



Smoothness on rank-order path graphs and its use in compressive spectral imaging with side information

Juan F. Florez-Ospina^{a,*}, Daniel L. Lau^b, Dominique Guillot^c, Kenneth Barner^a, Gonzalo R. Arce^a

^a Dept. of Electrical and Computer Engineering, University of Delaware, Newark, DE, 19716, United States

^b Dept. of Electrical Engineering, University of Kentucky, Lexington, KY, 40506, United States

^c Dept. of Mathematical Sciences, University of Delaware, Newark, DE, 19716, United States

ARTICLE INFO

Article history:

Received 26 April 2021

Revised 1 February 2022

Accepted 25 July 2022

Available online 26 July 2022

Keywords:

Rank-order path graph

Bilateral filter graph

Compressive spectral imaging

Side information

Graph Laplacian regularization

SD-CASSI

Rank-order statistics

Discrete cosine transform

ABSTRACT

This paper proposes a novel reconstruction approach to compressive spectral imaging (CSI) with panchromatic side information, which is based on the notion of approximate rank-order statistics. To that end, we assume that the signal of interest is sufficiently smooth on an unknown graph. When restricted to the family of path graphs, we show that the best path is indeed the *rank-order path graph* induced by the signal. That is, the path graph whose edge structure is given by the permutation that sorts the entries of the signal in ascending order. Our goal is to show that smoothness on rank-order path graphs inferred from the rank-order statistics of a co-registered panchromatic signal can be used to find accurate spectral image estimates from a compressive snapshot of the scene. We derive theoretical properties of rank-order path graphs and give illustrative examples of their use in signal recovery from undersampled measurements. Our approach leads to solutions with a closed-form, found efficiently by iterative inversion of highly sparse systems of linear equations. We evaluate our method through an experimental demonstration and extensive simulations. Our method performs notably better against a bilateral-filter graph model, adapted to the task, and some traditional and state-of-the-art algorithms.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

The recovery of a signal from an incomplete or limited number of linear observations appears in numerous applications, ranging from medical imaging and seismic exploration to baggage and cargo inspection in homeland security. Due to the underdetermined nature of the problem, infinitely many signals may produce the same set of observations, and therefore to identify the one that is closest to the underlying signal of interest, we need to rely on prior knowledge of the signal. The solution to the inverse problem is thus the signal that best satisfies the prior knowledge and, simultaneously, fits the observations to a certain level.

The premise that most signals have a sparse representation is a popular form of prior knowledge. In this setting, the signal of interest is assumed to be sparse (or compressible) on a pre-specified dictionary, and the solution to the inverse problem is given by the signal with the sparsest representation that complies with the

measurements. Sparsity has motivated the creation of prior models such as block-sparsity and union-of-subspace models, which integrate additional knowledge about sparsity-pattern disposition, thus leading to improved sparsity-based prior models [1].

In the last decade, numerous prior models that go beyond sparsity have been proposed. Plug and Play (PnP) prior models, for instance, exploit the abundance of denoising algorithms by enabling their integration as prior models for general inverse problems [2]. Since denoising algorithms are not necessarily cast as an optimization problem with a precise regularization function, PnP-based algorithms may have ill-defined convergence properties, leading to the development of regularization by denoising (RED), which aims at exploiting denoising priors through an explicit regularization function [3,4]. More recently, there has been interest towards developing data-driven prior models, e.g., deep priors [5]. The idea is to discover the necessary low-level statistical information from the available data so as to generate prior models, which require little to no human input, and are able to outperform traditional hand-crafted priors.

Due to the advances in the field of graph signal processing (GSP) [6], a lot of attention is being devoted recently to develop graph-based prior models, leading to notable performance in in-

* Corresponding author.

E-mail addresses: jfflorez@udel.edu (J.F. Florez-Ospina), dlau@uky.edu (D.L. Lau), dguillot@udel.edu (D. Guillot), barner@udel.edu (K. Barner), arce@udel.edu (G.R. Arce).

verse problems such as light field super-resolution [7], blind image deblurring [8], tomographic reconstruction [9], and medical image fusion [10] to name a few. This special class of prior models can be easily integrated into inverse problems, often invoked by minimizing a differentiable convex objective, referred to as the graph Laplacian quadratic form. The associated graph is frequently constructed by defining nearest neighbor relationships among the elements of a related complementary signal¹, weighted by a Gaussian kernel with a certain bandwidth parameter.

To advance on that front, this paper develops a novel graph-based model for compressive spectral imaging (CSI) with panchromatic side information that not only departs from the Gaussian kernel approach often used in other applications but also, to the best of our knowledge, is the first to exploit the concept of smoothness on graphs for CSI. In particular, by adopting a signal representation perspective, we demonstrate that when the graph is restricted to the family of path graphs with vertex set $V = \{1, 2, \dots, n\}$, a signal of interest $x \in \mathbb{R}^n$ on V is best represented by the path graph with edge set $E \subset V \times V$ given by:

$$E = \{(r_i^{-1}, r_{i+1}^{-1}) : i = 1, \dots, n-1\},$$

where $r^{-1} = r_1^{-1} r_2^{-1} \dots r_n^{-1}$ is the permutation of V that sorts the entries of x in ascending order, that is

$$x_{r_1^{-1}} \leq x_{r_2^{-1}} \leq \dots \leq x_{r_n^{-1}}.$$

Since the inverse permutation r of r^{-1} forms the rank-order statistics of x , the path graph constructed by using r^{-1} is referred to as the *rank-order path graph* induced by x . Since in practice we do not know the rank-order statistics of the signal to construct the rank-order path graph, we rely on the notion of approximate rank-order statistics. Specifically, we assume there is a mechanism to provide approximate ranks, $\hat{r}_1 \hat{r}_2 \dots \hat{r}_n$ in such a way that the induced order statistics of the signal x , that is

$$x_{\hat{r}_1}, x_{\hat{r}_2}, \dots, x_{\hat{r}_n}$$

are sufficiently smooth. We note that to understand this deeply, we may need to resort to the theory of concomitants of order statistics, or equivalently induced order statistics [11,12], which is out of the scope of the paper.

In addition to providing a theoretical study of rank-order path graphs in the context of signal recovery from undersampled linear measurements, we show the potential of rank-order path graphs to tackle the problem of reconstructing spectral images from compressive measurements. In imaging spectroscopy, we would like to characterize a scene of interest by sensing large amounts of spatial information across a multitude of frequency bands. Since the efficient sensing of all these bands is challenging, compressive spectral imaging (CSI) was developed to reconstruct a spectral image from a single coded snapshot of the scene [13,14]. However, when the scene of interest contains fine spatial details and diverse spectral content, a single coded snapshot may be insufficient to allow accurate spectral-image estimates, and we can often rely on broadband panchromatic detectors that provide side information to improve reconstruction quality [15–17]. This poses an ideal scenario where the rank-order statistics of the side information can be used to construct approximate rank-order path graphs, which are suitable to regularize the CSI inverse problem. Since the rank-order statistics are a statistical property of scalar data, there may be concern about their extension to multivariate data. We thus present a natural definition of multivariate rank-order statistics, which can be used in that setting.

¹ A signal related to the signal of interest, which contains desirable properties that we would like to impose on our signal estimate.

1.1. Paper organization

The remainder of the paper is organized as follows. Section 2 states the problem of signal recovery from under-sampled linear measurements, and how to solve it based on sparse representations and smoothness on graphs. Section 3 defines the concept of rank-order path graphs, states some of their properties, and explains how to use smoothness with respect to (approximate) rank-order path graphs to solve the problem. Section 4 develops an illustrative example. Section 5 states the problem of CSI with side information, explains some traditional approaches, and develops algorithms to tackle the problem based on smoothness on graphs, and in particular rank-order path graphs and the bilateral filter graph. Section 6 presents experimental results with simulated and real data. Section 7 concludes the paper.

2. Signal recovery from undersampled linear measurements

Consider the problem of signal recovery from a limited number of linear measurements. The goal is to reconstruct an unknown signal of interest $x \in \mathbb{R}^n$ from a vector of (noisy) measurements $y \in \mathbb{R}^m$ such that

$$y = Ax + e, \quad (2.1)$$

where $A \in \mathbb{R}^{m \times n}$ is a sensing matrix with fewer rows than columns, i.e., $m < n$, and e is an additive noise term. Since there are infinitely many signals that comply with the measurements, some form of regularity on the feasible set of signals must be imposed to be able to recover the original signal [18].

Under the assumption that x belongs to the set of s -sparse signals $\mathcal{M}_s = \{x \in \mathbb{R}^n, \|\Psi^T x\|_0 \leq s\}$, where s represents the sparsity level of a signal on a given dictionary Ψ , an accurate estimate $\hat{x}$ of the signal x can be obtained by solving the problem [1,18]:

$$\min_{x \in \mathbb{R}^n} \|\Psi^T x\|_1 \text{ s.t. } \|Ax - y\|_2 \leq \epsilon \quad (2.2)$$

provided that A satisfies the null-space property when the noise level $\epsilon = 0$ or similarly the restricted isometry property (RIP) when $\epsilon \geq 0$. In this work, however, we depart from the ℓ_1 -regularization problem and instead propose to reconstruct the signal of interest x by solving the problem:

$$\min_{x \in \mathbb{R}^n} x^T L_G x \text{ s.t. } \|Ax - y\|_2 \leq \epsilon, \quad (2.3)$$

where L_G is the graph Laplacian of an undirected graph G . As will be elaborated, the problem (2.3) suggests that the best solution to the underdetermined system of linear Eq. (2.1) is the signal that is smoothest with respect to G . Note that the problem (2.3) can be viewed from the point of view of signal recovery on graphs [19] for certain sensing matrices A . In our case, however, the matrix A does not necessarily perform an operation on the graph. The problem (2.3) can instead be regarded as a particular instance of kernel-based reconstruction, where L_G defines a topology-based kernel [20].

3. Rank-order path graphs

We now define formally the concept of rank-order path graph and smoothness with respect to a graph. Then, we derive important theoretical results, stated in the form of theorems and propositions, which can be leveraged to design graph-based smoothness models for signal recovery. Last, we study the problem of signal recovery from undersampled measurements and how to solve it using rank-order path graphs.

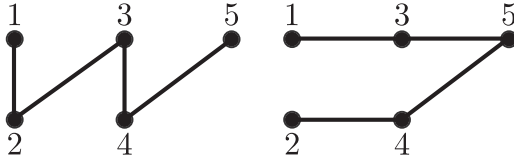


Fig. 1. Illustration of two different path graphs on $V = \{1, 2, 3, 4, 5\}$.

3.1. Graphs, permutations, and rank-order path graphs

An undirected graph $G = (V, E, w)$ is a triple, consisting of a vertex set $V = \{1, 2, \dots, n\}$, an edge set $E \subset V \times V$, and a nonnegative weight function $w: E \mapsto [0, \infty)$ such that $w(i, j) = w(j, i) > 0$ for $(i, j) \in E$, and $w(i, j) = 0$ for $(i, j) \notin E$. In this work, we assume that an edge (i, j) is an unordered pair of vertices $i, j \in V$, and graphs are not allowed to have self-loops, i.e., $(i, i) \notin E$ for any $i \in V$. Also, when w is unspecified, we assume the graph is unweighted, or equivalently that $w(i, j) = 1$ for $(i, j) \in E$ and zero otherwise.

The adjacency matrix $\mathbf{W}_G$ of G is an $n \times n$ matrix whose entries W_{ij} are given by the weight function $w(i, j)$ at the edge $(i, j) \in V \times V$, i.e., $W_{ij} = w(i, j)$. By definition $w(i, j) = w(j, i)$, and therefore we have that the adjacency matrix $\mathbf{W}_G$ is symmetric, i.e., $\mathbf{W}_G = \mathbf{W}_G^T$. The degree matrix $\mathbf{D}_G$ of G is an $n \times n$ diagonal matrix whose diagonal entries D_{ii} are given by $\sum_{j=1}^n w(i, j)$. The graph Laplacian $\mathbf{L}_G \in \mathbb{R}^{n \times n}$ of G is an $n \times n$ matrix whose diagonal entries L_{ii} are given by $\sum_{j=1}^n w(i, j)$ and off-diagonal entries L_{ij} are given by $-w(i, j)$, equivalently

$$\mathbf{L}_G = \mathbf{D}_G - \mathbf{W}_G. \quad (3.1)$$

An important characteristic of the graph Laplacian is that it is symmetric positive semidefinite, thus its eigenvalues are real and non-negative.

A permutation σ of the set V can be defined by (1) a linear ordering $\sigma = \sigma_1 \sigma_2 \dots \sigma_{n-1} \sigma_n$ where $\sigma_i \in V$ is listed exactly once; (2) a bijective map $\sigma: V \mapsto V$ such that $\sigma(1) = \sigma_1, \sigma(2) = \sigma_2, \dots, \sigma(n-1) = \sigma_{n-1}, \sigma(n) = \sigma_n$. Sometimes, we may need to represent σ as a permutation matrix $\mathbf{P} \in \{0, 1\}^{n \times n}$ such that

$$\mathbf{P} = \sum_{j=1}^n \mathbf{e}_j \mathbf{e}_{\sigma_j}^T,$$

where $\mathbf{e}_j, j = 1, \dots, n$ denote the standard basis. Equivalently, we can generate $\mathbf{P}$ by permuting the rows of the $n \times n$ identity matrix according to σ .

To simplify notation, let S_n be the set of all possible permutations of V .

Example 3.1. Consider the set S_3 of permutations of $V = \{1, 2, 3\}$. Then S_3 consists of the permutations 123, 132, 213, 231, 312, and 321. The permutation $231 \in S_3$ can be written as $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$, or

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Example 3.2. A path graph on the vertex set $V = \{1, 2, \dots, n\}$ is a graph $G = (V, E)$ with edge set $E \subset V \times V$ given by

$$E = \{(\sigma_1, \sigma_2), (\sigma_2, \sigma_3), \dots, (\sigma_{n-1}, \sigma_n)\}$$

where σ is a permutation of V . Fig. 1 shows pictorial representations of a pair of path graphs on $V = \{1, 2, 3, 4, 5\}$, which are associated with the permutations 12345 and 13542.

Further, let $\mathcal{L}_n$ be the set of possible path graphs on $V = \{1, \dots, n\}$. It follows that for $G \in \mathcal{L}_n$, there exist a permutation $\sigma \in S_n$ such that the edge set E of G is given by $\{(\sigma_i, \sigma_{i+1})\}_{i=1}^{n-1}$.

Definition 3.3. Let $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ be a signal on the vertex set $V = \{1, \dots, n\}$. The path graph $G \in \mathcal{L}_n$ is said to be the rank-order path graph induced by x if the permutation $r^{-1} = r_1^{-1} r_2^{-1} \dots r_n^{-1} \in S_n$ associated with the path graph G satisfies:

$$x_{r_1^{-1}} \leq x_{r_2^{-1}} \leq \dots \leq x_{r_n^{-1}}, \quad (3.2)$$

where $x_{r_i^{-1}}$ denotes the i -th order statistic $x_{(i)}$ of the entries of x .

The inverse permutation r of r^{-1} in S_n forms the rank-order statistics of x ; that is, x_j has rank r_j among the entries of x . To illustrate the concept, consider the following example.

Example 3.4. Let $x = (0.53, 0.25, 0.10, 0.77, 0.42)^T$ be a signal on the vertex set $V = \{1, \dots, 5\}$. Observe that

$$x_3 = 0.10 \leq x_2 = 0.25 \leq x_5 = 0.42 \leq x_1 = 0.53 \leq x_4 = 0.77.$$

Thus, we obtain that $r^{-1} = 32514$. The associated path graph is therefore $G \in \mathcal{L}_5$ with edge set E given by:

$$E = \{(3, 2), (2, 5), (5, 1), (1, 4)\}.$$

Also, note the rank-order statistics of x are given by $r = 42153 \in S_5$, which can be verified to be the inverse permutation of r^{-1} as mentioned above.

3.2. Smoothness with respect to rank-order path graphs

Consider a graph G on the vertex set $V = \{1, \dots, n\}$, and let $x \in \mathbb{R}^n$ be a signal on V . Then, the smoothness of x with respect to G is defined by Shuman et al. [21]:

$$x \mapsto x^T \mathbf{L}_G x. \quad (3.3)$$

As elaborated in [21], the smoothness of a graph signal varies depending on the underlying graph on which it resides. That is, a signal, which may be considered smooth on a given graph, may no longer be smooth on a different graph. In the following theorem, we show that when the underlying graph G is restricted to the set of path graphs $\mathcal{L}_n$, the path graph on which a signal x is the smoothest is the rank-order path graph induced by x .

Theorem 3.5. Let $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ and let $r^{-1} \in S_n$ be a permutation such that

$$x_{r_1^{-1}} \leq x_{r_2^{-1}} \leq \dots \leq x_{r_n^{-1}}. \quad (3.4)$$

Then the function $F(G) = x^T \mathbf{L}_G x$ achieves a global minimum on $\mathcal{L}_n$ at the graph G^* with edge set $E = \{(r_i^{-1}, r_{i+1}^{-1})\}_{i=1}^{n-1}$.

Proof. See Appendix A $\square$

This theorem states the fundamental principle of using rank-order path graphs to regularize inverse problems. As elaborated shortly, when we know the rank-order statistics of a signal of interest, and a few underdetermined linear observations of the signal, it is possible to find accurate estimates of a variety of signals using smoothness with respect to rank-order path graphs.

Rank-order statistics can naturally be extended to multivariate signals (e.g., spectral images or video). To generalize the concept, we rely on the notion of smoothness with respect to a collection of signals. More formally, the smoothness of a collection of signals, $x_1, x_2, \dots, x_L \in \mathbb{R}^n$ with respect to G is given by Kalofolias [22]:

$$X \mapsto \text{tr}(X^T \mathbf{L}_G X), \quad (3.5)$$

where $X \in \mathbb{R}^{n \times L}$ is such that $X = (x_1, x_2, \dots, x_L)$. Note that when $L = 1$, (3.5) reduces to (3.3).

By analogy to the case where X consists of only a signal (Theorem 3.5), we define the multivariate rank-order statistics of X to be the inverse permutation associated with the path solving the problem:

$$\min_{G \in \mathcal{L}_n} F(G) := \text{tr}(X^T \mathbf{L}_G X). \quad (3.6)$$

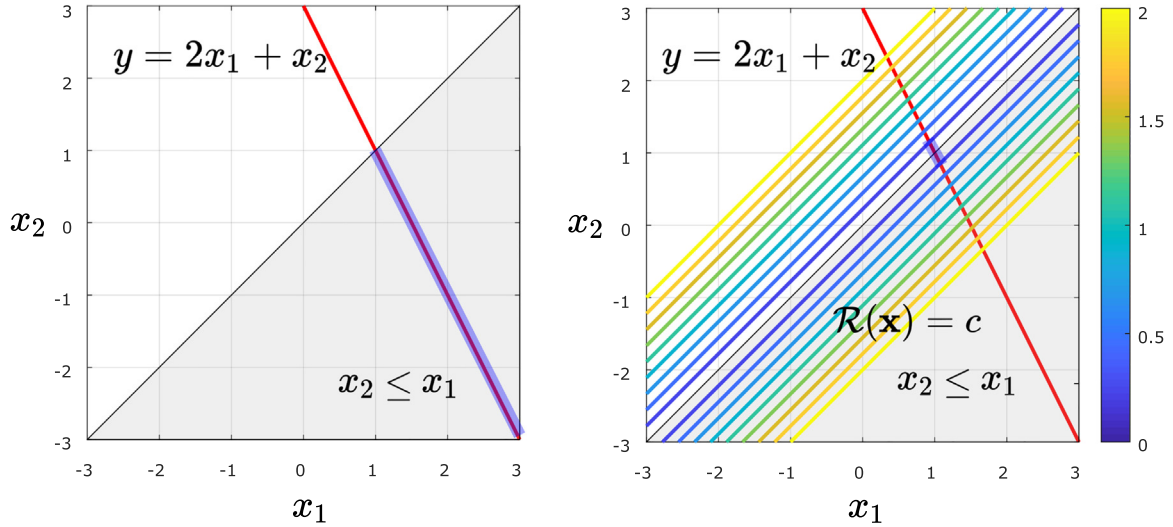


Fig. 2. Geometric illustration of underdetermined system of linear equations subject to ordering constraints. Left, the shaded region illustrates the ordering relationship between the components of $\mathbf{x} = (x_1, x_2)^T$, i.e., $x_2 \leq x_1$. Right, the contour lines of the objective function, which favor signals whose induced order statistics are smooth.

In particular, such a path is a Hamiltonian path of minimal length, and can be found by optimizing over the set of permutations $\mathcal{S}_n$ as stated in the following proposition.

Proposition 3.6. Problem (3.6) can be reformulated as follows:

$$\min_{G \in \mathcal{L}_n} \text{tr}(X^T \mathbf{L}_G X) = \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n-1} \|\mathbf{x}_{\sigma_i} - \mathbf{x}_{\sigma_{i+1}}\|_2^2, \quad (3.7)$$

where $\mathbf{x}_{\sigma_i} = (x_{\sigma_i 1}, \dots, x_{\sigma_i L})$ is the σ_i -th row of X .

Proof. See Appendix B $\square$

3.3. Spectral properties of rank-order path graphs

A smooth signal on the graph G is often described as having low-pass spectrum with respect to the spectral basis of $\mathbf{L}_G$. That is, most of the signal's energy is concentrated on the eigenvectors associated with the smallest eigenvalues. In the case of a (rank-order) path graph, the spectral basis has an explicit form, and therefore this can facilitate graph signal processing on the spectral domain of path graphs. As elaborated in [23,24], a path graph $\tilde{G}$ with vertex set $V = \{1, \dots, n\}$ and edge set $E = \{(i, i+1)\}_{i=1}^{n-1}$ has graph Laplacian $\mathbf{L}_G$ with eigenvalues $\tilde{\lambda}_k \geq 0$, $k = 0, \dots, n-1$ such that

$$\tilde{\lambda}_k = 2 - 2 \cos\left(\frac{\pi k}{n}\right), \quad (3.8)$$

and eigenvectors $\tilde{\mathbf{u}}_k \in \mathbb{R}^n$, $k = 0, \dots, n-1$ with entries $\tilde{u}_{jk}$ given by

$$\tilde{u}_{jk} = \sqrt{\frac{2}{n}} \cos\left(\frac{(j + \frac{1}{2})\pi}{n} k\right), \quad j = 0, \dots, n-1, \quad (3.9)$$

where $\tilde{\mathbf{u}}_0, \tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_{n-1}$ are also referred to as the DCT-2 basis. We note that in the case where $k=0$, the value of $\tilde{u}_{jk}$ has to be divided by $\sqrt{2}$ so that the eigenvectors form an orthonormal basis [24]. That is, the first eigenvector $\tilde{\mathbf{u}}_0$ is given by $\tilde{\mathbf{u}}_0 = \frac{1}{\sqrt{2}}(\sqrt{\frac{2}{n}}, \dots, \sqrt{\frac{2}{n}})^T$. More generally, the spectral basis of a rank-order path graph is a permuted version of the DCT-2 basis as stated in the following theorem.

Theorem 3.7. As before, let $\tilde{G} \in \mathcal{L}_n$ be the path graph with edge set given by $(1, 2), (2, 3), \dots, (n-1, n)$. Then for any $G \in \mathcal{L}_n$, its graph

Laplacian $\mathbf{L}_G$ is given by

$$\mathbf{L}_G = \mathbf{P}^T \mathbf{L}_{\tilde{G}} \mathbf{P} \quad (3.10)$$

for some permutation matrix $\mathbf{P}$. As a consequence, the eigenvectors $\mathbf{u}_k$ of $\mathbf{L}_G$ are given by

$$\mathbf{u}_k = \mathbf{P}^T \tilde{\mathbf{u}}_k, \quad (3.11)$$

where $\tilde{\mathbf{u}}_k$, $k = 0, \dots, n-1$ are defined in (3.9).

Proof. See Appendix C. $\square$

3.4. Signal recovery using rank-order path graphs

As mentioned in Section 2, we would like to recover a signal of interest $\mathbf{x} \in \mathbb{R}^n$ from a set of noisy measurements $\mathbf{y} \in \mathbb{R}^m$ obeying $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ as in (2.1), by solving the problem (2.3). In doing so, we now assume the graph G to belong to the family of path graphs $\mathcal{L}_n$, and in particular the path graph induced by available rank-order information of the signal of interest. The problem is thus to minimize $\mathbf{x}^T \mathbf{L}_G \mathbf{x}$ over the set $\{\mathbf{x} : \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}$.

At first, it would appear that the rank-statistics of a signal tell everything there is to know about $\mathbf{x}$, but there is an abundance of signals with the same rank-order statistics. So, even if we have a limited number of observations from the signal, we may still be unable to recover an accurate estimate of $\mathbf{x}$. Therefore, in addition to suitable rank-order information, the signal of interest has to be sufficiently smooth with respect to the given rank-order path graph. We illustrate the concept in the following example.

Example 3.8. Consider the recovery of a two-dimensional signal $\mathbf{x} = (x_1, x_2)^T$ from a scalar linear measurement $y = \mathbf{A}\mathbf{x}$ with $\mathbf{A} = (2, 1)$. Furthermore, assume that the rank statistics r of $\mathbf{x}$ are given by $r = (2, 1)$, so $\mathbf{x} \in \mathcal{M} = \{\mathbf{x} \in \mathbb{R}^2, x_2 \leq x_1\}$. Note that the set $\{\mathbf{x} \in \mathbb{R}^2, \mathbf{A}\mathbf{x} = y\} \cap \{\mathbf{x} \in \mathbb{R}^2, x_2 \leq x_1\}$, highlighted in blue in Fig. 2(left), does not allow us to distinguish a unique estimate of $\mathbf{x}$. Observe, however, that under the assumption that $\mathbf{x}$ is smooth with respect to the path graph $G \in \mathcal{L}_2$ with edge set $\{(2, 1)\}$, it is possible to identify unique estimates of $\mathbf{x}$ by minimizing $\mathcal{R}(\mathbf{x})$ over $\{\mathbf{x} \in \mathbb{R}^2, \mathbf{A}\mathbf{x} = y\} \cap \{\mathbf{x} \in \mathbb{R}^2, x_2 \leq x_1\}$. Since $\mathcal{R}(\mathbf{x}) = (x_1 - x_2)^2$ decreases along $y = 2x_1 + x_2$ as the entries of $\mathbf{x}$ take similar values as illustrated by the colored contour lines in Fig. 2(right), it can be said that $\mathcal{R}(\mathbf{x})$ will favor or prefer signals with smooth reordered entries.

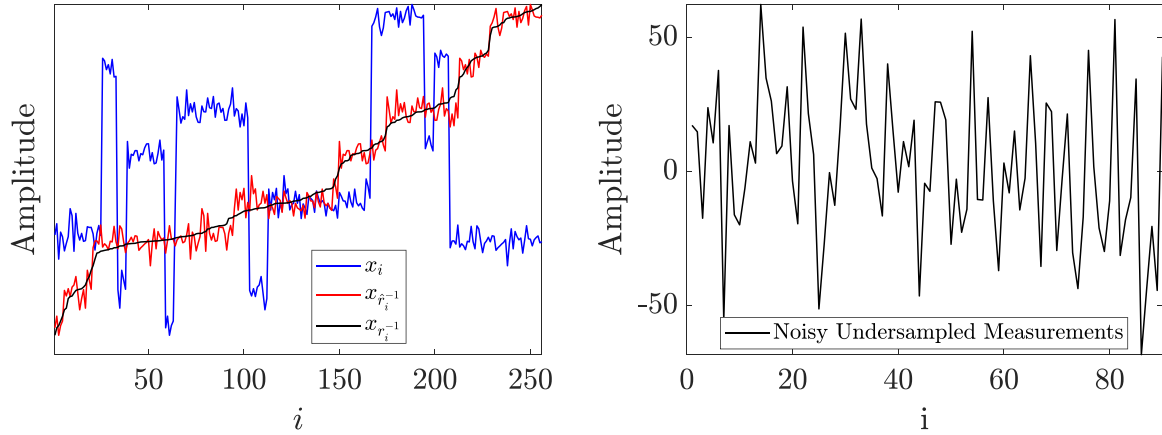


Fig. 3. Signal of interest in blue on the left and noisy undersampled measurements on the right. On the right, in red and black, we show the rearrangements of the signal of interest when permuting its entries using $\hat{r}^{-1}$ and r^{-1} , which encode approximate and exact rank-order statistics of the signal. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. Approximate (Multivariate) rank-order statistics

In practice, the (multivariate) rank-order statistics of a signal of interest have to be estimated from readily available data. A possible approach to this problem is to borrow the rank-order statistics of a related signal that can be inexpensively acquired. In CSI with side information, for instance, the rank-order statistics of a panchromatic image can be used as approximate multivariate rank-order statistics of the spectral image of interest. Another possible approach is to estimate the (multivariate) rank-order statistics from a collection of noisy estimates of the signal of interest by exploiting the notion of (multivariate) rank-order statistics in Section 3.2. This problem, however, does not have closed form, and its solution can be found as a particular instance of the Traveling Salesman Problem (TSP).

4. Illustrative examples

We now develop an example of signal recovery from noisy undersampled linear measurements, where a one-dimensional signal would like to be reconstructed using smoothness with respect to rank-order path graphs induced by approximate rank-order information.

Let $x = (x_1, \dots, x_n)^T$, depicted in blue in Fig. 3(left), be the signal of interest, and let $y = (y_1, \dots, y_m)$, depicted in Fig. 3(right), denote a set of noisy undersampled measurements from x such that $y = Ax + e$ where $A \in \mathbb{R}^{m \times n}$ is a sensing matrix with entries drawn from $\mathcal{N}(0, \text{SD} := 1)$, and $e \in \mathbb{R}^m$ is a noise term with entries drawn from $\mathcal{N}(0, \text{SD} := 0.3)$.

The idea is to reconstruct x from y through the program (2.3), where it is assumed that x is the smoothest with respect to a path graph $G \in \mathcal{L}_n$ whose structure depends on a given permutation as explained in Ex. 3.2. An important question to ask is: what path graph should we use to obtain accurate signal estimates? In this example, we use three different path graphs to show how this selection affect the reconstruction result.

The first path graph is constructed based on the permutation $\sigma \in S_n$ obtained by sorting the temporal position of the samples from smallest to largest. The second graph is constructed based on the permutation $\hat{r}^{-1}$ obtained by sorting the elements of a piecewise constant approximation of the original signal in ascending order. This permutation arranges the elements of the original signal depicted in blue in Fig. 3(left) in approximately ascending order as depicted by the signal in red in Fig. 3(left). The third path graph is the rank-order path graph induced by the original signal, which

is based on the permutation $r^{-1} \in S_n$ and sorts the entries of the original signal in ascending order as depicted by the black signal in Fig. 3(left).

For $n = 256$, $m = 90$, signal estimates and error signals are shown in Fig. 4; where the extent of the feasible set $\{x : \|Ax - y\|_2 \leq \epsilon\}$ is given by $\epsilon = \text{SD}\sqrt{m + \sqrt{2m}}$ with $\text{SD} = 0.3$ due to the fact that the norm of the error term e is a chi-square random variable [18]. In the figure, from left to right, we can observe that the estimate's accuracy is best when we have complete knowledge of the rank-order statistics of x . However, when we have partial knowledge, as illustrated in the center figure, the signal estimate may still be sufficiently accurate depending on the application. To run the numerical experiment, we used the CVX convex optimization toolbox [25].

In general, the estimation error $\|\hat{x} - x\|_2$ of a signal estimate $\hat{x}$, obtained by the program (2.3), is controlled by the smoothness of the signal of interest x with respect to the pre-specified graph G . Fig. 5 shows the graph spectrum of x on the path graphs induced by σ , $\hat{r}^{-1}$, r^{-1} . As Theorem 3.5 indicates, x is the smoothest on the path graph induced by r^{-1} , and thus its spectrum is concentrated the most at the lowest frequencies as shown in Fig. 5 on the right. The spectrum of x is not as concentrated on the remaining path graphs, but we can notice that partial knowledge of the rank-order statistics may also lead to high energy compaction as shown in Fig. 5 on the center.

5. Compressive spectral imaging using rank-order path graphs

We now consider the problem of compressive spectral imaging (CSI). Here the aim is to reconstruct a spectral image of L bands, $X_1, X_2, \dots, X_L \in \mathbb{R}^{n_1 \times n_2}$ from a single spatio-spectrally coded snapshot $Y \in \mathbb{R}^{n_1 \times n_2 + L - 1}$, captured by a CSI camera, where the relationship between $\{X_l\}_{l=1}^L$ and Y is given by

$$y \approx Hx, \quad (5.1)$$

where $y = \text{vec}(Y)$, $x = (\text{vec}(X_1)^T, \dots, \text{vec}(X_L)^T)^T$, and H denotes a sensing matrix, arising from the discretization of a continuous imaging model. In this work, we assume that the CSI camera is a single disperser coded aperture snapshot spectral imaging (SD-CASSI) system, depicted in the horizontal (or primary arm) of the dual-camera compressive spectral imager (DC-CSI) in Fig. 6. At a basic level, a SD-CASSI system encodes a collection of L images, captured at different wavelengths, into a single snapshot. In doing so, the images are first spatially modulated by a wavelength-independent random coded aperture τ , then a

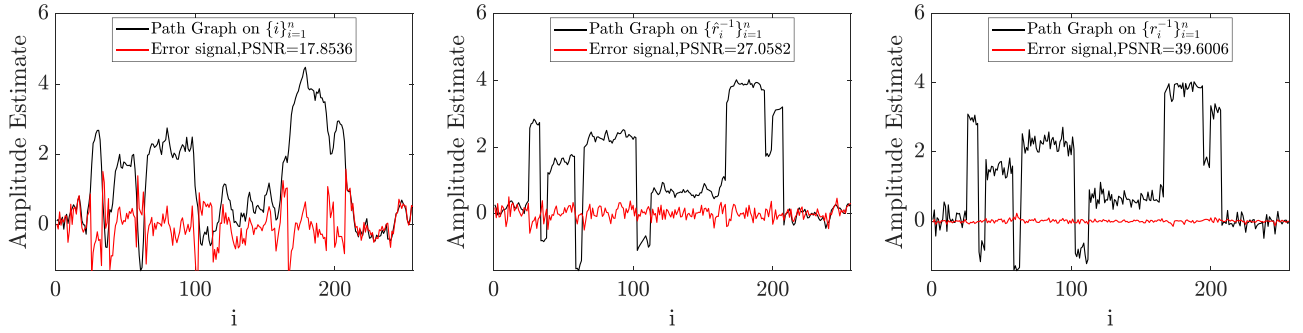


Fig. 4. Illustration of signal recovery from undersampled linear measurements using rank-order path graphs, encoding rank-order information of the signal of interest with different levels of accuracy. From left to right, we display signal estimates for the cases where there is no, partial, and complete knowledge of the rank-order statistics of the signal of interest.

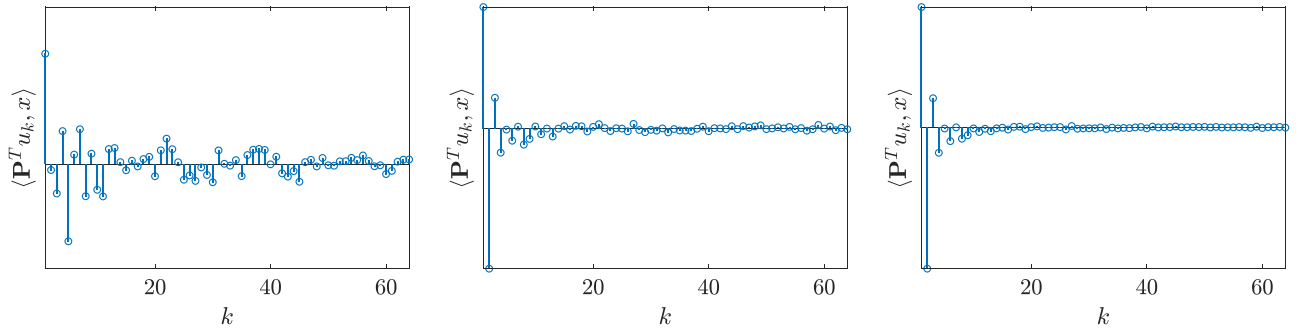


Fig. 5. Graph spectrum of the signal of interest x (depicted in blue in Fig. 5) on three different path graphs. From left to right, the first 64 coefficients of the spectrum of x on the temporal-order path graph, an approximate rank-order path graph, and the rank-order path graph induced by x . The matrix $\mathbf{P}$ is thus given by the identity matrix, the permutation matrix induced by $\hat{r}^{-1}$, and the permutation matrix induced by r^{-1} , respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

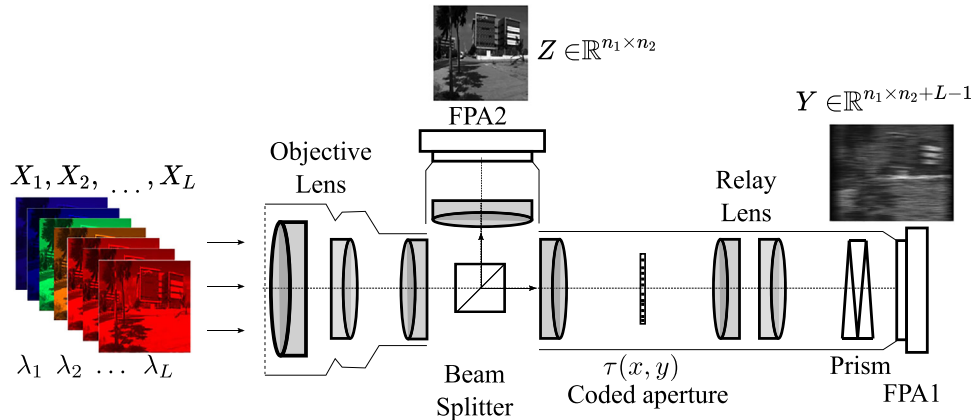


Fig. 6. Schematic of single disperser CASSI system with side information.

wavelength-dependent shift, along the x -axis, is introduced to each of the images by the prism, and finally the spatio-spectrally coded images are integrated along the wavelength domain by the focal plane array FPA1 [26,27].

As mentioned in Section 2, since (5.1) is underdetermined and approximate, we cannot simply take the inverse of $\mathbf{H}$ to obtain x . Instead, we search for the spectral image that best describes our prior knowledge of the scene, and simultaneously fits the measurements to a pre-specified noise level.

In the last decade, numerous approaches to CSI, which use different forms of prior knowledge, have been developed from the most traditional ones based on compressive sensing (CS) to the most recent ones based on deep learning. For a comprehensive review on the topic, we refer the reader to the recent signal processing magazine in snapshot compressive imaging [13,14]. For the pur-

pose of this paper, we only review state-of-the-art methods related to CS. In particular, approaches based on sparse representation and total variation.

Sparse representation approaches can be formulated as the problem (2.2), where it is assumed that the spectral image of interest has the sparsest representation on a given dictionary Ψ . A suitable dictionary can be either selected as a pre-specified set of functions, or designed based on signal examples [28]. In our experiments, we define $\Psi = \Psi_{2D-W} \otimes \Psi_{1D-DCT}$, where Ψ_{2D-W} denotes two-dimensional Symmlet-8 wavelet transform basis, and Ψ_{1D-DCT} denotes discrete cosine transform basis [13,29–31]. Hereafter, this approach is referred to as WDCT.

Similarly, the total variation model for spectral images assumes that the spectral image has minimal total-variation (TV) norm $\|x\|_{TV}$ over the feasible set defined by the measurements. Math-

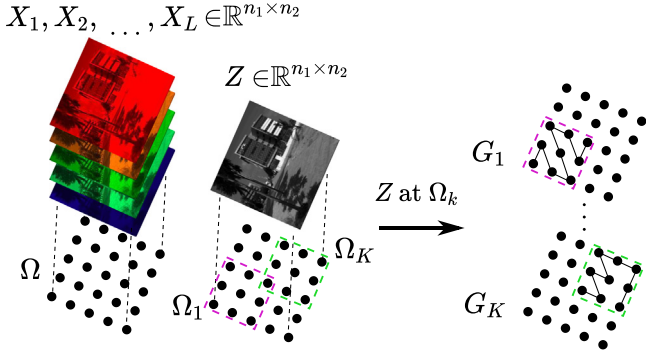


Fig. 7. Spectral and panchromatic images on a two dimensional domain Ω . The local rank-order information of Z at Ω_k leads to a collection of rank-order paths G_k on Ω . The G_k 's are then merged into a single graph G by (5.6).

ematically, the reconstruction problem can be formulated as:

$$\min_{x \in \mathbb{R}^{n_1 n_2 L}} \|x\|_{TV} \text{ s.t. } \|y - \mathbf{H}x\|_2 \leq \epsilon \quad (5.2)$$

where $\epsilon \geq 0$ is the noise level. There are several ways to define the total variation of a spectral image, but these often result from the extension of the total variation of a scalar image [32] to the case of vector (or multichannel) images [26,33,34]. In our experiments, $\|x\|_{TV}$ is given by $\sum_{l=1}^L \sum_{i,j} \|\nabla(\mathbf{F}_l)_{ij}\|_2$ where $\nabla \approx (\frac{d}{dx}, \frac{d}{dy})^T$ is a first order finite difference approximation of the gradient [33]. It is important to note that the TV model acts as a baseline model for CSI. Not only does it exhibit comparable performance with respect to recent state-of-the-art methods based on deep learning [35], but it also serves as a good initialization to more elaborate methods such as [36,37], which exploit non-local similarity and low-rank constraints.

5.1. CSI with side information

When the scene of interest contains fine spatial details and diverse spectral content, a single CASSI snapshot may not be sufficient to attain a desirable level of accuracy. As a result, in the last years, several dual-camera systems such the system in Fig. 6, have been proposed [16,17,38]. In this context, in addition to the primary arm, snapshot CSI systems contain a side-information camera, which captures a panchromatic or color projection of the spectral image. The extra information can be leveraged to design adaptive sparse representation reconstruction algorithms [39], leading to improved reconstructions. Also note that the relationship between the side image $Z \in \mathbb{R}^{n_1 \times n_2}$, captured by the focal plane array FPA2, and the spectral image $\{X_l\}_{l=1}^L$ can be modeled as:

$$z \approx \mathbf{R}x, \quad (5.3)$$

where $z = \text{vec}(Z)$, $x = (\text{vec}(X_1)^T, \dots, \text{vec}(X_L)^T)^T$, and $\mathbf{R}$ is such that

$$\mathbf{R} = [s_2(\lambda_1), \dots, s_2(\lambda_L)] \otimes \mathbf{I},$$

where $\mathbf{I}$ denotes the $n_1 n_2 \times n_1 n_2$ identity matrix, and $s_2(\lambda_l)$ denotes the spectral sensitivity of the FPA2 at the l th spectral band. As a result, we may add an additional constraint, i.e., $\|z - \mathbf{R}x\|_2 \leq \tilde{\epsilon}$, to the feasible set in (2.2), (2.3), and (5.2), which is referred to as using z as a direct measurement of $\{X_l\}_{l=1}^L$

Remark 5.1. In practice, the side information matrix $\mathbf{R}$ has to be estimated through a calibration process. This supposes an additional step that may be troublesome when the primary and secondary detectors do not have the same quantum efficiency as noted by the seminal work in CSI with side information [16]. In our experiments with simulated data, we study the effects on reconstruction accuracy of including the side information as a direct

measurement. But, in our experiments with real data, we don't use it as a direct measurement. As will be demonstrated, our approach does not need prior knowledge of the side information matrix $\mathbf{R}$ to benefit from the side information.

5.2. Spectral image estimation using rank-Order path graphs

We now address the CSI problem by using smoothness on rank-order path graphs. As mentioned in Section 3, a signal is the smoothest on a rank-order path graph induced by the signal itself, but of course, in this setting, we do not know the signal of interest x , and we will thus borrow the rank-order statistics from the side information image z to construct the graph.

Since the rank-order statistics of z may deviate from those of x , it is prudent to address the problem at a local level. This is a fairly common approach in imaging inverse problems [40,41]. In particular, we only use sets of local rank-order statistics from z , leading to a collection of local rank-order paths on V . To enable such a collection to be used in problem (2.3), a single graph G on V is constructed from the individual rank-order paths on V as follows:

5.2.1. Construction and aggregation of rank-order paths in CSI

Recall that a spectral image $X_1, X_2, \dots, X_L \in \mathbb{R}^{n_1 \times n_2}$ with vector representation $x \in \mathbb{R}^{n_1 n_2 L}$ such that $x = (\text{vec}(X_1)^T, \dots, \text{vec}(X_L)^T)^T$ can be assumed to reside on two-dimensional domain given by:

$$\Omega = \{1, \dots, n_1\} \times \{1, \dots, n_2\}.$$

Given a panchromatic image Z on Ω , the idea is to construct the graph G on Ω based on the local rank-order statistics of Z as follows.

Define a collection of overlapping neighborhoods Ω_k indexed by $S \subset \Omega$ such that $\Omega = \bigcup_{k \in S} \Omega_k$. Particularly, in this work, Ω_k is defined by:

$$\Omega_k = \{(i_1, i_2) \in \Omega : k_1 \leq i_1 \leq k_1 + w_1 - 1, k_2 \leq i_2 \leq k_2 + w_2 - 1\}, \quad (5.4)$$

where $w_1, w_2 \in \mathbb{N} \setminus \{0\}$ define the size of Ω_k , and S is defined by:

$$S = \{(k_1, k_2) \in \Omega : (k_1, k_2) \in S_1 \times S_2\}, \quad (5.5)$$

where $S_1 = \{1 + (i-1)ds_1\}_{i=1}^{\sqrt{|S|}}$ and $S_2 = \{1 + (i-1)ds_2\}_{i=1}^{\sqrt{|S|}}$, where ds_1 and ds_2 are selected such that adjacent Ω_k 's are overlapped to a certain degree. This is important because the collection of graphs may result disconnected otherwise, causing block artifacts during reconstruction.

For each Ω_k , perform the following procedure:

1. Find an ordering of the elements of Ω_k , i.e., $\Omega_k = \{s_i\}_{i=1}^{|\Omega_k|}$ such that the values of Z at Ω_k satisfy:

$$Z_{s_1} \leq Z_{s_2} \leq \dots \leq Z_{s_{|\Omega_k|-1}},$$

where Z_{s_i} denotes the value of Z indexed by two-dimensional coordinate associated with s_i .

2. Construct a graph $G_k = (\Omega, E_k)$ on Ω with edge set E_k given by

$$E_k = \{(s_1, s_2), (s_2, s_3), \dots, (s_{|\Omega_k|-1}, s_{|\Omega_k|})\},$$

As a last step, the collection of graphs $\{G_k\}_{k=1}^{|S|}$ are merged into a single graph G by means of the following operation:

$$G = (\Omega, \bigcup_{k=1}^{|S|} E_k). \quad (5.6)$$

To shed light on the main elements of the aforementioned graph-construction process, Fig. 7 illustrates the spectral and panchromatic images residing on Ω and the construction of rank-order paths on Ω as in Step 2. It can be observed that the rank-order statistics of Z at Ω_k for $k \in S$ lead to a collection of rank-order paths on Ω .

5.2.2. Rank-order paths (ROPs) algorithm for DC-CSI reconstruction

Given a compressive and panchromatic measurements $Y \in \mathbb{R}^{n_1 \times n_2 + L - 1}$ and $Z \in \mathbb{R}^{n_1 \times n_2}$, acquired through the DC-CSI system in Fig. 6, and provided a matrix model of the imaging system given by $\mathbf{H}$ and $\mathbf{R}$, a spectral image estimate of the scene can be obtained by measuring the smoothness of the estimate with respect to a collection of rank-order path graphs using the Algorithm 1.

Algorithm 1 ROPs.

- 1: **Input:** compressive and side-info measurements Y and Z ; primary and side-info sensing matrices $\mathbf{H}$ and $\mathbf{R}$; dimensions of the spectral image n_1, n_2, L ; noise level $\epsilon \geq 0$, and a side-information flag $w_{\text{side}} \in \{0, 1\}$.
 - 2: **Output:** spectral image estimate $\hat{x}$.
 - 3: Initialize S, w_1, w_2 .
 - 4: Obtain a graph $\tilde{G}$ by running the procedure in Sec. 5.2.1 provided Z and S, w_1, w_2 .
 - 5: **if** $w_{\text{side}} == 0$ **then**
 - 6: Set $\mathbf{A} := \mathbf{H}$, and $y := \text{vec}(Y)$.
 - 7: **else**
 - 8: Set $\mathbf{A} := [\mathbf{H}^T, \mathbf{R}^T]^T$, and $y := [\text{vec}(Y)^T, \text{vec}(Z)^T]^T$.
 - 9: **end if**
 - 10: Set $\hat{x}$ as the solution to the problem-(2.3) with G s.t. $\mathbf{L}_G := \mathbf{I}_L \otimes \mathbf{L}_{\tilde{G}}, \mathbf{A}, y$, and ϵ , where $\mathbf{I}_L$ denotes the identity matrix of size $L \times L$.
 - 11: Reshape $\hat{x}$ into a spectral image $\hat{X}_1, \dots, \hat{X}_L$.
 - 12: **Return** $\{\hat{X}_l\}_{l=1}^L$
-

We note that a side information flag w_{side} has been included in Algorithm 1 for the sake of completeness. This is because some approaches exploit side information as a direct measurement i.e., $w_{\text{side}} := 1$. However, integrating side information by means of the side information matrix $\mathbf{R}$ may be inconvenient in certain settings as explained in Section 5.1. Thus, in our experiments with real data, we use $w_{\text{side}} := 0$. As will be shown with simulated data, this setting does not affect the performance of our algorithm significantly, but instead the algorithm is able to run about 14 times faster than the case where $w_{\text{side}} := 1$.

In simulations, when $\epsilon = 0$, the problem (2.3) can be posed as a quadratic program with equality constraints, as follows which reduces to a system of linear equations [42]. Similarly, when $\epsilon > 0$, there exists $\alpha > 0$ such that the problem (2.3) can be reformulated as a particular instance of Tikhonov regularization [42]. Specifically, problem (2.3) can be reduced to the unconstrained program $\min_x \|y - \mathbf{A}x\|_2^2 + \alpha x^T (\mathbf{I}_L \otimes \mathbf{L}_{\tilde{G}})x$, whose solution satisfies:

$$(\mathbf{A}^T \mathbf{A} + \alpha (\mathbf{I}_L \otimes \mathbf{L}_{\tilde{G}}))x = \mathbf{A}^T y, \quad (5.7)$$

where $\mathbf{I}_L$ denotes an identity matrix of size $L \times L$, and $\tilde{G}$ denotes the graph with vertex set Ω obtained from the side information as in Step 4 of Algorithm 1. The regularization term $x^T (\mathbf{I}_L \otimes \mathbf{L}_{\tilde{G}})x$ can be expressed as $\sum_{l=1}^L x_l^T \mathbf{L}_{\tilde{G}} x_l$, where $x_1, \dots, x_L$ form x . The regularization term thus encourages spectral images which are sufficiently smooth with respect to $\tilde{G}$ over the set of signals defined by the measurements.

A typical approach to parameter selection is to optimize a certain metric, e.g., PSNR, SSIM, or SAM, over the space of parameters. This approach in practice may not be straightforward because we often lack of a reference datacube to compute such metrics. To

overcome this difficulty, we can rely on simulations with noiseless data using a spectral scene, hopefully with similar spatio-spectral characteristics to that of the scene of interest. In the case of our method, this would avoid the need to tune a regularization parameter, which may also be troublesome.

We have noticed that the most influential parameter to select in our method is the size of subdomain Ω_k determined by w_1, w_2 even though the degree to which adjacent Ω_k 's overlap is also relevant to avoid block artifacts. We suggest the practitioner keep the overlap constant, say between 25 and 75 percent, and play with the subdomain's size. Particularly, we assume that $w_1 = w_2$ and notice that the value of w_1 for good reconstructions with simulated data depends on the size of the spectral scene of interest as well as its spatial content. As will be evident in the experimental results, a scene of size $512 \times 512 \times 31$ may require a larger w_1 than a scene of size $256 \times 256 \times 31$. We note that, in our experiments, the same set of parameters were used for all spectral datacubes regardless of their spatio-spectral content. In the experiments with real data, however, the regularization parameter may vary depending on the signal.

5.2.3. Bilateral filter graph (BFG) algorithm for DC-CSI reconstruction

A reasonable question to ask is whether a different graph can be used for reconstruction or not. And the answer is yes indeed. In general, we need a graph on which the spectral image is sufficiently smooth. Instead of using local rank-ordering information from Z to construct a collection of paths on Ω , we can exploit, instead, distance-ordering information from Z by using a collection of bilateral-filter (BF) graphs on Ω . The BF graph arises in the context of graph signal processing by interpreting the bilateral filter as a kernel-based graph [43]. Although the BF graph is well known for image denoising and various computer vision applications, e.g., [44–47], presumably such a graph has not been used before for compressive spectral imaging. In the context of our paper, we thus have adapted it to the CSI task to provide a comparison with a different graph-based smoothness model.

Unlike in Section 5.2.1, here we set $S := \Omega$, and define $\Omega_k = \{(i_1, i_2) \in \Omega : \max\{|i_1 - k_1|, |i_2 - k_2|\} \leq r\}$, where $r \in \mathbb{N} \setminus \{0\}$ denotes the radius of Ω_k around $k \in S$. This definition of neighborhood is amenable to the fact that a bilateral filter graph has a star-like network structure, where the center vertex is given by k .

We now can adapt the graph construction procedure in Section 5.2.1 as follows:

1. Construct a graph $G_k = (\Omega, E_k, w_k)$ on Ω with edge set given by:

$$E_k = \bigcup_i \{(k, j) : j \neq k \in \Omega_k\},$$

and the weight function $w_k : \Omega \times \Omega \mapsto [0, \infty)$ given by:

$$w_k(i, j) = \exp\left(-\frac{\|i - j\|_2^2}{2\delta_s^2}\right) \exp\left(-\frac{|Z_i - Z_j|^2}{2\delta_r^2}\right)$$

for $(i, j) \in E_k$, and $w_k(i, j) = 0$ otherwise, where $\delta_s, \delta_r > 0$ are bandwidth parameters, and Z_j denote the value of Z at $j \in \Omega_k$. In our experiments, the bandwidth parameters are set as $\delta_s = 2r$, and $\delta_r = \text{mean}(\{|Z_i - Z_j|^2\}_{(i, j) \in \bigcup_k E_k})$.

As a last step, the collection of graphs $\{G_k\}_{k \in S}$ are merged into a single graph G by means of the following operation:

$$G = (V, \bigcup_{k \in S} E_k, w), \quad (5.8)$$

where the weight function $w : \bigcup_k E_k \mapsto [0, \infty)$ is defined by

$$w(i, j) = \frac{\sum_k w^{(k)}(i, j)}{|\{k \in S : w_k(i, j) > 0\}|}. \quad (5.9)$$

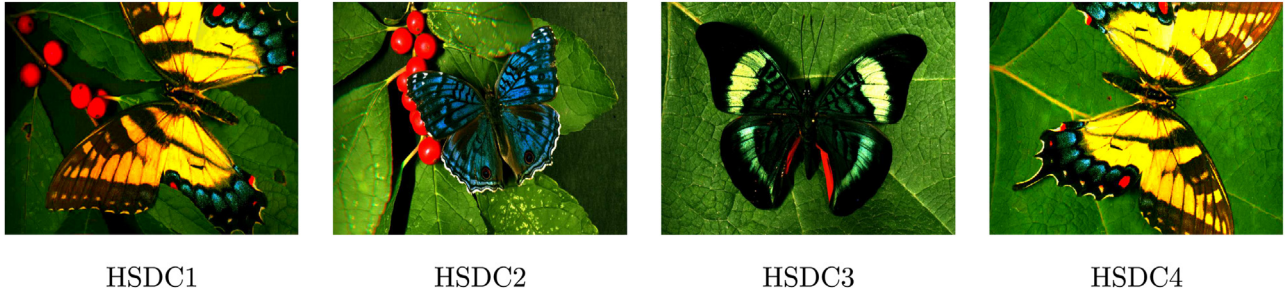


Fig. 8. UDEL hyperspectral image database. RGB renderings of four hyperspectral datacubes (HSDC) used for performance evaluation of the signal recovery algorithms. The hyperspectral datacubes HSDC1, HSDC2, HSDC3, and HSDC4 consist of 31 spectral bands of size 2064-by-3088 pixels, ranging from 400 to 700 nm.

6. Experimental results

We now evaluate the performance of the proposed graph-based model ROPs (described in Section 5.2.2) and compare it against the aforementioned traditional approaches TV and WDCT and a more recent method DeSCI [36], which are based on sparse representation and low-rank minimization. In doing so, we reconstruct a collection of spectral images of various scenes from simulated and real measurements. In the case of simulated measurements, the performance is assessed in terms of the structural similarity index (SSIM) and the spectral angle mapper (SAM), which measure the spatial and spectral fidelity of the reconstructed datacubes. In addition, we include average reconstruction and graph-construction times (R-TIME and GC-TIME). In the case of real measurements, we evaluate the performance of all approaches except the WDCT model due to its inferior performance in simulations. In this context, the spatial and spectral fidelity is judged by visual inspection of the reconstructed RGB renderings. In addition, we evaluate the reconstructed spectral signatures at certain locations with respect to the reflectance spectra measured by a non-imaging spectrometer, Ocean Optics USB2000+, brought in close proximity with the target location.

6.1. Noiseless simulated measurements

For simulated data, we considered a total of 20 hyperspectral image datacubes, consisting of four datacubes from the UDEL database, captured at our lab and shown in Fig. 8, eight datacubes selected at random from Harvard [48], and eight datacubes selected at random from ICVL [49]. Specifics about the used databases can be found in the supplementary materials. The goal is to reconstruct such datacubes from their compressive and panchromatic snapshots. For convenience, the datacubes were downsampled to the size $512 \times 512 \times L$, where $L = 31$ represents the number of frequency bands of a given datacube.

By assuming the sensing matrices $\mathbf{H}$ and $\mathbf{R}$ are known, the compressive and panchromatic images $\mathbf{y}$ and $\mathbf{z}$ of a datacube can be generated using the forward models (5.1) and (5.3). As aforementioned, $\mathbf{H}$ is set as a SD-CASSI matrix with binary random coded aperture of size 512×512 with 50% transmittance, and the spectral sensitivities of the FPA1 and FPA2 are set to be one for all wavelengths. Fig. 9 shows simulated compressive and panchromatic snapshots of the scene HSDC2 displayed in Fig. 8.

Provided the snapshots and sensing matrices, estimates of the spectral datacubes were obtained by running the reconstruction algorithms in two regimes. In the first regime, the side information signal $\mathbf{z}$ is not considered as a direct measurement, meaning that the feasible set of signals, or data-fidelity term, is determined by the CASSI's snapshot $\mathbf{y}$ and sensing matrix $\mathbf{H}$, i.e., $\{\mathbf{x} \in \mathbb{R}^{n_1 n_2 L} : \mathbf{H}\mathbf{x} = \mathbf{y}\}$. In the second regime, the side information is considered as a direct measurement, and thus the feasible set of signals is further

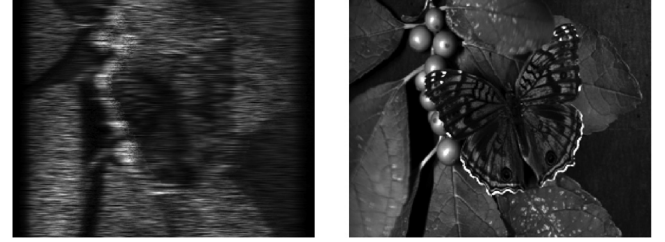


Fig. 9. Compressive and panchromatic snapshots of the HSDC2 in Fig. 8, generated by simulated sensing matrices of the DCCSI in Fig. 6.

constrained as $\{\mathbf{x} \in \mathbb{R}^{n_1 n_2 L} : \mathbf{H}\mathbf{x} = \mathbf{y}, \mathbf{R}\mathbf{x} = \mathbf{z}\}$. It is important to say that, regardless of the regime, the graph-based algorithms always rely on statistical properties (i.e., rank-order statistics, or radiometric distances) of the side signal $\mathbf{z}$ to construct the graphs.

Table 1 shows the average reconstruction performance of the above mentioned methods ROPs, BFG, TV, WDCT, and DeSCI across different datacubes. The best metrics are boldfaced. The column w/o indicates that side information is not included as a direct measurement in any of the methods, while the column w/ indicates the opposite. To reconstruct the datacubes based on the TV and WDCT models, we used a fast algorithm based on alternating direction of multipliers, referred to as C-SALSA [50], and the noise level ϵ , the maximum number of iterations MAXIT, and the tolerance TOL were set to 10^{-8} , 10000, and 10^{-6} , respectively. To reconstruct the datacubes based on the ROPs and BFG models, we used the Algorithm 1 in Section 5.2.2 and its adaptation with the bilateral-filter graph in Section 5.2.3. For ROPs, the parameters of the neighborhood Ω_k were set to $w_1 = w_2 = 9$, and adjacent Ω_k 's were set to be at most 50 percent overlapped. For BFG, the neighborhood parameter was set to $r = 3$. In step 10 of the Algorithm 1, the problem (2.3) was solved by a method based on conjugate gradients for sparse systems of linear equations, referred to as SYMMLQ [51]; here, TOL and MAXIT were also set to 10000, and 10^{-6} . For DeSCI, we tried our best to provide the DeSCI with a good initialization via GAP-TV, and ran the rank-minimization part of the algorithm using the default settings. DeSCI was run without side information due to the time it takes for reconstructions, but obviously we can expect it to improve with side information.

As indicated by the columns w/o and w/ in Table 1, when side information is used as a direct measurement, the reconstruction performance of all the methods increases, but this, in turn, increases the reconstruction complexity, leading to longer reconstruction times. In contrast, when side information is not used as a direct measurement, the reconstruction performance of all methods decreases, but this, in turn, leads to faster reconstruction time. However, note that our approach is able to maintain reconstruction accuracy in both regimes, and therefore we may want to run our algorithm in the first regime, w/o. To illustrate, consider the recon-

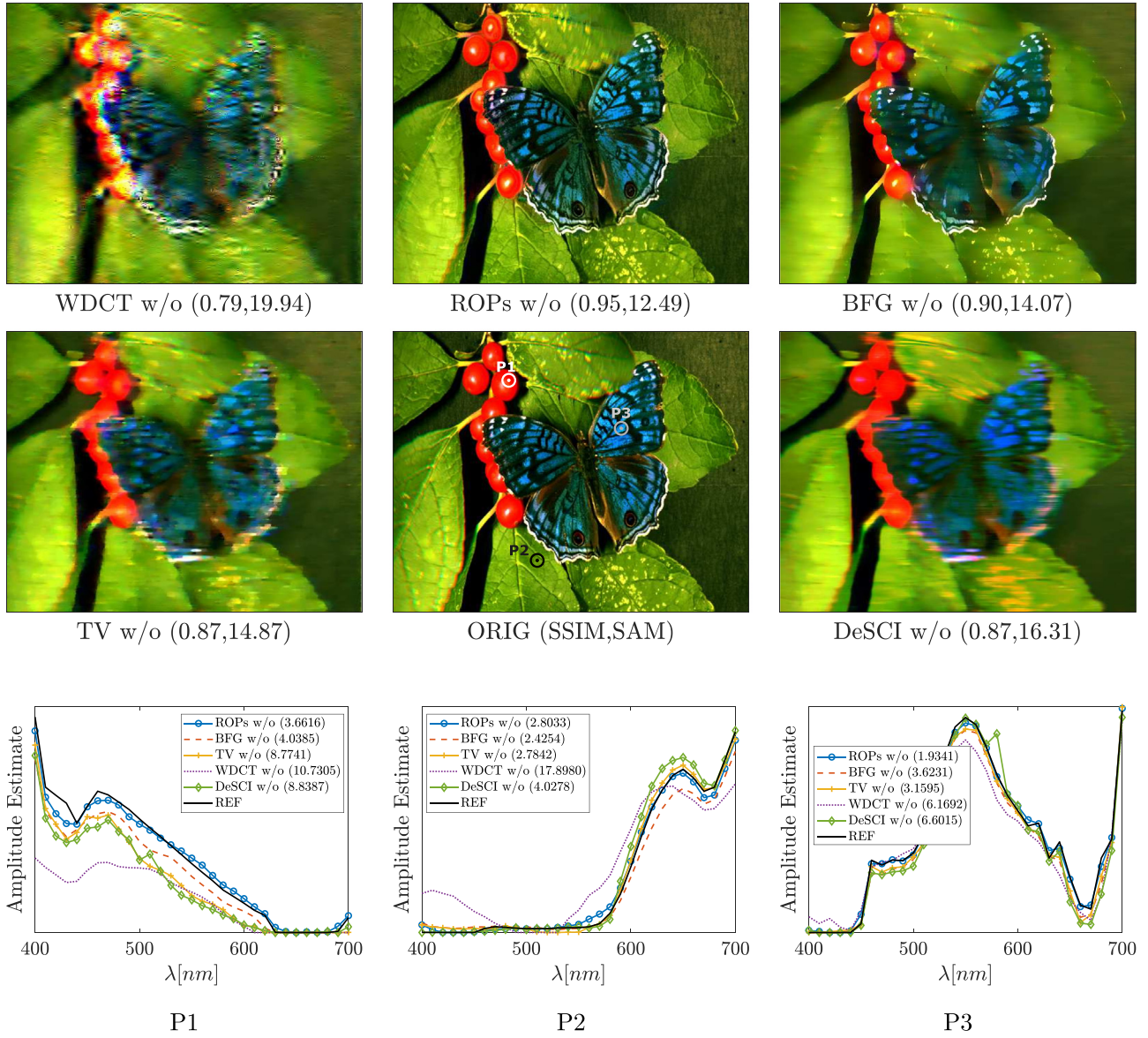


Fig. 10. Reconstructions from simulated measurements. RGB rendering of the original spectral datacube, and reconstructed RGB renderings obtained by the ROPs, BFG, TV, WDCT, and DeSCI models without side information as direct measurement. Also in the figure, original and reconstructed spectral signatures at locations P1, P2, and P3 displayed on the original RGB image. In the legends, the values in parenthesis are SAM values in degrees; the smaller the better.

Table 1

Average performance metrics of our approach based on smoothness on graphs (ROPs and BFG), and two traditional approaches based on compressive sensing and low-rank minimization (TV, WDCT, and DeSCI) across different databases UDEL(4), HRVRD(8), and ICVL(8), where the number in parenthesis denotes the number of datacubes within each database. The columns w/ and w/o indicate whether the reconstruction algorithm uses side information or not as a direct measurement of the spectral image of interest. Regardless of the column w/ and w/o, ROPs and BFG always require the side information image to construct the graphs as described in Section 5.

	SIDE INFO	ROPs		BFG		TV		WDCT		DeSCI
		w/o	w/	w/o	w/	w/o	w/	w/o	w/	w/o
UDEL(4)	SSIM	0.96	0.97	0.88	0.95	0.82	0.97	0.76	0.91	0.82
	SAM (°)	12.13	11.53	15.74	13.51	16.58	12.16	20.14	17.39	18.09
	R-TIME (min)	7.65	116.16	245.41	689.00	146.94	408.27	365.48	626.71	1214.01
	GC-TIME (sec)	1.06	1.22	196.15	191.42	-	-	-	-	-
HRVRD(8)	SSIM	0.99	0.99	0.93	0.98	0.89	0.98	0.83	0.94	0.89
	SAM (°)	2.91	2.93	3.79	4.00	4.39	3.14	7.82	7.43	5.98
	R-TIME (min)	9.31	126.57	222.81	690.56	156.60	366.20	516.58	641.20	1330.18
	GC-TIME (sec)	1.15	1.12	217.69	193.65	-	-	-	-	-
ICVL(8)	SSIM	0.98	0.98	0.89	0.97	0.81	0.98	0.76	0.94	0.81
	SAM (°)	2.35	2.26	3.22	3.02	3.89	2.46	5.39	4.69	4.82
	R-TIME (min)	7.99	107.98	181.10	679.83	150.24	379.94	486.80	772.99	1328.41
	GC-TIME (sec)	1.10	1.07	237.56	204.33	-	-	-	-	-

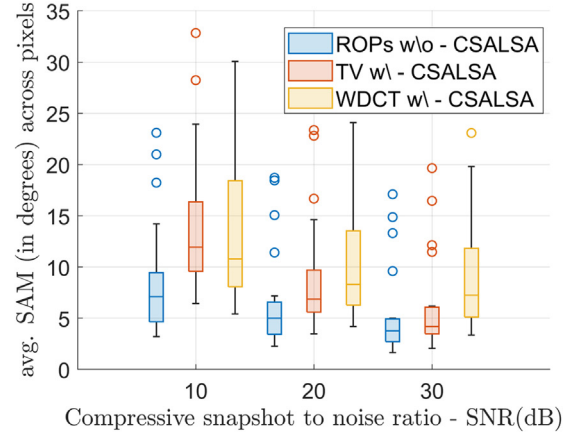
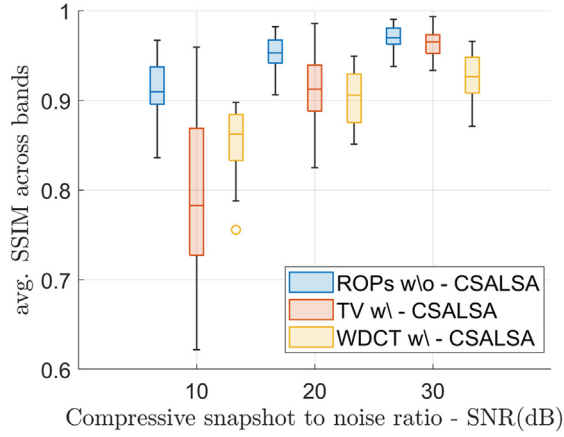


Fig. 11. Box plots of SSIM and SAM values of reconstructed spectral images for different noise levels added to the compressive measurements. The higher the SNR, the less noisy the measurements. Our method ROPs w/o exhibits significantly higher performance than the comparing approaches over the entire collection of spectral images.

structed RGB renderings in Fig. 10; here the reconstructed spectral images were obtained without side information as a direct measurement. In such a regime, we can observe that our approach still produces spectral image estimates, which preserve spatial details and contain minor spectral distortion as opposed to the traditional approaches. Note that, in the case of our approach, the SSIM value in both regimes w/o and w/ does not change significantly. This is because the graph-based prior encodes the spatial structure of the side information, and is able to transfer it to the reconstructions without relying explicitly on the side-information sensing matrix $\mathbf{R}$.

6.2. Noisy simulated measurements

We now consider a more realistic simulated scenario to show the performance of our method, ROPs w/o, and the comparing approaches, TV w/, and WDCT w/, in the presence of noisy measurements. In doing so, we reconstruct the same collection of spectral images from their compressive snapshots contaminated with additive white Gaussian noise. More precisely, let $x \in \mathbb{R}^{(512^2)31}$ represent a spectral image of interest, then its coded snapshot $y \in \mathbb{R}^{512(512+31-1)}$ satisfies $y = \mathbf{H}x + e$, where e is a noise term with entries drawn from a zero-mean Gaussian distribution with standard deviation

$$SD := \sqrt{\frac{\text{std}(\mathbf{H}x)^2}{10^{\text{SNR}/10}}},$$

where $\text{std}(\mathbf{H}x)^2$ denotes variance of entries of $\mathbf{H}x$, i.e., clean coded snapshot, and SNR denotes the signal-to-noise ratio in dBs. In our experiment, SNR takes values in the set $\{10, 20, 30\}$.

Provided z and $\mathbf{R}$ as in the previous experiment, TV w/ and WDCT w/ reconstruct x from y by minimizing $\|x\|_{\text{TV}}$ and $\|\Psi^T x\|_1$ respectively over $\{x : \|[\mathbf{H}^T, \mathbf{R}^T]^T x - [y^T, z^T]^T\|_2 \leq \epsilon\}$. While our method, ROPs w/o, reconstructs x from y by minimizing $x^T \mathbf{L}_G x$ over $\{x : \|\mathbf{H}x - y\|_2 \leq \epsilon\}$, where the graph G is constructed from z as in Section 5.2.1. Unlike in Section 6.1 that we used SYMMLQ to solve the optimization problem associated with our method, here we use the same constrained optimization solver C-SALSA to solve all the optimization problems. We should note that even though our approach can be solved much more efficiently by reformulating the problem as an unconstrained problem as will be done in the experiments with real measurements, here it is convenient to work with the constrained formulation because the extent of the feasible set can be defined as $\epsilon := SD\sqrt{m} + \sqrt{2m}$ as explained in Section 4, thus avoiding the search for a regularization parameter.

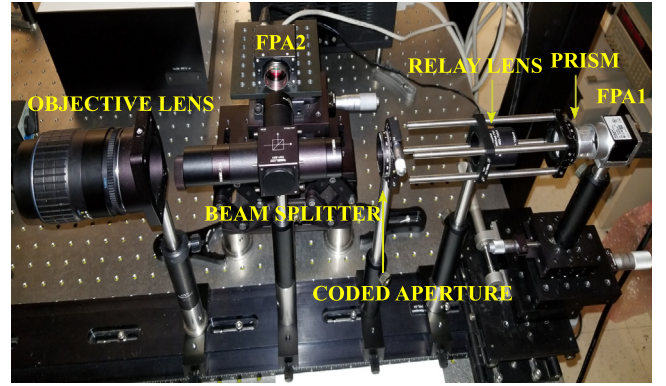


Fig. 12. Dual-camera compressive spectral imager.

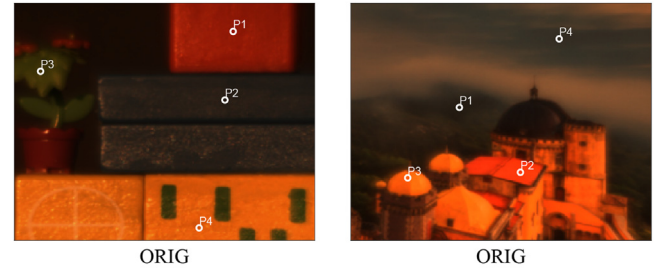


Fig. 13. Color images of the spectral scenes of interest. The color images were captured by the side-information camera in Fig. 12 under the same illumination conditions used for the CASSI snapshot. The spectral signatures at the points P1, P2, P3, and P4 will be evaluated.

Fig. 11 summarizes the SSIM and SAM values for different levels of noise added to the measurements, where the variability of the metrics is due to the number of spectral images considered for reconstruction. As indicated by Fig. 11(left), our method obtains SSIM values, which are not only higher but less variable than the comparing approaches. Similarly, as indicated by Fig. 11(right), our method obtains SAM values that are smaller and less variable than the comparing methods. The figures thus exhibit that our approach produces consistent spectral image estimates with high spatial fidelity and low spectral distortion across different noise levels for various spectral scenes.

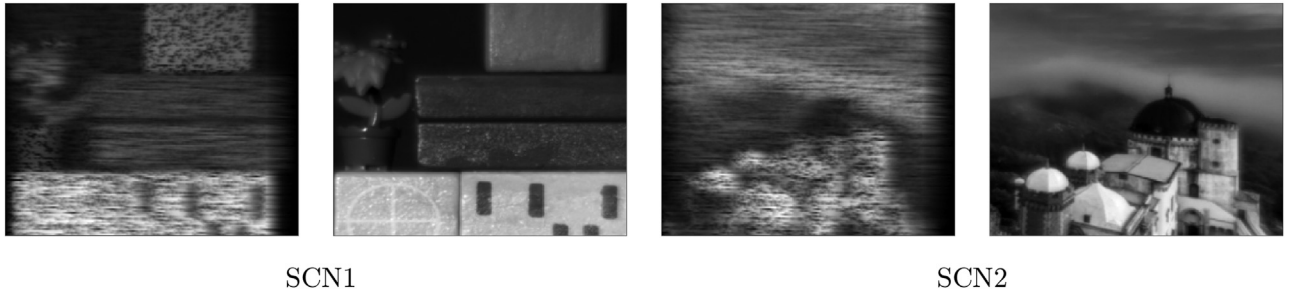


Fig. 14. Compressive and panchromatic snapshots of the scenes SCN1, SCN2 as seen by our DCCSI system Fig. 12. The snapshots of SCN3, and SCN4 can be found in the supplementary materials.

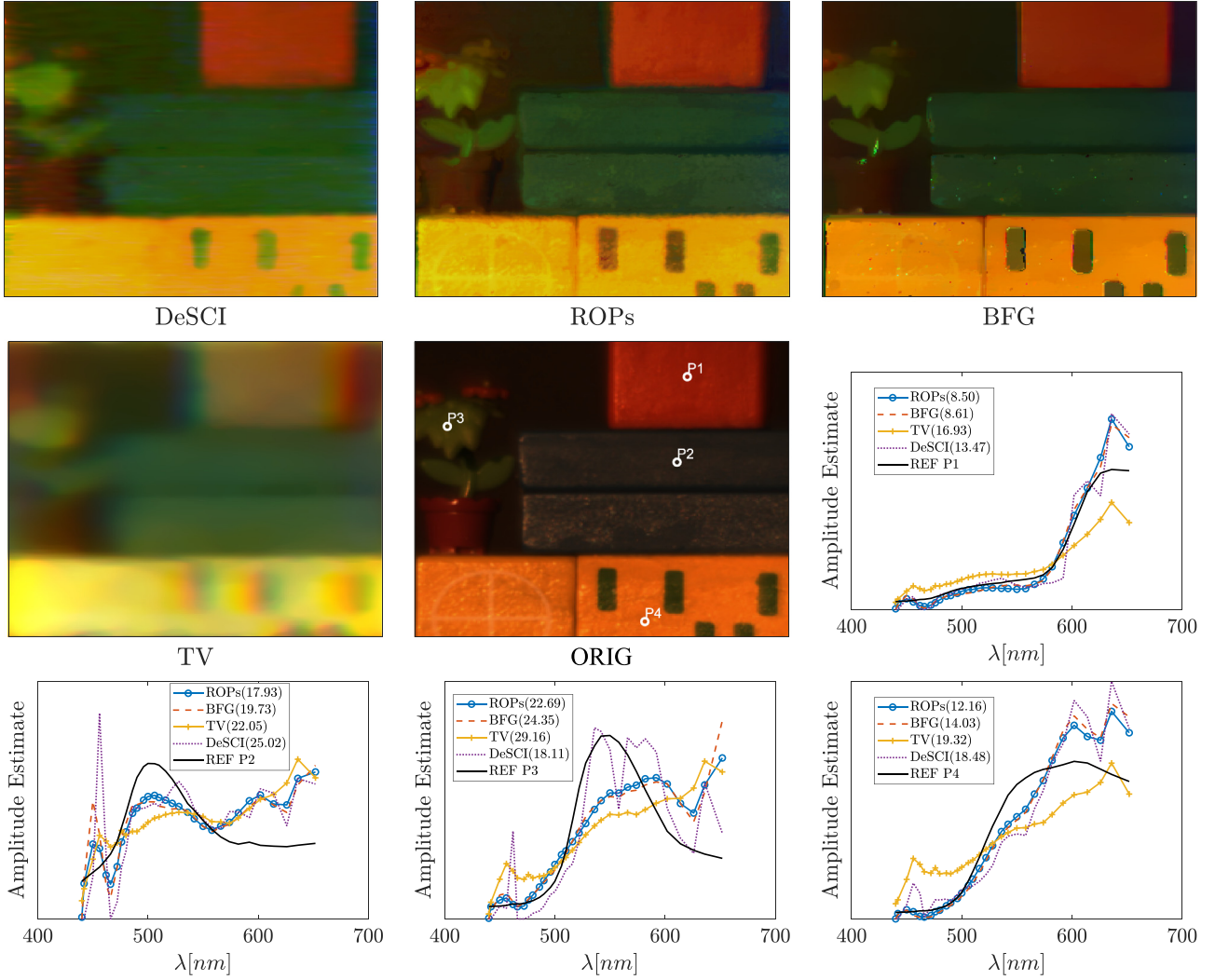


Fig. 15. Reconstructions from real measurements. RGB image of the original scene obtained by a color camera, and reconstructed RGB renderings obtained by the ROPs, BFG, TV, and DeSCI models without side information as direct measurement. Also in the figure, reconstructed spectral signatures at a few locations P1, P2, P3, and P4 displayed in ORIG. In the legends, the values in parenthesis indicate the accuracy of the reconstruction in terms of the SAM metric in degrees. The average values over the four signatures are **15.32** for ROPs, 16.76 for BFG, 20.40 for TV, and 18.77 for DeSCI; the smaller, the better.

6.3. Real measurements

For real data, we implemented the dual-camera compressive spectral imaging (DC-CSI) system in Fig. 12 whose schematic is displayed in Fig. 6. We then configured the DC-CSI system to encode 31 spectral bands of size 256×256 from a scene of interest onto a compressive snapshot of size 256×286 , and a panchromatic snapshot of size 256×256 . The interested reader is referred to the sup-

plementary materials for implementation details of our DC-CSI system. Figs. 13 and 14 show the scenes of interest, and some compressive and panchromatic snapshots as seen by our DC-CSI system in Fig. 12.

Since the primary and secondary cameras of our DC-CSI system have different quantum efficiencies, using side information as a direct measurement may not be straightforward as discussed in Section 5.1. Thus, the reconstruction algorithms ROPs, BFG, TV, and

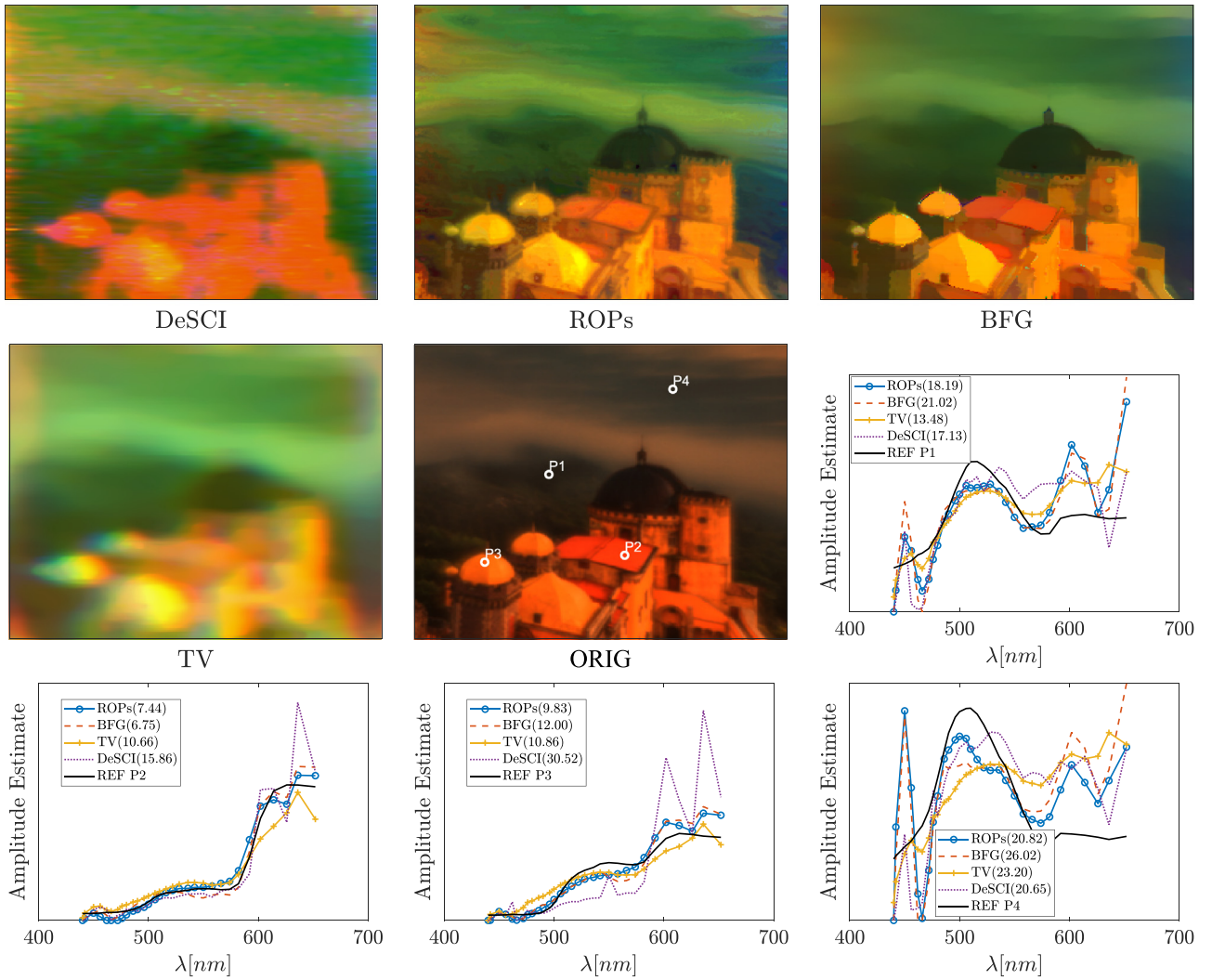


Fig. 16. Reconstructions from real measurements. Reconstructed RGB renderings obtained by the ROPs, BFG, TV, and DeSCI models without side information as direct measurement. Also in the figure, reconstructed spectral signatures at a few locations P1, P2, P3, and P4 displayed in ORIG. In the legends, the values in parenthesis indicate the accuracy of the reconstruction in terms of the SAM value in degrees. The average values over the four signatures are **13.14** for ROPs, 16.44 for BFG, 14.55 for TV, and 21.04 for DeSCI; the smaller, the better.

DeSCI were only run in the first regime; side information is not included as a direct measurement. That is, the feasible set of signals, or data-fidelity term, is determined by $\{\mathbf{x} \in \mathbb{R}^n : \|\hat{\mathbf{H}}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}$, where the CASSI's sensing matrix $\hat{\mathbf{H}}$ was estimated by a calibration procedure similar to that proposed in [27]. We should note that, in this setting, the graph-based algorithms are expected to produce spectral image estimates with better spatial resolution than the comparing approaches since they leverage the statistical properties of the panchromatic data for reconstruction.

Color renderings of the scenes (SCN1 and SCN2) and spectral signatures reconstructed by ROPs, BFG, TV, DeSCI can be observed in Figs. 15, and 16. Unlike the TV, DeSCI, and BFG approaches, the proposed graph-based approach ROPs produces color images that not only contain as much spatial details as the associated panchromatic images but also display almost as many colors as the color images in Fig. 13. Even though BFG preserves sharp edges on relatively large spatial structures, it smooths out small details on the scene, leading to piece-wise constant approximations. In contrast, ROPs preserves better both large and small details.

In the figures, we can also observe reconstructed and reference spectral signatures at four different location P1, P2, P3, and P4. For comparison purposes, such signatures have been normalized

by their total intensity. To clarify, let $X_1(i, j), X_2(i, j), \dots, X_L(i, j)$ be a spectral signature at location $(i, j) \in \Omega$. The normalized spectral signature is given by

$$\tilde{X}_l(i, j) = \frac{X_l(i, j)}{\sum_{p=1}^L X_p(i, j)}, \quad l = 1, \dots, L.$$

In the figure's legends, the SAM value of the reconstructed signature with respect to the reference is displayed in parenthesis for each approach. As suggested by the average SAM value over the four spectral signatures, the proposed method ROPs outperforms the other methods on average. Specifically, the methods ROPs, BFG, TV, and DeSCI reach average SAM values of 15.32, 16.76, 20.40, and 18.77 for SCN1 in Fig. 15 and 13.14, 16.44, 14.55, and 21.04 for SCN2 in Fig. 16, respectively. In addition, Figs. 17 and 18 display the intensity estimates of the scenes across the spectrum. Although the graph-based estimates exhibit high intensity artifacts at the image boundaries and so do the comparing approaches, in particular at the shorter wavelengths, they perform significantly better than the competing approaches, which do not exploit structural information from the side information.

Regarding parameter settings, for all algorithms, we set the parameter TOL to 10^{-3} . We noticed that the stability of the solutions

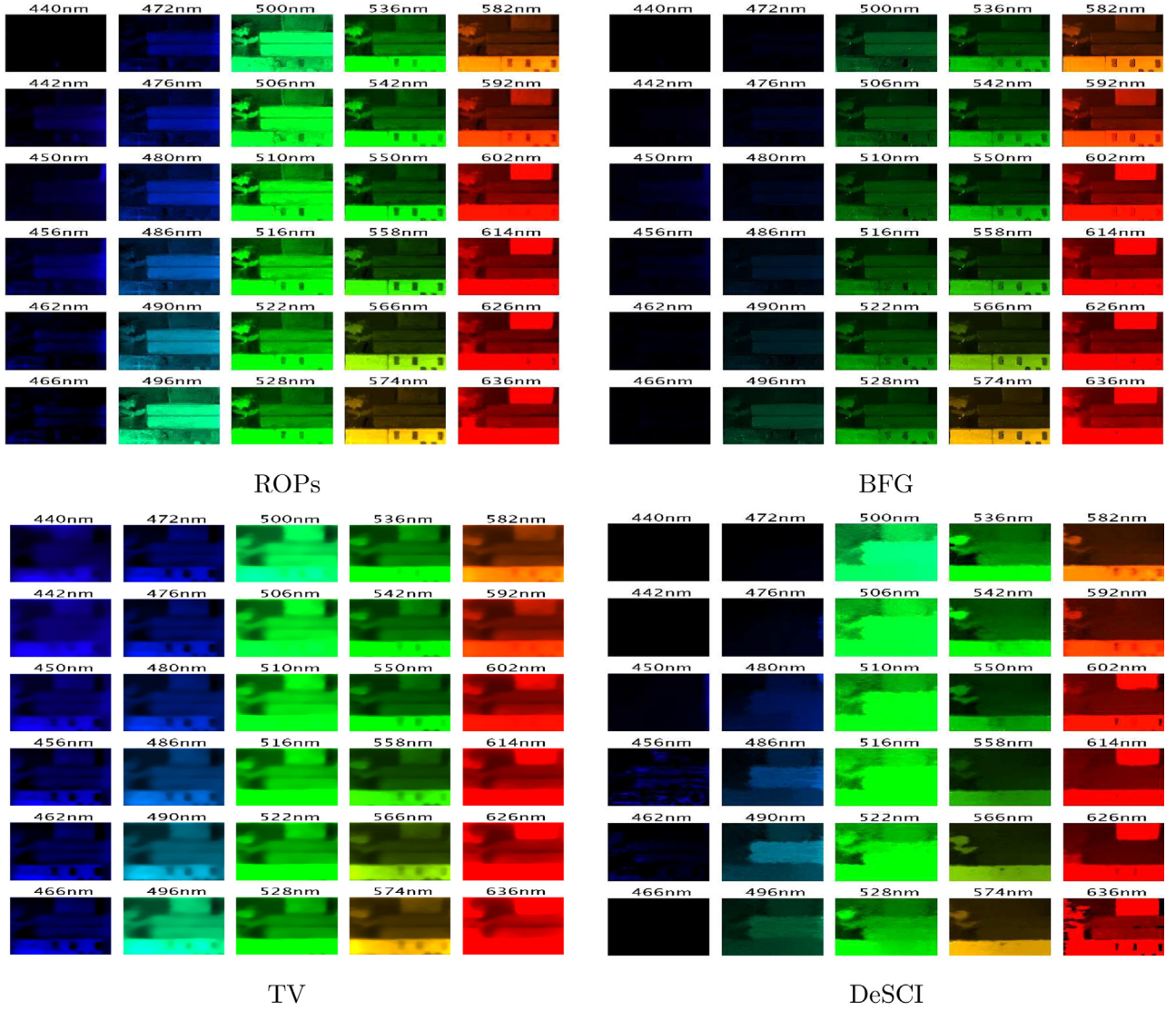


Fig. 17. Spectral image estimates of the scene SCN 1 in Fig. 13. We display 30 out of 31 spectral bands of size 256 by 256 obtained by the graph-based approaches ROPs, BFG, and the comparing approaches TV, and DeSCI.

may be affected by setting this parameter too small, as we did in the simulated data case. For ROPs, the parameters of the neighborhood Ω_k were set to $w_1 = w_2 = 7$, and adjacent Ω_k 's were set to be at most 66 percent overlapped. For BFG, the neighborhood parameter was set to $r = 2$, and the bandwidth parameters were set to $\delta_s = 2r$, and $\delta_r/3$, where δ_r is set as suggested in Section 5.2.3. Additionally, the regularization parameter α in (5.7) was selected by a grid search over $\alpha > 0$. To that end, we obtained first an estimate x_α for each α , and recorded the values of the data fidelity term $\|\hat{\mathbf{H}}x_\alpha - y\|_2^2$ and regularization term $x_\alpha^T \mathbf{L}_G x_\alpha$. Last, we selected as regularization parameter the one at which the curves given by $(\alpha, \frac{x_\alpha^T \mathbf{L}_G x_\alpha}{\max_\alpha (x_\alpha^T \mathbf{L}_G x_\alpha)})$ and $(\alpha, \frac{\|\hat{\mathbf{H}}x_\alpha - y\|_2^2}{\max_\alpha (\|\hat{\mathbf{H}}x_\alpha - y\|_2^2)})$ intersected; specifically, for ROPs $\alpha = 2.9764$ for both SCN1 and SCN2, and for BFG $\alpha = 1.2743$ for both SCN1 and SCN2. For TV, we defined the noise level parameter $\epsilon > 0$ in (5.2) by the same procedure; specifically, $\epsilon = 9.0418$ for SCN1 and $\epsilon = 7.0330$ for SCN2. The DeSCI's parameters were adjusted as in the case of simulated measurements.

7. Discussion

In this paper, we introduced the concept of rank-order path graphs and examined the problem of signal recovery from under-

sampled linear measurements by using smoothness with respect to rank-order path graphs. In particular, we showed the usefulness of rank-order path graphs for compressive spectral imaging, and demonstrated its advantages over some traditional approaches. Despite having a relatively simple sparse edge structure, rank-order path graphs are capable of capturing structural details in the data, leading to efficient graph representations, and fast iterative inversion. The key idea is that by rearranging the original feature space, the complexity of the rearranged space can be captured by using fundamental graphs, i.e., path graphs.

We note that all methods have an inherent limitation that may affect their reconstruction accuracy. In particular, the data fidelity term $\|y - \hat{\mathbf{H}}x\|_2^2$ is well suited to handle additive Gaussian noise in y , but in practice, the statistical nature of the noise affecting y is far from additive Gaussian. There is non-linear noise due to quantization and non-uniform illumination of the scene besides signal-dependent noise due to the approximate nature of the sensing matrix. Consequently, we should not have a perfect match between our reconstructed spectral signatures and the reference spectra measured by the non-imaging spectrometer. In our experiment, however, some fixes may lead to overall improvements. First, the CASSI matrix is derived based on the assumption that the coded

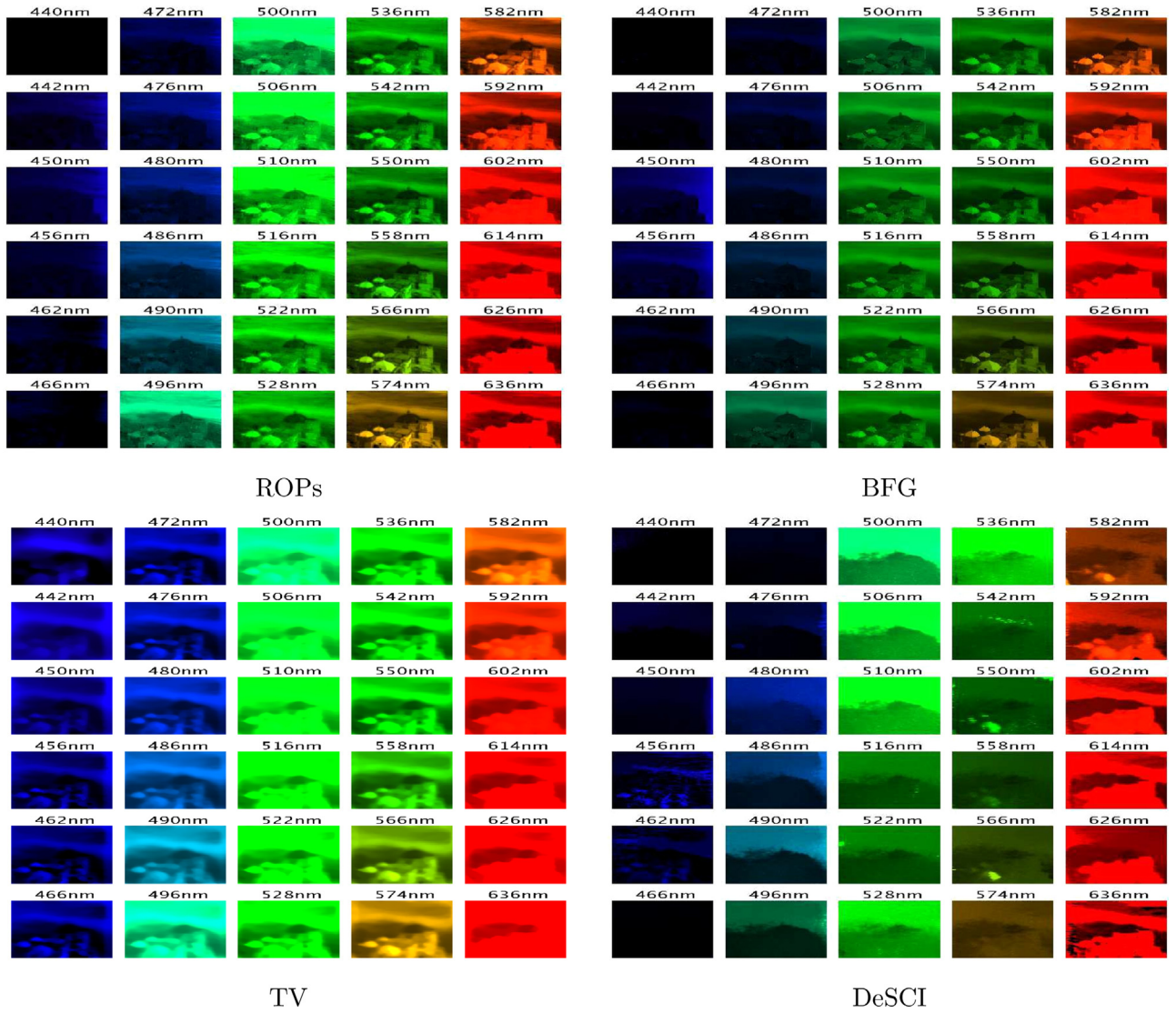


Fig. 18. Spectral image estimates of the scene SCN 2 in Fig. 13. We display 30 out of 31 spectral bands of size 256 by 256 obtained by the graph-based approaches ROPs, BFG, and the comparing approaches TV, and DeSCI.

aperture's pixel size and detector's pixel size are the same, and this assumption does not hold in our experiment. This is because 128×128 coded aperture pixels map onto 1024×1024 detector pixels, which degrades the quality of our compressive measurement. We may thus enhance spectral reconstruction by replacing the coded aperture with a higher resolution one. Second, on the calibration front, more careful estimation of the shift introduced by the prism to the coded aperture on the detector and the use of a calibration source, i.e., monochromator and light guide, with a more uniform illumination pattern may also lead to improved reconstructions.

The fact that a one-dimensional projection (e.g. a panchromatic image) from a high-dimensional object (or a hyperspectral image) can be used to rank order its elements smoothly appears to be rooted in the theory of manifolds. In our application scenario, we could say that if the hyperspectral image lives close to a low dimensional linear manifold, then the ordering information from the panchromatic image is useful to infer a smooth multivariate ordering of the high dimensional vectors (or spectral signatures). We are currently exploring this hypothesis.

In recent years, the idea of patch ordering has been used as a means to regularizing various image restoration problems, e.g,

[52] and some references therein. Specifically, the method by Vaksman et al. [52] builds on the assumption that a patch-based permutation obtained from a noisy image induces a smooth reordering of the associated pixels in the original image. Consequently, a restored image can be obtained by encouraging image estimates with sparse second derivative along the linear order of the pixels induced by the permutation. A simplified version of their regularization term is given by:

$$\mathcal{R}(x) = \|\mathbf{M}\mathbf{L}_G\mathbf{P}x\|_1,$$

where $\mathbf{M}$ denotes a weight diagonal matrix, $\mathbf{L}_G$ denotes the graph Laplacian of a path graph G , and $\mathbf{P}$ denotes the patch-based permutation matrix inferred from the noisy image.

In our context, the assumption that the signal of interest is sufficiently smooth with respect to a path graph, leads fundamentally to regularization terms of the form:

$$\mathcal{R}(x) = \|\mathbf{L}_G^{1/2}\mathbf{P}x\|_2^2,$$

where G is the path graph with edge set $\{(i, i+1)\}_{i=1}^n$, and $\mathbf{P}$ is computed from the rank-order statistics of the panchromatic image.

Even though both regularization terms bear a similarity, not only do they lead to different solutions with different characteristics, but our regularization term is simpler to minimize. It is important to say, however, that the work by Vaksman et al. [52] can indeed be instrumental to extend our approach to a broader class of compressive spectral imaging (CSI) systems even if we do not have access to a panchromatic side information image. We note that several other CSI systems, e.g., [53–55], capture compressive images that are co-registered with the spectral image of interest and look almost like a panchromatic image with random coded noise. When a co-registered compressive image is available, unlike in SD-CASSI, it may be possible to estimate the rank-order information of a panchromatic image from the compressive image itself. As a result, a suitable set of rank-order path graphs can be used for CSI reconstruction without an additional camera. Of course, the statistics of the noise may deviate quite a lot from those of traditional image restoration problems. Thus, we may need to adapt significantly to the available methods or invent new ones.

If there is, on the other hand, access to a panchromatic image, which is aligned to the spectral image of interest, since our method does not rely on the side information sensing matrix to integrate such a panchromatic image in the reconstruction process, our method can be easily used to regularize a variety of recent spectral imaging systems, e.g., [56–60], without increasing calibration complexity nor adding significant hardware design considerations.

Rank-order path graphs could be extended to applications such as depth map restoration [61,62], positron emission tomography enhancement [63], and spectral X-Ray computer tomography reconstruction [64] to name a few, where side information is readily available. To construct the graphs, however, we should not use rank ordering information directly extracted from the side information itself. Instead, rank orderings have to consider the geometrical aspects of the multimodal feature space [65–67]. To do so, the notion of multivariate rank-order statistics presented in Section 3 can be explored.

Adaptive compressive spectral imaging architectures may benefit from the use of rank-order path graphs because not only does its construction take little computational overhead but their edge structure is also highly sparse. To do so, we can show that the error accuracy is inversely related to the smallest eigenvalue of $(\mathbf{A}^T \mathbf{A} + \alpha(\mathbf{I}_L \otimes \mathbf{L}_C))$ in (5.7). As a result by allowing the coded aperture to modify the distribution of zeros and ones in the matrix $\mathbf{A}$ such that the smallest eigenvalue is maximized, we may obtain richer spatio-spectrally coded measurements and therefore better reconstructions than random coded apertures. As suggested by the recent magazine in snapshot spectral imaging this is a direction worth exploring [14].

We conclude by saying that rank-order path graphs have the potential to become a valuable asset not only in modern-day computer vision and computational imaging problems, but also as models to learn efficient graph representations of data.

Declaration of Competing Interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

CRedit authorship contribution statement

Juan F. Florez-Ospina: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Daniel L. Lau:** Conceptualization, Resources, Writing – review & editing, Supervision. **Dominique Guillot:** Conceptualization, Formal analysis, Writing – review & editing. **Kenneth Barner:** Conceptualization, Resources, Writing – review & editing, Funding acquisition. **Gonzalo R. Arce:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants NSF 1815992 and NSF 1816003. Juan F. Florez thanks Fulbright Colombia and Ministry of Science, Technology and Innovation of Colombia for his doctoral fellowship. Also, Juan thanks Hoover Rueda and Alejandro Parada for insightful technical discussions and their important role as mentors.

Appendix A. Proof of Theorem 3.5

The proof of Theorem 3.5 depends on the following lemma.

Lemma A.1. Let $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$. Then

$$\min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n-1} (x_{\sigma_{i+1}} - x_{\sigma_i})^2 = \sum_{i=1}^{n-1} (x_{r_{i+1}^{-1}} - x_{r_i^{-1}})^2,$$

where $r^{-1} \in \mathcal{S}_n$ is such that $x_{r_1^{-1}} \leq x_{r_2^{-1}} \leq \dots \leq x_{r_n^{-1}}$.

Proof. Replacing $x = (x_1, x_2, \dots, x_n)^T$ by $(x_{r_1^{-1}}, x_{r_2^{-1}}, \dots, x_{r_n^{-1}})^T$, we may assume without loss of generality that $x_1 \leq x_2 \leq \dots \leq x_n$. Let $\sigma \in \mathcal{S}_n$ be a permutation of $\{1, \dots, n\}$ such that $\sigma = \sigma_1 \sigma_2 \dots \sigma_n$, and $1 \leq i \leq n-1$.

Choose $\sigma_i, \sigma_{i+1} \in \{\sigma_k\}_{k=1}^n$ such that $x_{\sigma_i} \leq x_{\sigma_{i+1}}$. For $\beta \in [x_{\sigma_i}, x_{\sigma_{i+1}}]$, it is easy to see that $(x_{\sigma_{i+1}} - x_{\sigma_i})^2 \geq (x_{\sigma_{i+1}} - \beta)^2$. More generally, let $t_{i+1} := \max(\sigma_{i+1}, \sigma_i)$. Since by hypothesis $x_{t_{i+1}} \geq x_{t_{i+1}-1}$ and $x_{t_{i+1}-1}$ belongs in the interval formed by $x_{\sigma_{i+1}}, x_{\sigma_i}$, we have that $(x_{\sigma_{i+1}} - x_{\sigma_i})^2 \geq (x_{t_{i+1}} - x_{t_{i+1}-1})^2$. Summing over i , we conclude that

$$\sum_{i=1}^{n-1} (x_{\sigma_{i+1}} - x_{\sigma_i})^2 \geq \sum_{i=1}^{n-1} (x_{t_{i+1}} - x_{t_{i+1}-1})^2.$$

Therefore, the left hand side reaches a minimum over $\mathcal{S}_n$ at the permutation $t_1 t_2 \dots t_n$, which is equivalent to the permutation that sorts the entries of x in ascending order, i.e., $r^{-1} = r_1^{-1} r_2^{-1} \dots r_n^{-1}$. This concludes the proof. $\square$

Proof of Theorem 3.5.. Recall that for any graph $G = (V, E)$ and any vector $z = (z_1, \dots, z_n)^T \in \mathbb{R}^n$, we have

$$z^T \mathbf{L}_G z = \sum_{(i,j) \in E(G)} (z_i - z_j)^2.$$

By definition, for $G \in \mathcal{L}_n$, there is a permutation $\sigma \in \mathcal{S}_n$ such that $E = \{(\sigma_1, \sigma_2), (\sigma_2, \sigma_3), \dots, (\sigma_{n-1}, \sigma_n)\}$, and we have $x^T \mathbf{L}_G x = \sum_{(i,j) \in \{(\sigma_k, \sigma_{k+1})\}_{k=1}^{n-1}} (x_i - x_j)^2 = \sum_{k=1}^{n-1} (x_{\sigma_{k+1}} - x_{\sigma_k})^2$. Thus,

$$\min_{G \in \mathcal{L}_n} x^T \mathbf{L}_G x = \min_{\sigma \in \mathcal{S}_n} \sum_{k=1}^{n-1} (x_{\sigma_{k+1}} - x_{\sigma_k})^2.$$

The result now follows from Lemma Appendix A.1. $\square$

Appendix B. Proof of Proposition 3.6

The proof of Proposition 3.6 depends on the following lemma.

Lemma B.1. Let $G = (V, E, W)$ be an undirected weighted graph on $V = \{1, \dots, n\}$, and let $X \in \mathbb{R}^{n \times L}$. Then

$$\text{tr}(X^T L_G X) = \sum_{(i,j) \in E} W_{ij} \|x_i - x_j\|_2^2,$$

where $x_i = (x_{i1}, x_{i2}, \dots, x_{iL})$ denotes the i th row of X .

Proof. Let x_l denote the l th column of X . By definition of $\text{tr}()$, we have that $\text{tr}(X^T L_G X) = \sum_{l=1}^L x_l^T L_G x_l$. Since $x_l^T L_G x_l$ can be expressed as $\sum_{(i,j) \in E} W_{ij} (x_{il} - x_{jl})^2$, where x_{il} denotes the i th entry of x_l , we obtain $\text{tr}(X^T L_G X) = \sum_{l=1}^L \sum_{(i,j) \in E} W_{ij} (x_{il} - x_{jl})^2$. Interchanging the order of the summations, we get

$$\text{tr}(X^T L_G X) = \sum_{(i,j) \in E} W_{ij} \sum_{l=1}^L (x_{il} - x_{jl})^2 = \sum_{(i,j) \in E} W_{ij} \|x_i - x_j\|_2^2,$$

as claimed. $\square$

Proof of Proposition 3.6. Applying Lemma Appendix B.1, we obtain

$$\min_{G \in \mathcal{L}_n} \text{tr}(X^T L_G X) = \min_{G \in \mathcal{L}_n} \sum_{(i,j) \in E} \|x_i - x_j\|_2^2.$$

Observe that each permutation $\sigma = \sigma_1 \sigma_2 \dots \sigma_n \in \mathcal{S}_n$ can be associated with a path graph $G = (V, E) \in \mathcal{L}_n$ whose edge set is given by

$$E = \{(\sigma_i, \sigma_{i+1}) : i = 1, \dots, m-1\}.$$

Therefore, we can optimize over the set of permutations $\mathcal{S}_m$ instead of $\mathcal{L}_m$, and this completes the proof. $\square$

Appendix C. Proof of Theorem 3.7

The proof of Theorem 3.7 depends on the following definition and lemma.

Definition C.1. The graphs G and H are *isomorphic* if there exist a bijective map $\tau : V(G) \mapsto V(H)$ such that

$$(i, j) \in E(G) \text{ iff } (\tau(i), \tau(j)) \in E(H).$$

Lemma C.2. The graphs G and H are isomorphic if and only if their adjacency matrices are related by

$$W_G = P^T W_H P$$

for some permutation matrix P .

Proof. See [68] Theorem 4.2.1. $\square$

Proof of Theorem 3.7. By definition, $G \in \mathcal{L}_n$ has edge set $E(G)$ given by $\{(\sigma_k, \sigma_{k+1}) : k = 1, \dots, n-1\}$ for some permutation $\sigma \in \mathcal{S}_n$. Note that the inverse permutation σ^{-1} of σ can be regarded as a bijection from V to V , and satisfies

$$(i, j) \in E(G) \text{ iff } (\sigma^{-1}(i), \sigma^{-1}(j)) \in E(\tilde{G}).$$

This is because (i, j) can be written as (σ_k, σ_{k+1}) for some $k \in \{1, \dots, n-1\}$, and thus $(\sigma^{-1}(i), \sigma^{-1}(j)) = (k, k+1) \in E(\tilde{G})$. Therefore G and $\tilde{G}$ are isomorphic.

By applying Lemma Appendix C.2 with G and $\tilde{G}$, we obtain that $W_G = P^T W_{\tilde{G}} P$ for some permutation matrix P . In particular, P can be constructed from σ as follows. Recall that $P^T = P^{-1}$, we thus have that the adjacency matrix $W_{\tilde{G}}$ of $\tilde{G}$ satisfies $W_{\tilde{G}} = P W_G P^T$. By assumption, the edge set of $\tilde{G}$ is given by $\{(k, k+1) : k = 1, \dots, n-$

$1\}$. So, we have that $e_k^T W_{\tilde{G}} e_{k+1} = e_k^T P W_G P^T e_{k+1} = 1$, where e_k denotes k th standard basis vector. This shows that $P^T e_k = e_{\sigma_k}$ and $P^T e_{k+1} = e_{\sigma_{k+1}}$, and we thus obtain

$$P^T \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} = \begin{bmatrix} e_{\sigma_1} & e_{\sigma_2} & \dots & e_{\sigma_n} \end{bmatrix} \Rightarrow P = \begin{bmatrix} e_{\sigma_1}^T \\ e_{\sigma_2}^T \\ \vdots \\ e_{\sigma_n}^T \end{bmatrix}.$$

Now, it remains to show the relationship between the graph Laplacians of $\tilde{G}$ and G . By definition, $L_{\tilde{G}} = D_{\tilde{G}} - W_{\tilde{G}}$. So, we have that $P^T L_{\tilde{G}} P = P^T D_{\tilde{G}} P - W_G$. The result follows by verifying that the degree matrix of G is given by $P^T D_{\tilde{G}} P$. To do so, note that the degree of the vertex $\sigma_k \in V(G)$ is $e_{\sigma_k}^T D_G e_{\sigma_k} = e_{\sigma_k}^T P^T D_{\tilde{G}} P e_{\sigma_k} = e_k^T D_{\tilde{G}} e_k$ since $P e_{\sigma_k} = e_k$. As a consequence, the degree of the vertex σ_k can be computed exactly from $e_{\sigma_k}^T P^T D_{\tilde{G}} P e_{\sigma_k}$ and this concludes the proof. $\square$

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.sigpro.2022.108707](https://doi.org/10.1016/j.sigpro.2022.108707).

References

- [1] Y.C. Eldar, G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.
- [2] S.V. Venkatakrishnan, C.A. Bouman, B. Wohlberg, Plug-and-play priors for model based reconstruction, in: 2013 IEEE Global Conference on Signal and Information Processing, IEEE, 2013, pp. 945–948.
- [3] Y. Romano, M. Elad, P. Milanfar, The little engine that could: regularization by denoising (RED), SIAM J. Imaging Sci. 10 (4) (2017) 1804–1844.
- [4] E.T. Reehorst, P. Schniter, Regularization by denoising: clarifications and new interpretations, IEEE Trans. Comput. Imaging 5 (1) (2018) 52–67.
- [5] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep image prior, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.
- [6] A. Ortega, P. Frossard, J. Kovačević, J.M.F. Moura, P. Vanderghenst, Graph signal processing: overview, challenges, and applications, Proc. IEEE 106 (5) (2018) 808–828.
- [7] M. Rossi, P. Frossard, Geometry-consistent light field super-resolution via graph-based regularization, IEEE Trans. Image Process. 27 (9) (2018) 4207–4218.
- [8] Y. Bai, G. Cheung, X. Liu, W. Gao, Graph-based blind image deblurring from a single photograph, IEEE Trans. Image Process. 28 (3) (2018) 1404–1418.
- [9] F. Mahmood, N. Shahid, U. Skoglund, P. Vanderghenst, Adaptive graph-based total variation for tomographic reconstructions, IEEE Signal Process. Lett. 25 (5) (2018) 700–704.
- [10] B.A. Brooksby, S. Jiang, H. Dehghani, B.W. Pogue, K.D. Paulsen, J.B. Weaver, C. Kogel, S.P. Poplack, Combining near-infrared tomography and magnetic resonance imaging to study in vivo breast tissue: implementation of a Laplacian-type regularization to incorporate magnetic resonance structure, J. Biomed. Opt. 10 (5) (2005) 051504.
- [11] S.-S. Yang, Concomitants of Order Statistics, Iowa State University, 1976.
- [12] N. Balakrishnan, C.R. Rao, N. Limnios, C. Papadopoulos, A.P. Basu, S.E. Rigdon, C.D. Lai, M. Xie, D. Murthy, A. Chatterjee, et al., Handbook of statistics, Order 16 (1998) 3–731.
- [13] G.R. Arce, D.J. Brady, L. Carin, H. Arguello, D.S. Kittle, Compressive coded aperture spectral imaging: an introduction, IEEE Signal Process. Mag. 31 (1) (2014) 105–115.
- [14] X. Yuan, D.J. Brady, A.K. Katsaggelos, Snapshot compressive imaging: theory, algorithms, and applications, IEEE Signal Process. Mag. 38 (2) (2021) 65–88.
- [15] M.-Y. Chen, F. Renna, M.R.D. Rodrigues, Compressive sensing with side information: how to optimally capture this extra information for GMM signals? IEEE Trans. Signal Process. 66 (9) (2018) 2314–2329.
- [16] X. Yuan, T.-H. Tsai, R. Zhu, P. Lull, D. Brady, L. Carin, Compressive hyperspectral imaging with side information, IEEE J. Sel. Top. Signal Process. 9 (6) (2015) 964–976.
- [17] L. Galvis, D. Lau, X. Ma, H. Arguello, G.R. Arce, Coded aperture design in compressive spectral imaging based on side information, Appl. Opt. 56 (22) (2017) 6332–6340.
- [18] E.J. Candes, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.
- [19] S. Chen, A. Sandryhaila, J.M.F. Moura, J. Kovačević, Signal recovery on graphs: variation minimization, IEEE Trans. Signal Process. 63 (17) (2015) 4609–4624.
- [20] D. Romero, M. Ma, G.B. Giannakis, Kernel-based reconstruction of graph signals, IEEE Trans. Signal Process. 65 (3) (2016) 764–778.

- [21] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains, *IEEE Signal Process. Mag.* 30 (3) (2013) 83–98.
- [22] V. Kalofolias, How to learn a graph from smooth signals, in: *Artificial Intelligence and Statistics*, 2016, pp. 920–929.
- [23] G. Strang, The discrete cosine transform, *SIAM Rev.* 41 (1) (1999) 135–147.
- [24] D.I. Shuman, B. Ricaud, P. Vandergheynst, Vertex-frequency analysis on graphs, *Appl. Comput. Harmon. Anal.* 40 (2) (2016) 260–291.
- [25] M. Grant, S. Boyd, Y. Ye, *CVX: matlab software for disciplined convex programming*, 2008.
- [26] A. Wagadarikar, R. John, R. Willett, D. Brady, Single disperser design for coded aperture snapshot spectral imaging, *Appl. Opt.* 47 (10) (2008) B44–B51.
- [27] A.A. Wagadarikar, N.P. Pitsianis, X. Sun, D.J. Brady, Spectral image estimation for coded aperture snapshot spectral imagers, in: *Image Reconstruction from Incomplete Data V*, vol. 7076, International Society for Optics and Photonics, 2008, p. 707602.
- [28] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.* 54 (11) (2006) 4311–4322.
- [29] M.F. Duarte, R.G. Baraniuk, Kronecker compressive sensing, *IEEE Trans. Image Process.* 21 (2) (2011) 494–504.
- [30] C.V. Correa, H. Arguello, G.R. Arce, Snapshot colored compressive spectral imager, *JOSA A* 32 (10) (2015) 1754–1763.
- [31] H. Rueda-Chacon, F. Rojas, H. Arguello, Compressive spectral image fusion via a single aperture high throughput imaging system, *Sci. Rep.* 11 (1) (2021) 1–12.
- [32] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vis.* 20 (1) (2004) 89–97.
- [33] X. Bresson, T.F. Chan, Fast dual minimization of the vectorial total variation norm and applications to color image processing, *Inverse Probl. Imaging* 2 (4) (2008) 455–484.
- [34] X. Yuan, Generalized alternating projection based total variation minimization for compressive sensing, in: *2016 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2016, pp. 2539–2543.
- [35] S. Zheng, Y. Liu, Z. Meng, M. Qiao, Z. Tong, X. Yang, S. Han, X. Yuan, Deep plug-and-play priors for spectral snapshot compressive imaging, *Photonics Res.* 9 (2) (2021) B18–B29.
- [36] Y. Liu, X. Yuan, J. Suo, D.J. Brady, Q. Dai, Rank minimization for snapshot compressive imaging, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (12) (2018) 2990–3006.
- [37] Y. Fu, Y. Zheng, I. Sato, Y. Sato, Exploiting spectral-spatial correlation for coded hyperspectral image restoration, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3727–3736.
- [38] L. Wang, Z. Xiong, D. Gao, G. Shi, F. Wu, Dual-camera design for coded aperture snapshot spectral imaging, *Appl. Opt.* 54 (4) (2015) 848–858.
- [39] L. Wang, Z. Xiong, G. Shi, F. Wu, W. Zeng, Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (10) (2016) 2104–2111.
- [40] G. Yu, G. Sapiro, S. Mallat, Solving inverse problems with piecewise linear estimators: from gaussian mixture models to structured sparsity, *IEEE Trans. Image Process.* 21 (5) (2012) 2481–2499.
- [41] J. Yang, X. Liao, X. Yuan, P. Lull, D.J. Brady, G. Sapiro, L. Carin, Compressive sensing by learning a gaussian mixture model from measurements, *IEEE Trans. Image Process.* 24 (1) (2014) 106–119.
- [42] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [43] A. Gadde, S.K. Narang, A. Ortega, Bilateral filter: graph spectral interpretation and extensions, in: *Image Processing (ICIP)*, 2013 20th IEEE International Conference on, IEEE, 2013, pp. 1222–1226.
- [44] J. Hu, S. Li, The multiscale directional bilateral filter and its application to multisensor image fusion, *Inf. Fusion* 13 (3) (2012) 196–206.
- [45] Y. Wang, A. Ortega, D. Tian, A. Vetro, A graph-based joint bilateral approach for depth enhancement, in: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2014, pp. 885–889.
- [46] X. Yan, H. Qin, J. Li, H. Zhou, J.-g. Zong, Infrared and visible image fusion with spectral graph wavelet transform, *JOSA A* 32 (9) (2015) 1643–1652.
- [47] H. Salehi, J. Vahidi, An ultrasound image despeckling method based on weighted adaptive bilateral filter, *Int. J. Image Graph.* 20 (03) (2020) 2050020.
- [48] A. Chakrabarti, T. Zickler, Statistics of real-world hyperspectral images, in: *CVPR 2011*, IEEE, 2011, pp. 193–200.
- [49] B. Arad, O. Ben-Shahar, Sparse recovery of hyperspectral signal from natural RGB images, in: *European Conference on Computer Vision*, Springer, 2016, pp. 19–34.
- [50] M.V. Afonso, J.M. Bioucas-Dias, M.A.T. Figueiredo, An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems, *IEEE Trans. Image Process.* 20 (3) (2010) 681–695.
- [51] C.C. Paige, M.A. Saunders, Solution of sparse indefinite systems of linear equations, *SIAM J. Numer. Anal.* 12 (4) (1975) 617–629.
- [52] G. Vaksman, M. Zibulevsky, M. Elad, Patch ordering as a regularization for inverse problems in image processing, *SIAM J. Imaging Sci.* 9 (1) (2016) 287–319.
- [53] M.E. Gehm, R. John, D.J. Brady, R.M. Willett, T.J. Schulz, Single-shot compressive spectral imaging with a dual-disperser architecture, *Opt. Express* 15 (21) (2007) 14013–14027.
- [54] X. Lin, Y. Liu, J. Wu, Q. Dai, Spatial-spectral encoded compressive hyperspectral imaging, *ACM Trans. Graphics (TOG)* 33 (6) (2014) 233.
- [55] Y. Li, A. Majumder, H. Zhang, M. Gopi, Optimized multi-spectral filter array based imaging of natural scenes, *Sensors* 18 (4) (2018) 1172.
- [56] K. Monakhova, K. Yanny, N. Aggarwal, L. Waller, Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array, *Optica* 7 (10) (2020) 1298–1307.
- [57] O.F. Kar, F.S. Oktem, Compressive spectral imaging with diffractive lenses, *Opt. Lett.* 44 (18) (2019) 4582–4585.
- [58] P. Wang, R. Menon, Computational multispectral video imaging, *JOSA A* 35 (1) (2018) 189–199.
- [59] M. Nimmer, G. Steidl, R. Riesenberger, A. Wuttig, Spectral imaging based on 2D diffraction patterns and a regularization model, *Opt. Express* 26 (22) (2018) 28335–28348.
- [60] S.-H. Baek, I. Kim, D. Gutierrez, M.H. Kim, Compact single-shot hyperspectral imaging using a prism, *ACM Trans. Graphics (TOG)* 36 (6) (2017) 1–12.
- [61] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, W. Gao, Depth restoration from RGB-D data via joint adaptive regularization and thresholding on manifolds, *IEEE Trans. Image Process.* 28 (3) (2018) 1068–1079.
- [62] H. Rueda-Chacon, J.F. Florez-Ospina, D.L. Lau, G.R. Arce, Snapshot compressive ToF+ spectral imaging via optimized color-coded apertures, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (10) (2019) 2346–2360.
- [63] M.J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J.S. Duncan, L. Pizarro, D. Atkinson, B.F. Hutton, S. Ourselin, et al., Pet reconstruction with an anatomical MRI prior using parallel level sets, *IEEE Trans. Med. Imaging* 35 (9) (2016) 2189–2199.
- [64] Q. Yang, W. Cong, Y. Xi, G. Wang, Spectral X-ray CT image reconstruction with a combination of energy-integrating and photon-counting detectors, *PLoS ONE* 11 (5) (2016) e0155374.
- [65] R.C. Hardie, G.R. Arce, Ranking in $\mathbb{R}^P$ and its use in multivariate image estimation, *IEEE Trans. Circuits Syst. Video Technol.* 1 (2) (1991) 197–209.
- [66] G.R. Arce, *Nonlinear Signal Processing: A Statistical Approach*, John Wiley & Sons, 2005.
- [67] M. McLoughlin, G. Arce, Deterministic properties of the recursive separable median filter, *IEEE Trans. Acoust.* 35 (1) (1987) 98–106.
- [68] S.M. Cioabă, *A First Course in Graph Theory and Combinatorics*, vol. 55, Springer, 2009.