

Target-Based Hyperspectral Demixing via Generalized Robust PCA

Sirisha Rambhatla, Xingguo Li, and Jarvis Haupt

Department of Electrical and Computer Engineering,

University of Minnesota – Twin Cities, Minneapolis, MN-55455

{rambh002, lixx1661, jdhaupt}@umn.edu.

Abstract—Localizing targets of interest in a given hyperspectral (HS) image has applications ranging from remote sensing to surveillance. This task of target detection leverages the fact that each material/object possesses its own characteristic spectral response, depending upon its composition. As *signatures* of different materials are often correlated, matched filtering based approaches may not be appropriate in this case. In this work, we present a technique to localize targets of interest based on their spectral signatures. We also present the corresponding recovery guarantees, leveraging our recent theoretical results. To this end, we model a HS image as a superposition of a low-rank component and a dictionary sparse component, wherein the dictionary consists of the *a priori* known characteristic spectral responses of the target we wish to localize. Finally, we analyze the performance of the proposed approach via experimental validation on real HS data for a classification task, and compare it with related techniques.

Index Terms—Hyperspectral imaging, Robust-PCA, dictionary sparse, target localization, and remote sensing.

I. INTRODUCTION

Hyperspectral (HS) imaging is an imaging modality which senses the intensities of the reflected electromagnetic waves corresponding to different wavelengths of the electromagnetic spectra, often invisible to the human eye. As the spectral response associated with an object/material is dependent on its composition, HS imaging lends itself very useful in identifying target objects/materials via their characteristic spectra or *signature* responses, often referred to as "endmembers" in the literature. Typical applications of HS imaging range from monitoring agricultural use of land, catchment areas of rivers and water bodies, food processing and surveillance, to detecting various minerals, chemicals, and even presence of life sustaining compounds on distant planets. However, often, these spectral *signatures* are highly correlated, making it difficult to detect region of interest based on these endmembers. Analysis of these characteristic responses to localize a target material/object serves as the primary motivation of this work.

A. Model

We begin by formalizing the problem. Each HS image $\mathbf{I} \in \mathbb{R}^{n \times m \times f}$, is a stack of f 2-D images each of size $n \times m$, as shown in Fig. 1(a). Here, f is determined by the number of frequencies or frequency bands across which we measure the reflectances. Overall, each volumetric element or *voxel*, of a HS image is a vector of length f , and represents the response of the material in a pixel of the 2-D grid to f frequencies.

As a particular scene is composed of only a small number of objects/materials, the corresponding characteristic responses

of each voxel are low-rank [1]. For example, while imaging an agricultural area, we would expect to record responses from materials like biomass, farm vehicles, roads, houses and water bodies and so on. Further, the spectra of complex materials can be assumed to be a linear mixture of the constituent materials [1], [2], i.e. HS image voxels can be viewed as being generated by a linear mixture model [3]. Since, a particular scene does not contain a large number of distinct materials, we can decompose the responses into a low-rank part and a component which is sparse in a known dictionary – a *dictionary sparse* part.

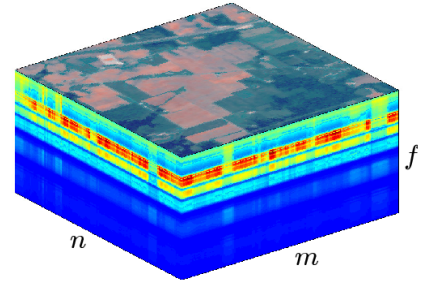


Fig. 1: The Hyperspectral image data-cube corresponding to the Indian Pines dataset. The top layer is depicted in RGB for visualization, however each slice is a grayscale image of reflectances.

Formally, let $\mathbf{Y} \in \mathbb{R}^{f \times nm}$ be formed by *unfolding* the HS image \mathbf{I} , such that, each column of \mathbf{Y} corresponds to a voxel. Now, let \mathbf{Y} arise as a result of a superposition of a low-rank component $\mathbf{X} \in \mathbb{R}^{f \times nm}$ with rank r , and a dictionary sparse component expressed here as $\mathbf{R}\mathbf{A}$, i.e.,

$$\mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A}. \quad (1)$$

Here, $\mathbf{R} \in \mathbb{R}^{f \times d}$ is an *a priori* known dictionary composed of normalized characteristic responses of the material/object, or of the constituents of the material we wish to localize, and $\mathbf{A} \in \mathbb{R}^{d \times nm}$ is the *sparse* coefficient matrix (also referred to as "abundances" in the literature) with s non-zero elements globally, i.e. $\|\mathbf{A}\|_0 = s$. Note that, \mathbf{R} can also be constructed by learning a dictionary based on the known spectral signatures of a target [4]–[6].

B. Prior Art

The model shown in (1) is a close counterpart of a number of well-known problems. To start, in the absence of the dictionary sparse part, (1) reduces to the popular problem of principal component analysis (PCA) [7]. This problem also shares its structure with variants of PCA such as robust-PCA [8], [9]

where \mathbf{R} is absent, outlier pursuit [10] where $\mathbf{R} = \mathbf{I}$ and \mathbf{A} is column sparse, and others [11]–[19]. On the other hand, the problem can be identified as that of sparse recovery [20]–[22], in the absence of the low-rank part \mathbf{X} . Following which, sparse recovery methods for analysis of HS images have been explored in [23]–[26]. In addition, in a recent work [27], the authors further impose a low-rank constraint on the coefficient matrix \mathbf{A} . Further, applications of compressive sampling have been explored in [28], while [3] analyzes the case where HS images are noisy and incomplete.

The model described in (1) was introduced in [29] as a means to detect traffic anomalies in a network, wherein, the authors focus on a case where the dictionary \mathbf{R} is *overcomplete*, i.e., *fat*, and the rows of \mathbf{R} are orthogonal, e.g., $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$. Further, the coefficient matrix \mathbf{A} is assumed to possess at most k nonzero elements per row and column and s nonzero elements overall.

In a recent work [30], we analyze the extension of [29] to a case where the dictionary has more rows than columns, i.e., is *thin*, while removing the orthogonality constraint for both the *thin* and the *fat* dictionary cases, when s is small. The *thin* dictionary case is amenable to a number of applications, as often, we are looking for only a small number of *signatures*. For example, in our motivating application of HS imaging, we are interested in detecting the presence of a small number of spectral *signatures* corresponding to the target of interest. To this end, we focus our attention on the *thin* case, although a similar analysis applies for the *fat* case [30].

Note that, when the known dictionary \mathbf{R} is thin, we can alternatively multiply (1) on the left by the pseudo-inverse of \mathbf{R} , say \mathbf{R}^\dagger , in which case, the model shown in (1) reduces to that of robust PCA, i.e.,

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}} + \mathbf{A}, \quad (2)$$

where $\tilde{\mathbf{Y}} = \mathbf{R}^\dagger \mathbf{Y}$ and $\tilde{\mathbf{X}} = \mathbf{R}^\dagger \mathbf{X}$. Therefore, in this case, we can adopt a two-step procedure: (1) recover the sparse matrix \mathbf{A} by robust PCA and, (2) estimate the low-rank part using the estimate of $\mathbf{R}\mathbf{A}$.¹ Models considered in (1) and (2) are two alternate methods of recovering the components for thin dictionaries. These two methods inherently solve different optimization problems. The proposed method is a one-shot procedure with guarantees to solve the problem, while the conditions under which the pseudo-inverse based procedure succeeds will depend on the interaction between \mathbf{R}^\dagger and the low-rank part. Further, our experiments indicate that solving for model in (1) is more *robust* as compared to RPCA^\dagger for the classification problem at hand, across different choice of dictionaries.

C. Our Contribution

In this work, we present a one-step method for target detection in a hyperspectral image with the underlying model of (1). This is done by building upon the results of [30] for the underlying application. In particular, we show the conditions under which target detection is possible with thin dictionaries. In general, the method is also applicable to fat dictionaries

and for other applications, which follow the underlying model. The rest of the paper is organized as follows. We formulate the problem, introduce the notation and present the theoretical result for the case at hand, i.e., the *thin* case, in section II. Next, we present the description of the dataset along with the experimental evaluations in section III. Finally, we conclude this discussion in section IV.

II. PROBLEM FORMULATION

Our aim is to recover the low-rank component \mathbf{X} , and the sparse coefficient matrix \mathbf{A} , given the dictionary \mathbf{R} , and samples \mathbf{Y} generated according to the model described in (1). We utilize the structure of the components, \mathbf{X} and \mathbf{A} , to arrive at the following convex problem for $\lambda \geq 0$,

$$\underset{\mathbf{X}, \mathbf{A}}{\text{minimize}} \|\mathbf{X}\|_* + \lambda \|\mathbf{A}\|_1 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A}. \quad (3)$$

where, $\|\cdot\|_*$ denotes the nuclear norm, and $\|\cdot\|_1$ refers to the l_1 -norm, which serve as convex relaxations of rank and sparsity (i.e. l_0 -norm), respectively. For the application at hand, we present the theoretical result for the case when the number of dictionary elements $d \leq f$. Analysis of a more general case can be found in [30]. For the this case, we assume the dictionary \mathbf{R} to be a *frame*, i.e. for any vector $\mathbf{v} \in \mathbb{R}^d$, we have

$$\mathbf{F}_L \|\mathbf{v}\|_2^2 \leq \|\mathbf{R}\mathbf{v}\|_2^2 \leq \mathbf{F}_U \|\mathbf{v}\|_2^2, \quad (4)$$

where \mathbf{F}_L and \mathbf{F}_U are the lower and upper *frame bounds*, respectively, with $0 < \mathbf{F}_L \leq \mathbf{F}_U$.

We begin by defining a few relevant subspaces, similar to those used in [29].

Let the pair $\{\mathbf{X}_0, \mathbf{A}_0\}$ be the solution to the problem shown in (3). We define Φ as the space of matrices spanning either the row or the column space of the low-rank component \mathbf{X}_0 . Specifically, let $\mathbf{U}\Sigma\mathbf{V}^\top$ denote the singular value decomposition of \mathbf{X}_0 , then the space Φ is defined as

$$\Phi := \{\mathbf{U}\mathbf{W}_1^\top + \mathbf{W}_2\mathbf{V}^\top, \mathbf{W}_1 \in \mathbb{R}^{nm \times r}, \mathbf{W}_2 \in \mathbb{R}^{f \times r}\}.$$

Next, let Ω be the space spanned by the matrices which have the same support (locations of non-zero elements) as \mathbf{A}_0 , and let Ω_R be defined as

$$\Omega_R := \{\mathbf{Z} = \mathbf{R}\mathbf{H}, \mathbf{H} \in \Omega\}.$$

Next, let $\mathcal{P}_\Phi(\cdot)$, $\mathcal{P}_\Omega(\cdot)$ and $\mathcal{P}_{\Omega_R}(\cdot)$ be the orthogonal projection operator(s) onto the space of matrices defined above. In addition, we will use \mathbf{P}_U and \mathbf{P}_V to denote the projection matrices for the column and row spaces of \mathbf{X}_0 , respectively, i.e., implying the following for any matrix \mathbf{X} ,

$$\mathcal{P}_\Phi(\mathbf{X}) = \mathbf{P}_U \mathbf{X} + \mathbf{X} \mathbf{P}_V - \mathbf{P}_U \mathbf{X} \mathbf{P}_V$$

Indeed, there are situations under which we cannot hope to recover the matrices \mathbf{X} and \mathbf{A} . To characterize these scenarios, we employ various notions of incoherence, out of which, the first is the incoherence between the low-rank part, \mathbf{X}_0 , and the dictionary sparse part, $\mathbf{R}\mathbf{A}_0$,

$$\mu := \max_{\mathbf{Z} \in \Omega_R \setminus \{\mathbf{0}_{d \times nm}\}} \frac{\|\mathcal{P}_\Phi(\mathbf{Z})\|_F}{\|\mathbf{Z}\|_F},$$

where $\mu \in [0, 1]$ is small when these components are incoherent (which is good for recovery). The next two measures of incoherence can be interpreted as a way to avoid the cases where for $\mathbf{X}_0 = \mathbf{U}\Sigma\mathbf{V}^\top$, (a) \mathbf{U} resembles the dictionary \mathbf{R} ,

¹We denote this method of estimating \mathbf{A} as RPCA^\dagger in our experiments.

TABLE I: Simulation results for the proposed approach (XpRA), robust-PCA based approach (RPCA[†]), matched filtering (MF) on original data \mathbf{Y} , and matched filtering on transformed data $\mathbf{R}^\top \mathbf{Y}$ (MF[†]), across dictionary elements d , and the regularization parameter for initial dictionary learning step ρ . Threshold selects columns with column-norm greater than threshold such that AUC is maximized. Further, “*” denotes the case where ROC curve was “flipped” (i.e. classifier output was inverted to improve performance).

(a) Learned dictionary, $d = 4$

d	ρ	Method	Threshold	Performance at best operating point		AUC
				TPR	FPR	
4	0.01	XpRA	0.30	0.979	0.023	0.989
		RPCA [†]	0.65	0.957	0.049	0.974
		MF ₊	N/A	0.957	0.036	0.994
		MF [†]	N/A	0.914	0.104	0.946
	0.1	XpRA	0.8	0.989	0.017	0.997
		RPCA [†]	0.8	0.989	0.014	0.997
		MF	N/A	0.989	0.016	0.998
		MF [†]	N/A	0.989	0.010	0.998
	0.5	XpRA	0.6	0.968	0.031	0.991
		RPCA [†]	0.6	0.935	0.067	0.988
		MF	N/A	0.548	0.474	0.555
		MF [†]	N/A	0.849	0.119	0.939

(b) Learned dictionary, $d = 10$

d	ρ	Method	Threshold	Performance at best operating point		AUC
				TPR	FPR	
10	0.01	XpRA	0.6	0.935	0.060	0.972
		RPCA [†]	0.7	0.978	0.023	0.990
		MF ₊	N/A	0.624	0.415	0.681
		MF [†]	N/A	0.569	0.421	0.619
	0.1	XpRA	0.5	0.968	0.029	0.993
		RPCA [†]	0.5	0.871	0.144	0.961
		MF ₊	N/A	0.688	0.302	0.713
		MF [†]	N/A	0.527	0.469	0.523
	0.5	XpRA	1	0.978	0.031	0.996
		RPCA [†]	2.2	0.849	0.113	0.908
		MF	N/A	0.807	0.309	0.781
		MF [†]	N/A	0.527	0.465	0.539

(c) Dictionary by sampling voxels, $d = 15$

d	Method	Threshold	Performance at best operating point		AUC
			TPR	FPR	
15	XpRA	0.3	0.989	0.021	0.998
	RPCA [†]	3	0.849	0.146	0.900
	MF	N/A	0.957	0.085	0.978
	MF [†]	N/A	0.796	0.217	0.857

(d) Average performance

Method	TPR		FPR		AUC	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
XpRA	0.972	0.019	0.030	0.014	0.991	0.009
RPCA [†]	0.919	0.061	0.079	0.055	0.959	0.040
MF	0.796	0.179	0.234	0.187	0.814	0.178
MF [†]	0.739	0.195	0.258	0.192	0.775	0.207

and (b) \mathbf{V} as the sparse coefficient matrix \mathbf{A}_0 . To this end, we define respectively, the following to measure these properties,

$$\gamma_{UR} := \max_i \frac{\|\mathbf{P}_U \mathbf{R} \mathbf{e}_i\|^2}{\|\mathbf{R} \mathbf{e}_i\|^2} \text{ and } \gamma_V := \max_i \|\mathbf{P}_V \mathbf{e}_i\|^2,$$

where $\gamma_V \in [r/nm, 1]$. Also, we define $\xi := \|\mathbf{R}^\top \mathbf{U} \mathbf{V}^\top\|_\infty$. Finally, for a matrix \mathbf{M} , we use $\|\mathbf{M}\| := \sigma_{\max}(\mathbf{M})$ for the spectral norm, where $\sigma_{\max}(\mathbf{M})$ denotes the maximum singular value of the matrix \mathbf{M} , and $\|\mathbf{M}\|_\infty := \max_{\{i, j\}} |M_{ij}|$.

A. Theoretical Result

We now present our main theoretical result for the *thin* case, i.e., when $d \leq f$, which was developed in [30]. We begin by introducing the following definitions and assumptions.

Definition D.1. $\lambda_{\min} = \frac{1+C}{1-C} \xi$, where C is defined as

$$C := \frac{c}{\mathbf{F}_L(1-\mu)^2 - c},$$

$$c := \frac{\mathbf{F}_U}{2} [(1 + 2\gamma_{UR})(\min(s, d) + s\gamma_V) + 2s\gamma_V] - \frac{\mathbf{F}_L}{2} [\min(s, d) + s\gamma_V].$$

Definition D.2. $\lambda_{\max} := \frac{1}{\sqrt{s}} (\sqrt{\mathbf{F}_L} (1 - \mu) - \sqrt{r\mathbf{F}_U\mu})$.

Assumption A.1. $\lambda_{\max} \geq \lambda_{\min}$

Assumption A.2. Let $s_{\max} := \frac{(1-\mu)^2 nm}{2r}$, then

$$\gamma_{UR} \leq \begin{cases} \frac{(1-\mu)^2 - 2s\gamma_V}{2s(1+\gamma_V)}, & \text{for } s \leq \min(d, s_{\max}) \\ \frac{(1-\mu)^2 - 2s\gamma_V}{2(d+s\gamma_V)}, & \text{for } d < s \leq s_{\max} \end{cases}.$$

The following theorem establishes the conditions under which we can recover the pair $\{\mathbf{X}_0, \mathbf{A}_0\}$ successfully.

Theorem 1. Consider a superposition of a low-rank matrix $\mathbf{X}_0 \in \mathbb{R}^{f \times nm}$ of rank r , and a dictionary sparse component $\mathbf{R} \mathbf{A}_0$, wherein the dictionary $\mathbf{R} \in \mathbb{R}^{f \times d}$, with $d \leq f$, obeys the frame condition with frame bounds $[\mathbf{F}_L, \mathbf{F}_U]$, the sparse coefficient matrix \mathbf{A}_0 has at most s non-zeros, i.e., $\|\mathbf{A}_0\|_0 = s$, where $s \leq s_{\max}$, and $\mathbf{Y} = \mathbf{X}_0 + \mathbf{R} \mathbf{A}_0$, with parameters γ_{UR} , ξ , $\gamma_V \in [r/nm, 1]$ and $\mu \in [0, 1]$. Then, solving the formulation shown in (3) will recover matrices \mathbf{X}_0 and \mathbf{A}_0 if assumptions A.1 and A.2 hold for any $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, as defined in D.1 and D.2, respectively.

In other words, Theorem 1 establishes the sufficient conditions for the existence of λ s to guarantee recovery of $\{\mathbf{X}_0, \mathbf{A}_0\}$. We see that these conditions are closely related to the various incoherence measures γ_{UR} , γ_V and μ , between \mathbf{X} , \mathbf{R} and \mathbf{A} .

Note that, we have here an upper-bound on the global sparsity, i.e., $s \leq s_{\max}$, however, our simulation results in [30] on synthetic data reveal that the recovery is indeed possible even for higher s ; See [30] for the details of the analysis. To this end, we propose a randomized analysis of the problem in (3) to improve these results as a future work.

III. EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of the proposed technique on real HS data. We begin by describing the dataset used for the simulations, following which we describe the experimental set-up and conclude by presenting the results.

A. Data

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [31] sensor is one of the popular choices for collecting HS images for various remote sensing applications. In this present exposition, we consider the “Indian Pines” dataset [32], which was collected over the Indian Pines test site in North-western Indiana in the June of 1992. This dataset consists of spectral reflectances across 224 bands in wavelength of ranges 400 – 2500 nm from a scene which is composed mostly of agricultural land, along with two major dual lane highways, a rail line and some built structures. This dataset is further processed by removing the bands corresponding to those of water absorption, and the resulting HS data-cube with dimensions $\{145 \times 145 \times 200\}$ is as visualized in Fig. 1. This modified dataset is available as “corrected Indian Pines” dataset [32], with the ground-truth containing 16 classes; Henceforth, referred to as the “Dataset”.

We will analyze the performance of the proposed technique for the identification of the stone-steel towers (class 16 in the Dataset), shown as ground-truth in Fig. 2(b), which constitutes about 93 voxels in the Dataset. We consider two cases for choosing the known dictionary \mathbf{R} – (1) when a (thin) dictionary is learned [4]–[6] based on the voxels (class-16), and (2) when the dictionary is formed by randomly sampling voxels (from class-16). This is to emulate the two ways in which we can arrive at the characteristics of a target – (1) the case where the *exact signatures* are not available, and/or there is noise, and (2) where we have access to the exact signatures we are looking for. Next, we form the data matrix

$\mathbf{Y} \in \mathbb{R}^{f \times nm}$ by stacking each voxel of the image side-by-side, which results in a $\{200 \times 145^2\}$ data matrix \mathbf{Y} .

Normalization: For normalization, we divide the data matrix \mathbf{Y} and the dictionary \mathbf{R} by $\|\mathbf{Y}\|_\infty$ to preserve the inter-voxel scaling. Next, we normalize the columns of \mathbf{R} , such that they are unit-norm. In practice, we can store the un-normalized dictionary \mathbf{R} , consisting of actual *signatures* of the target material, and can normalize it, as described above, after the HS image has been acquired.

B. Experimental Setup

We evaluate and compare the performance of the proposed method (XpRA) with RPCA^\dagger using transformed data $\tilde{\mathbf{Y}} = \mathbf{R}^\dagger \mathbf{Y}$ (as described in section I-B), matched filtering (MF) based approach on data \mathbf{Y} (denoted by MF), and a MF based approach on $\tilde{\mathbf{Y}}$, denoted by MF^\dagger , via receiver operating characteristic (ROC), shown in Table I(a)-(d).

For the proposed technique, we employ the accelerated proximal gradient (APG) algorithm presented in [29] to solve the optimization problem shown in (3). Similarly, for RPCA^\dagger we employ the APG with transformed data matrix $\tilde{\mathbf{Y}}$, while setting $\mathbf{R} = \mathbf{I}$; See (2). With reference to selection of tuning parameters for the APG solver for XpRA (RPCA^\dagger , respectively), we choose $v = 0.95$, $\nu = \|\mathbf{Y}\|$ ($\nu = \|\tilde{\mathbf{Y}}\|$), $\bar{\nu} = 10^{-4}$, and scan through 100 λ s in the range $\lambda \in (0, |\mathbf{R}^\top \mathbf{Y}|_\infty / \|\mathbf{Y}\|)$ ($\lambda \in (0, \|\tilde{\mathbf{Y}}\|_\infty / \|\tilde{\mathbf{Y}}\|)$), to generate the ROCs. The output of XpRA and RPCA^\dagger is a data-cube corresponding to the low-rank part and the dictionary sparse part (which contains the region of interest). We threshold the resulting estimates of the sparse part \mathbf{A} , based on their column norms. We choose the threshold such that the area under the curve (AUC) is maximized for both cases (XpRA and RPCA^\dagger).

In case of MF, we form the inner-product of the column-normalized data matrix \mathbf{Y} , say \mathbf{Y}_n , with the dictionary \mathbf{R} , i.e., $\mathbf{R}^\top \mathbf{Y}_n$, and select the maximum absolute inner-product per column. For MF^\dagger , matched filtering corresponds to finding maximum absolute entry for each column of the column-normalized $\tilde{\mathbf{Y}}$. Next, in both cases we scan through 1000 threshold values between $(0, 1]$ to generate the ROC results shown in Table I.

Table I(a)-(b) show the ROC characteristics for the aforementioned methods for a dictionary learned using all voxels of class-16. We evaluate the performance of the aforementioned techniques for two dictionary sizes, $d = 4$ and $d = 10$, for three values of the regularization parameter ρ , used for the initial dictionary learning step, i.e., $\rho = 0.01$, 0.1 and 0.5 . (The parameter ρ controls the sparsity during the initial dictionary learning step.)

Table I(c) shows the ROC results for the case where we randomly select 15 voxel from class-16 to form the known dictionary. Further, we show the overall performance in terms of the mean and standard deviation of ROC metrics (TPR, FPR and AUC), for each algorithm in Table I(d), i.e., across different dictionaries considered in Table I(a)-(c).

C. Analysis

As described above, Table I shows the ROCs for the classification performance of the proposed method as compared to RPCA^\dagger , MF and MF^\dagger . We note that when we use \mathbf{R} which has been learned based on samples of class-16, matched filtering based approaches under perform (with an exception for $d = 4$ and $\rho = 0.1$). RPCA^\dagger performs well for $d = 10$ and $\rho = 0.01$. For all other choices the proposed method performs the best. Even for cases in which it is not the best, it is still comparable to the best performer. Following this, we conclude that the performance of XpRA is most reliable across different the choices of the dictionaries. This is also evident by Table I(d) which shows the mean and standard deviation of ROC metrics (TPR, FPR and AUC). We observe that the mean and standard deviation metrics for the proposed method are superior to other methods across the different dictionary choices.

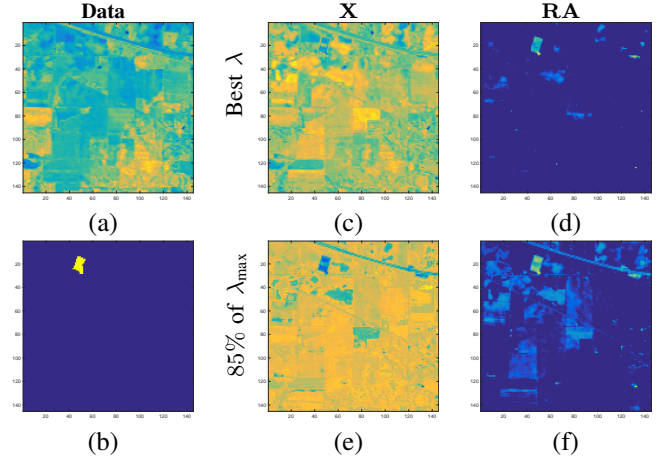


Fig. 2: Recovery of the low-rank component \mathbf{X} and the dictionary sparse component \mathbf{RA} for different values of λ for the proposed technique at $f = 50$ slice of the data (a), (b) corresponding ground truth for class-16. Panel (c) and (d), and (e) and (f) show the recovery of the low-rank part and the dictionary part, for the λ at the best operating point and for smaller λ , respectively, dictionary formed from data voxels (Table I(c)).

There are other interesting recovery results which warrant our attention. For example, Fig. 2 shows the low-rank component and the dictionary sparse component recovered by the proposed method for two different values of λ , for the case where we form the dictionary by randomly sampling the voxels (Table I(c)). Interestingly, we recover the rail tracks/roads running diagonally on the top-right corner, along with some low-density housing; See Fig 2 (f). This is because the *signatures* we seek (stone-steel towers) are similar to those of the materials used in these structures. This further corroborates the applicability of the proposed approach in detecting the presence of a particular spectral *signature* in a HS image. However, this also highlights potential drawback of this technique. As XpRA is based on identifying materials with similar composition, it may not be effective in distinguishing between very closely related classes, say two agricultural crops.

IV. CONCLUSIONS

We present a generalized robust PCA-based technique to localize a target in a HS image, based on the *a priori* known spectral *signature* of the material we wish to localize. We model the data as being composed of a low-rank component and a dictionary sparse component. Here, the dictionary contains the *a priori* known spectral *signatures* of the target. In this work, we adapt the theoretical results of [30] and present the conditions under which such a decomposition recovers the two components for the HS demixing task. Further, we evaluate and compare the performance of the proposed method via experimental evaluations on real HS image data for a classification task.

V. ACKNOWLEDGEMENT

The authors graciously acknowledge support from the DARPA YFA, Grant N66001-14-1-4047.

REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan 2002.
- [2] J. B. Greer, "Sparse demixing of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 219–228, Jan 2012.
- [3] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral images," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 33–56, 2012.
- [4] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.
- [5] B. A. Olshausen and D. J. Field, "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [6] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: Design of Dictionaries for Sparse Representation," in *Proceedings of SPARS*, pp. 9–12, 2005.
- [7] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.
- [8] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.
- [9] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.
- [10] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," in *Advances in Neural Information Processing Systems*, 2010, pp. 2496–2504.
- [11] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma, "Stable principal component pursuit," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, IEEE, 2010, pp. 1518–1522.
- [12] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3419–3430, 2011.
- [13] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Information and Inference*, vol. 2, no. 1, pp. 32–68, 2013.
- [14] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis, "Low-rank matrix recovery from errors and erasures," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4324–4337, 2013.
- [15] X. Li and J. Haupt, "Identifying outliers in large matrices via randomized adaptive compressive sampling," *Trans. Signal Processing*, vol. 63, no. 7, pp. 1792–1807, 2015.
- [16] X. Li and J. Haupt, "Locating salient group-structured image features via adaptive compressive sensing," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015.
- [17] X. Li and J. Haupt, "Outlier identification via randomized adaptive compressive sampling," in *IEEE International Conference on Acoustic, Speech and Signal Processing*, 2015.
- [18] X. Li and J. Haupt, "A refined analysis for the sample complexity of adaptive compressive outlier sensing," in *IEEE Workshop on Statistical Signal Processing*, 2016.
- [19] X. Li, J. Ren, Y. Xu, and J. Haupt, "An efficient dictionary based robust pca via sketching," *Technical Report*, 2016.
- [20] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.
- [21] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [22] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [23] Y. Moudden, J. Bobin, J. L. Starck, and J. M. Fadili, "Dictionary learning with spatio-spectral sparsity constraints," in *Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, 2009.
- [24] J. Bobin, Y. Moudden, J. L. Starck, and J. Fadili, "Sparsity constraints for hyperspectral data analysis: Linear mixture model and beyond," 2009.
- [25] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y. W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011, pp. 2329–2336.
- [26] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 963–978, Sept 2011.
- [27] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4775–4789, Aug 2016.
- [28] M. Golbabaee, S. Arberet, and P. Vanderghenst, "Distributed compressed sensing of hyperspectral images via blind source separation," in *Forty Fourth Asilomar Conference on Signals, Systems and Computers*, Nov 2010, pp. 196–198.
- [29] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 5186–5205, 2013.
- [30] S. Rambhatla, X. Li, and J. Haupt, "A dictionary based generalization of robust PCA," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, IEEE, 2016.
- [31] Jet Propulsion Laboratory, NASA and California Institute of Technology, "Airborne Visible/Infrared Imaging Spectrometer," 1987, Available at <http://aviris.jpl.nasa.gov/>.
- [32] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, "220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3, dataset available via http://www.ehu.eu/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes," Sept 2015.