64
Max pooling		
Convolution	3x3x2	128
Convolution	3x3x2	128
Max pooling		
Convolution	3x3x2	256
Convolution	3x3x2	256
Max pooling		
Convolution	3x3x2	512
Convolution	3x3x2	512
Max pooling		
Fully connected		2048
Fully connected		2048
Fully connected		10

of tasks, we explore how state-of-the-art classification methods based on the paradigm shifting framework of deep learning are affected when presented with recovered imagery during inference. Specifically, we consider the influential framework of convolutional neural networks (CNNs) for the classification of Sentinel-2 imagery recovery by the proposed and state-of-the-art methods using the recently developed EuroSAT dataset [44], which is composed of 3000 13-band hyperspectral images of 64 × 64 pixels from ten classes, namely *Annual Crop, Forest, Herbaceous Vegetation, Highway, Industrial, Pasture, Permanent Crop, Residential, River* and *SeaLake*.

Although CNNs have been employed for a number of remote sensing image analysis tasks like land cover classification [45], scene classification [46], multi-source data fusion [47] and object localization [48], this work in the first to explore the capabilities of 3D CNNs for the classification of Sentinel-2 imagery, in contrast to the original work associated with the EuroSAT dataset [44] which employed available pre-trained networks and thus focuses on the classification capabilities using only three channel corresponding to red, green and blue wavelengths. The proposed CNN is composed of a sequence of convolutions with relu activations and max pooling layers which are fed to two fully connected layers for classification. More specifically, the architecture characteristics are shown in Table VIII.

To train the proposed CNN architecture, we selected 100 images from each class and validate the accuracy in a set of 20 images per class. The network was training using the Stochastic Gradient Descend (SGD) optimization and the categorical cross accuracy was selected as the loss function. Fig. 15 present the accuracy as a function of training epoch (one epoch is one pass over a training data) for both training and validation examples. These results indicate that very high accuracy can be achieved even from a limited number of examples and training time, while increasing these two training aspect will also lead to a smaller gap between training and validation error, a manifestation of overfitting.

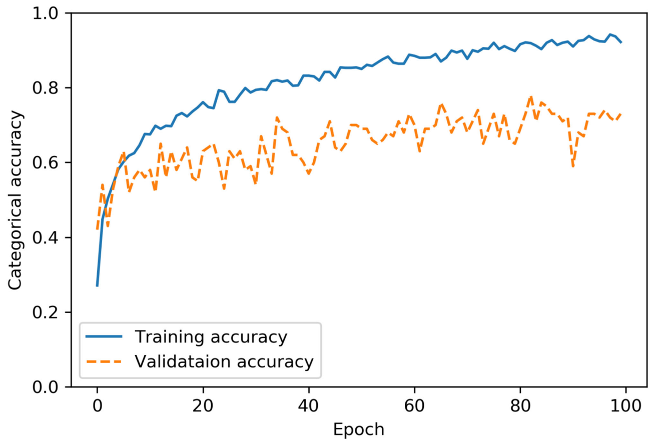


Fig. 15. Classification accuracy with full resolution input.

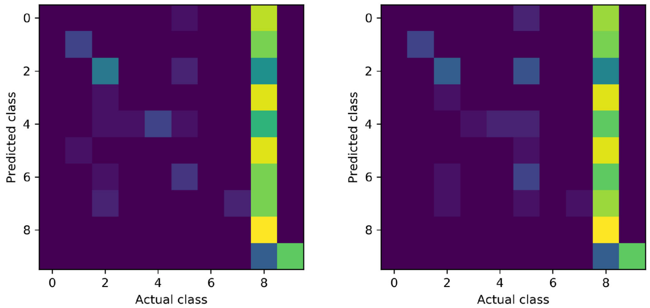
TABLE IX
RECOVERY QUALITY (PSNR) USING REAL FILTER PROFILES FROM 3 FRAMES OF SENTINEL 2 IMAGES

	GRMR	BTES	WB	PPID	ItSD
Annual crop	**44.1**	35.8	41.7	43.0	16.3
Forest	**43.1**	38.1	42.4	42.9	13.0
Herb/us veg.	**41.7**	34.9	39.9	41.6	15.7
Highway	**38.1**	30.9	36.0	37.2	15.8
Industrial	34.3	27.1	31.9	**35.2**	22.2
Pasture	**41.6**	35.3	40.2	40.9	14.0
Perm crop	39.9	32.4	37.6	**40.1**	16.1
Residential	**36.2**	30.3	34.5	36.0	19.4
Rivel	**39.5**	31.7	37.3	38.5	17.4
Sea lake	45.7	41.7	45.4	**45.9**	18.9

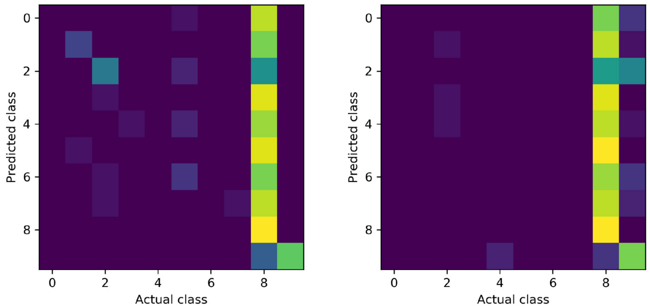
To understand the impact of recovery on classification, we assumed that the validation images were acquired with a 3 × 3 SSI camera employing the real Sentinel-2 spectral filter profiles. The reconstruction quality in terms of PSNR is given in Table IX for the different classes of imagery. We observe that in most cases, the proposed method achieves the best results, followed by the PPID method. Furthermore, we also notice that a significant variation exists among the different classes in term of reconstruction quality where for example the case of Forest imagery, high quality is achieved by all methods, unlike Industrial imagery where methods perform must worse.

Once the images were recovered, the training CNN was employed for classification. The accuracy for proposed and state-of-the-art methods is: GRMR: 26.5%, PPID: 24.5%, WB: 24.0%, ItSD: 18.0%, and BTES: 25.0%.

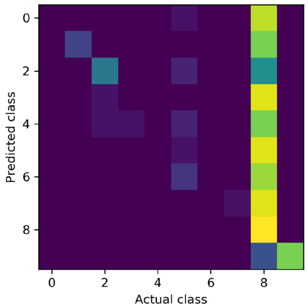
To gain some further insights into the impact of each method on the classification, Fig. 16 present the corresponding confusion matrices for all methods. As a general remark, we observe that although the proposed GRMR method achieved the best performance in this context, there is a significant gap between the average accuracy achieved by the demosaicing method and the accuracy when the full resolution images are used as shown in Fig. 16. The confusion matrices indicate that most of the examples are classified as Residential, which to an extend is



(a) GRMR (b) PPID



(c) WB (d) ItSD



(e) BTES

Fig. 16. Confusion matrices for different recovery methods.

expected since this particular class is showing the largest diversity in terms of content.

X. CONCLUSION

In this work, we address the problem of recovering high spatial and spectral resolution hyperspectral imagery from a limited number of sub-sampled spatio-spectral exposures acquired using a Snapshot Spectral Imaging architecture. Unlike existing approaches which assume that each detector elements acquires a ”perfectly” defined spectral band, we propose a framework where the non-ideal behavior of the spectral filters is seamlessly incorporated into the recovery process. To recover the full spectral and spatial resolution hyperspectral image from a limited number of snapshots, we propose a novel regularized optimization scheme which simultaneously exploits redundancies within image patches through a low rank term and spatial smoothness through a graph-based regularization term.

An in-depth investigation of the behavior of the proposed GRMR framework reveals that: (i) introducing both rank and graph regularization leads to higher quality reconstruction, (ii) the proposed framework can gracefully exploit information from multiple exposures and (iii) introducing knowledge regarding the spectral profile of each filter lead to significantly increased quality. The validate the merits of the proposed method, we explore the reconstruction quality of three diverse datasets, namely the benchmark CAVE dataset, a dataset from imagery acquired by the Sentinel 2 satellite and imagery acquired by an 25 band XIMEA acquired with a 5 × 5 pattern IMEC SSI sensor. The experimental results indicate that that proposed method achieves superior performance compared to state-of-the-art methods when a realistic sampling processing is considered, while high quality reconstruction is possible from a limited number of exposures, e.g., 2 exposures for estimating 25 spectral bands.

The proposed framework follows a “pure” signal processing perspective where prior knowledge is introduced into the recovery process. Future work will explore how machine learning based signal processing methodologies can be applied to the problem of SSI recovery. A key challenge that needs to be address in that respect is the lack of training data since SSI cameras acquire snapshot imagery only. Furthermore, more research is needed in term of demosaicing which can also lead to high classification accuracy in addition to reconstruction quality. A potential avenue for exploration is the utilization of recovered imagery for both training and validation, in an effort to mitigate shortcomings of the recovery to the classifier.

REFERENCES

[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2007.

[2] M. T. Eismann, *Hyperspectral Remote Sensing*. Bellingham, WA, USA: SPIE, 2012.

[3] D. Manolakis, R. Lockwood, and T. Cooley, *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[4] N. Hagen and M. W. Kudenov, “Review of snapshot spectral imaging technologies,” *Opt. Eng.*, vol. 52, no. 9, 2013, Art. no. 090901.

[5] P.-J. Lapray, X. Wang, J.-B. Thomas, and P. Gouton, “Multispectral filter arrays: Recent advances and practical implementation,” *Sensors*, vol. 14, no. 11, pp. 21626–21659, 2014.

[6] B. Geelen, N. Tack, and A. Lambrechts, “A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic,” in *Proc. SPIE* , 2014, Paper 89740L.

[7] B. Geelen, C. Blanch, P. Gonzalez, N. Tack, and A. Lambrechts, “A tiny VIS-NIR snapshot multispectral camera,” in *Proc. SPIE*, 2015, Paper 937414.

[8] R. W. Schafer and R. M. Mersereau, “Demosaicking: Color filter array interpolation,” *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 44–54, Jan. 2005.

[9] D. Menon and G. Calvagno, “Color image demosaicking: An overview,” *Signal Process., Image Commun.*, vol. 26, no. 8, pp. 518–533, 2011.

[10] Y. Monno, M. Tanaka, and M. Okutomi, “Multispectral demosaicking using guided filter,” in *Proc. IS&T/SPIE Electron. Imag.*, 2012, Paper 82990O.

[11] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, “Residual interpolation for color image demosaicking,” in *Proc. 20th IEEE Int. Conf. Image Process*, 2013, pp. 2304–2308.

[12] Y. Monno, D. Kiku, M. Tanaka, and M. Okutomi, “Adaptive residual interpolation for color image demosaicking,” in *Proc. IEEE Int. Conf. Image Process*, 2015, pp. 3861–3865.

[13] J. Brauers and T. Aach, “A color filter array based multispectral camera,” in *Proc. 12th Workshop Farbbildverarbeitung*, Ilmenau, Germany, 2006.

[14] J. Mizutani, S. Ogawa, K. Shinoda, M. Hasegawa, and S. Kato, "Multispectral demosaicking algorithm based on inter-channel correlation," in *Proc. Vis. Commun. Image Process. Conf.*, 2014, pp. 474–477.

[15] S. Mihoubi, O. Losson, B. Mathon, and L. Macaire, "Multispectral demosaicing using pseudo-panchromatic image," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 982–995, Dec. 2017.

[16] P. Getreuer, "Image demosaicking with contour stencils," *Image Process. On Line*, vol. 2, pp. 22–34, 2012.

[17] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.

[18] Y. Zhao, J. Yang, Q. Zhang, L. Song, Y. Cheng, and Q. Pan, "Hyperspectral imagery super-resolution by sparse representation and spectral regularization," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 1–10, 2011.

[19] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral resolution enhancement of hyperspectral images via sparse representations," *Electron. Imag.*, vol. 2016, no. 19, pp. 1–6, 2016.

[20] K. Degraux, V. Cambareri, L. Jacques, B. Geelen, C. Blanch, and G. Lafruit, "Generalized inpainting method for hyperspectral image acquisition," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 315–319.

[21] M. Golbabaee and P. Vandergheynst, "Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 2741–2744.

[22] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[23] T. Hu, H. Zhang, H. Shen, and L. Zhang, "Robust registration by rank minimization for multiangle hyper/multispectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2443–2457, Jun. 2014.

[24] J. Yang, Y.-Q. Zhao, J. C.-W. Chan, and S. G. Kong, "Coupled sparse denoising and unmixing with low-rank constraint for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1818–1833, Mar. 2016.

[25] Y. Xu, Z. Wu, and Z. Wei, "Spectral–spatial classification of hyperspectral image based on low-rank decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2370–2380, Jun. 2015.

[26] Q. Tan, Y. Liu, X. Chen, and G. Yu, "Multi-label classification based on low rank representation for image annotation," *Remote Sens.*, vol. 9, no. 2, 2017, Art. no. 109.

[27] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, 2009, Art. no. 717.

[28] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[29] M. Signoretto, R. Van de Plas, B. De Moor, and J. A. Suykens, "Tensor versus matrix completion: A comparison with application to spectral data," *IEEE Signal Process. Lett.*, vol. 18, no. 7, pp. 403–406, Jul. 2011.

[30] R. Zhu, M. Dong, and J.-H. Xue, "Spectral nonlocal restoration of hyperspectral images with low-rank property," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3062–3067, Jun. 2015.

[31] G. Tsagkatakis, M. Jayapala, B. Geelen, and P. Tsakalides, "Non-negative matrix completion for the enhancement of snapshot mosaic multispectral imagery," in *Proc. Electron. Imag.*, 2016, pp. 1–6.

[32] M. H. Kamal, B. Heshmat, R. Raskar, P. Vandergheynst, and G. Wetzstein, "Tensor low-rank and sparse light field photography," *Comput. Vis. Image Understanding*, vol. 145, pp. 172–181, 2016.

[33] E. J. Candes and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.

[34] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[35] N. Shahid, N. Perraudin, V. Kalofolias, G. Puy, and P. Vandergheynst, "Fast robust PCA on graphs," *IEEE J. Select. Topics Signal Process.*, vol. 10, no. 4, pp. 740–756, Jun. 2016.

[36] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pac. J. Optim.*, vol. 6, pp. 615–640, 2010.

[37] J. Tanner and K. Wei, "Normalized iterative hard thresholding for matrix completion," *SIAM J. Sci. Comput.*, vol. 35, no. 5, pp. S104–S125, 2013.

[38] N. Parikh *et al.*, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[39] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 504–517, Mar. 2016.

[40] N. Shahid, V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "Robust principal component analysis on graphs," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2812–2820.

[41] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

[42] L. Miao, H. Qi, R. Ramanath, and W. E. Snyder, "Binary tree-based generic demosaicking algorithm for multispectral filter arrays," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3550–3558, Nov. 2006.

[43] D. Selva and D. Krejci, "A survey and assessment of the capabilities of Cubesats for earth observation," *Acta Astronaut.*, vol. 74, pp. 50–68, 2012.

[44] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Int. Geosci. Remote Sens. Symp., Valencia*, pp. 204–207, 2018.

[45] K. Karalas, G. Tsagkatakis, M. Zervakis, and P. Tsakalides, "Deep learning for multi-label land cover classification," in *Proc. SPIE*, 2015, vol. 9643, Paper 96430Q.

[46] K. Nogueira, O. A. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.

[47] G. Scarpa, M. Gargiulo, A. Mazza, and R. Gaetano, "A CNN-based fusion method for feature extraction from sentinel data," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 236.

[48] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

Authors' photographs and biographies not available at the time of publication.