

Reconstruction of Hyperspectral Imagery From Random Projections Using Multihypothesis Prediction

Chen Chen, *Student Member, IEEE*, Wei Li, *Member, IEEE*, Eric W. Tramel, *Member, IEEE*, and James E. Fowler, *Senior Member, IEEE*

Abstract—Reconstruction of hyperspectral imagery from spectral random projections is considered. Specifically, multiple predictions drawn for a pixel vector of interest are made from spatially neighboring pixel vectors within an initial non-predicted reconstruction. A two-phase hypothesis-generation procedure based on partitioning and merging of spectral bands according to the correlation coefficients between bands is proposed to fine-tune the hypotheses. The resulting prediction is used to generate a residual in the projection domain. This residual being typically more compressible than the original pixel vector leads to improved reconstruction quality. To appropriately weight the hypothesis predictions, a distance-weighted Tikhonov regularization to an ill-posed least-squares optimization is proposed. Experimental results demonstrate that the proposed reconstruction significantly outperforms alternative strategies not employing multihypothesis prediction.

Index Terms—Compressed sensing, hyperspectral data, multihypothesis prediction, principal component analysis, Tikhonov regularization.

I. INTRODUCTION

HYPERSPECTRAL imagery (HSI) captures a dense spectral sampling of reflectance values over a wide range of the spectrum. While the rich spectral information is expected to improve the performance of image-analysis techniques, this improvement often comes with the cost of computational complexity, over-dimensionality, and statistical ill-conditioning. Furthermore, high-dimensional HSI data also increases communication costs associated with the transmission of data from the remote sensor. As a consequence, some form of spectral dimensionality reduction is almost always required before HSI can be used in image-analysis applications; if this can be accomplished prior to downlink of the data set from

the remote sensor, communications costs can be significantly alleviated.

There has been interest in using random projections effectuated directly within the hardware of the hyperspectral sensor to simultaneously reduce dimensionality as the HSI data set is sensed, and sensor architectures for such random-projection-based spectral imaging have been proposed (e.g., [1]). In such a paradigm, the computational cost of dimensionality reduction is shifted from the resource-constrained remote-sensing platform to a more-capable ground-based receiver which is then burdened with the task of reconstruction of the HSI data set from the random projections. There has been, of course, significant recent interest in using compressed sensing (CS) (e.g., [2]) to provide such reconstruction. Alternatively, [3] proposed a reconstruction strategy—compressive-projection principal component analysis (CPPCA)—which recovers an HSI data set using principal component analysis (PCA). Specifically, from the random projections, the CPPCA receiver recovers not only the coefficients associated with the PCA transform, but also an approximation to the PCA transform basis itself. Experimental results [3], [4] have demonstrated significant advantage for CPPCA as compared to various CS reconstructions for HSI data sets in terms of both reconstruction quality as well as computational complexity.

Irrespective of the specific reconstruction algorithm used, there has been a focus in recent literature in improving reconstructions from random projections via the use of prediction followed by reconstruction from the resulting projection-domain residual. This approach has been particularly prevalent in the CS reconstruction of video wherein predictions are made temporally from one frame to the next (see [5] for a comprehensive survey). However, it is also possible to apply prediction and residual reconstruction spatially within a single image—[6], for example, reconstructs still images by deriving a prediction for an image block from the spatially surrounding blocks within an initial reconstruction.

In this paper, we propose to incorporate this idea of spatial prediction and residual reconstruction into the reconstruction from random projections of HSI data sets. The focus is on the use of multihypothesis (MH) prediction [7] in which multiple hypotheses are created and then combined to yield a composite prediction superior to any of the constituent single-hypothesis predictions. Central to our discussion is a formulation of the MH prediction procedure in the domain of the random

Manuscript received September 19, 2012; revised December 17, 2012; accepted January 9, 2013. This work was supported by the National Science Foundation under Grant CCF-0915307.

C. Chen is with the University of Texas at Dallas, Richardson, TX 75083-0688 USA (e-mail: chenchen870713@gmail.com).

W. Li is with the University of California, Davis, Davis, CA 95616 USA (e-mail: liwei089@ieee.org).

E. W. Tramel and J. E. Fowler are with the Department of Electrical and Computer Engineering and the Geosystems Research Institute, Mississippi State University, Starkville, MS 39759-3468 USA (e-mail: eric.tramel@gmail.com; fowler@ece.msstate.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2013.2240307

projections as well as a hypothesis-generation procedure based on spectral-band partitioning; as this formulation results in an ill-posed optimization, we resort to Tikhonov regularization [8] which is widely used to yield tractable solutions to such ill-posed problems. In experimental results, we compare our proposed Tikhonov-based MH regularization against several straightforward reconstruction techniques which do not employ MH prediction. We find that our proposed approach yields superior reconstruction across a broad range of subsampling rates.

The remainder of the manuscript is organized as follows. In Section II, several techniques—including those based on CS and CPPCA—for the reconstruction from random projections of hyperspectral data are reviewed. Then, in Section III, we employ the general strategy of MH prediction to the problem of HSI reconstruction, focusing as well on the generation of hypotheses for HSI data. In Section IV, experimental comparison of various algorithms is presented. Finally, in Section V, some concluding remarks are made.

II. RECONSTRUCTION FROM RANDOM PROJECTIONS

Consider a data set of M vectors $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_M]$, where each $\mathbf{x}_m \in \mathbb{R}^N$. We assume that the signal-acquisition device applies an $N \times K$ orthonormal random projection \mathbf{P} to obtain random projections $\tilde{\mathbf{Y}} = [\tilde{\mathbf{y}}_1 \dots \tilde{\mathbf{y}}_M]$ where each $\tilde{\mathbf{y}}_m = \mathbf{P}^T \mathbf{x}_m$ has dimension K ($K \ll N$); K/N is referred to as the *sub-rate* hereafter. To reconstruct an approximation $\hat{\mathbf{X}}$ from the projections $\tilde{\mathbf{Y}}$, we could apply one of the several existing algorithms; we overview three below, specifically, multitask Bayesian compressive sensing (MT-BCS) [9] in Section II-A, CPPCA [3] in Section II-B, and the recently proposed class-dependent CPPCA (C-CPPCA) [10] in Section II-C.

A. Multitask Bayesian Compressive Sensing

CS, in brief, produces a sparse signal representation directly from a small number of projections onto another basis, recovering the sparse transform coefficients via nonlinear reconstruction. The main tenet of CS theory holds that, if signal $\mathbf{x} \in \mathbb{R}^N$ can be sparsely represented (i.e., using only L nonzero coefficients) with some basis, then we can recover \mathbf{x} from K -dimensional projections $\tilde{\mathbf{y}} = \mathbf{P}^T \mathbf{x}$ under certain conditions. For recovery of a set of multiple, possibly correlated vectors $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_M]$, there have been proposals for multi-vector extensions of CS under the name of “multitask” [9] CS; these, in turn, link closely to a larger body of literature on “simultaneous sparse approximation” (e.g., [11]–[15]). Below, we focus on MT-BCS [9] which introduces a hierarchical Bayesian framework into the multi-vector CS-recovery problem to share prior information across the multiple vectors.

B. Compressive-Projection Principal Component Analysis

CPPCA [3] is driven by projections at the sensor onto lower-dimensional subspaces chosen at random. The CPPCA receiver, given only these random projections, recovers not only the

coefficients associated with the PCA transform, but also an approximation to the PCA transform basis itself.

CPPCA reconstruction of a set of randomly projected vectors first entails an eigenvector-reconstruction process based on an approximation that uses Ritz vectors [16] as close representations of orthonormal projections of eigenvectors. Specifically, the PCA transform of \mathbf{x}_m in data set $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_M]$ is $\hat{\mathbf{x}}_m = \mathbf{U}^T \mathbf{x}_m$, where $N \times N$ transform matrix \mathbf{U} emanates from the eigendecomposition of $\Sigma_{\mathbf{X}}$; i.e.,

$$\Sigma_{\mathbf{X}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (1)$$

where \mathbf{U} contains the N unit eigenvectors of $\Sigma_{\mathbf{X}}$ column-wise. CPPCA reconstruction uses the first L Ritz vectors (essentially, the eigenvectors of $\tilde{\Sigma}_{\tilde{\mathbf{Y}}} = \mathbf{P}^T \Sigma_{\mathbf{X}} \mathbf{P}$) to obtain approximations of the first L principal eigenvectors in \mathbf{U} corresponding to the L largest eigenvalues in $\mathbf{\Lambda}$. These approximate eigenvectors are then assembled into $N \times L$ matrix $\tilde{\Psi}$. The reconstruction of the data set \mathbf{X} is then produced in a pseudoinverse-based recovery of the PCA coefficients

$$\hat{\mathbf{X}} = \tilde{\Psi} (\mathbf{P}^T \tilde{\Psi})^+ \tilde{\mathbf{Y}}. \quad (2)$$

The number of eigenvectors to recover is determined by the heuristic $L = \text{round}(K / \log N)$ as proposed in [17]. The reader is referred to [3] for greater detail.

C. Class-Dependent Compressive-Projection Principal Component Analysis

C-CPPCA [10] can be viewed as an extension of the original CPPCA that incorporates classification into the reconstruction. For this approach, the sender-side sensing procedure is exactly the same as that employed in original CPPCA. At the receiver, the random projections $\tilde{\mathbf{Y}}$ are first classified into one of several groups using a pixel-wise classification algorithm, such as a support vector machine (SVM). After the grouping procedure, CPPCA reconstruction is employed for each class independently.

More specifically, a small set of “exemplars” is chosen randomly from the projections $\tilde{\mathbf{Y}}$; these exemplars are recovered using MT-BCS (which is efficient at recovering a small number of samples, unlike CPPCA which requires more samples to be effective). These reconstructed samples are then clustered into different groups, producing “pseudo” *a priori* label information (training data). A trained pixel-wise classifier is then employed to classify each pixel in each $\tilde{\mathbf{Y}}$ into one of K_0 classes. Thus, each $\tilde{\mathbf{Y}}$ is further partitioned into K_0 groups, $\tilde{\mathbf{Y}}_k$, $k = 1, \dots, K_0$, based on this classification. Finally, traditional CPPCA reconstruction is employed individually on each class.

It was shown in [10] that the C-CPPCA strategy outperforms the original CPPCA for HSI. This can be attributed to multiple modes in the data distribution arising from the presence of multiple classes/objects in the image. Hence, the C-CPPCA method employs statistics pertinent to each class as opposed to the average statistics over all classes. The resulting reconstruction then exploits individual local geometrical distributions as provided by the covariance matrix of each class as opposed to a single aggregated distribution as is employed in

the original CPPCA. A detailed discussion for C-CPPCA can be found in [10].

III. RECONSTRUCTION USING MH PREDICTION

In this section, we propose a reconstruction algorithm based on MH prediction. The algorithm is driven by the idea that, if a prediction is similar to the target sample, then the prediction residual will be more amenable to reconstruction since the residual is typically more compressible than the original pixel vector. Specifically, suppose that \mathbf{x} is a pixel vector (hyperspectral signature) from an HSI data set, and $\bar{\mathbf{x}}$ is a prediction which satisfies $\bar{\mathbf{x}} \approx \mathbf{x}$. The residual is $\mathbf{r} = \mathbf{x} - \bar{\mathbf{x}}$. With a $N \times K$ orthonormal random projection \mathbf{P} , the projection of \mathbf{r} is $\mathbf{q} = \mathbf{P}^T \mathbf{r} = \tilde{\mathbf{y}} - \mathbf{P}^T \bar{\mathbf{x}}$. The final reconstruction of $\tilde{\mathbf{y}}$ is calculated as $\hat{\mathbf{x}} = \bar{\mathbf{x}} + \text{Reconstruct}(\mathbf{q}, \mathbf{P})$, where $\text{Reconstruct}(\cdot)$ is some suitable signal reconstruction (e.g., MT-BCS or CPPCA). The general idea of reconstruction from prediction residuals in the random-projection domain has been explored extensively in prior literature, particularly for the CS reconstruction of video; the reader is referred to [5] for a comprehensive overview. Below, we describe the process that we propose for generating predictions specifically for HSI.

A. MH Prediction

To produce a highly compressible residual, one should create a prediction that is as close as possible to \mathbf{x} , which implies that the following optimization problem is desired:

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{z} \in \mathcal{P}(\mathbf{X}_{\text{ref}})} \|\mathbf{x} - \mathbf{z}\|_2^2 \quad (3)$$

where $\mathcal{P}(\mathbf{X}_{\text{ref}})$ is the set of predictions that can be made from some reference data set, \mathbf{X}_{ref} (more on \mathbf{X}_{ref} below). However, since \mathbf{x} is unknown in CS or CPPCA reconstruction, solving (3) directly is infeasible. Instead, one approach is to reformulate (3) as

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{z} \in \mathcal{P}(\mathbf{X}_{\text{ref}})} \|\hat{\mathbf{x}} - \mathbf{z}\|_2^2 \quad (4)$$

wherein some initial reconstruction, $\hat{\mathbf{x}}$, is used as a proxy for \mathbf{x} in (3). Many forms of CS reconstruction for video take this general approach, e.g., [18], [19].

An alternative is to recast the optimization of (3) from the ambient signal domain of \mathbf{x} into the measurement domain of $\tilde{\mathbf{y}}$; specifically

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{z} \in \mathcal{P}(\mathbf{X}_{\text{ref}})} \|\mathbf{P}^T \mathbf{x} - \mathbf{P}^T \mathbf{z}\|_2^2 = \arg \min_{\mathbf{z} \in \mathcal{P}(\mathbf{X}_{\text{ref}})} \|\tilde{\mathbf{y}} - \mathbf{P}^T \mathbf{z}\|_2^2. \quad (5)$$

The Johnson–Lindenstrauss (JL) lemma [20] holds that L points in \mathbb{R}^N can be projected into a K -dimensional subspace while approximately maintaining pairwise distances as long as K is sufficiently large with respect to L . Specifically, for $\epsilon > 0$ and every set \mathcal{Q} of L points in \mathbb{R}^N , there exists mapping $f: \mathbb{R}^N \rightarrow \mathbb{R}^K$ such that, for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{Q}$

$$(1 - \epsilon) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \leq \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|_2^2 \leq (1 + \epsilon) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \quad (6)$$

as long as $K \geq O(\epsilon^{-2} \log L)$. This suggests that the solution of (5) will likely coincide with that of (3).

Instead of choosing a single prediction, or *hypothesis*, to reformulate (5), we aim to find an optimal linear combination of all hypotheses contained in the reference set $\mathcal{P}(\mathbf{X}_{\text{ref}})$; i.e., (5) becomes $\bar{\mathbf{x}} = \mathbf{H}\hat{\mathbf{w}}$ where

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|\tilde{\mathbf{y}} - \mathbf{P}^T \mathbf{H}\mathbf{w}\|_2^2. \quad (7)$$

Here, \mathbf{H} is a matrix of dimensionality $N \times K_H$ whose columns are K_H possible hypotheses in $\mathcal{P}(\mathbf{X}_{\text{ref}})$, and $\hat{\mathbf{w}}$ is a column vector which represents a linear combination of the columns of \mathbf{H} . However, because usually $K \neq K_H$, the ill-posed nature of the problem requires some kind of regularization to differentiate among the infinite number of possible linear combinations which lie in the solution space of (7).

The most common approach to regularizing a least-squares problem is Tikhonov regularization [8] which imposes an ℓ_2 penalty on the norm of $\hat{\mathbf{w}}$

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|\tilde{\mathbf{y}} - \mathbf{P}^T \mathbf{H}\mathbf{w}\|_2^2 + \lambda \|\Gamma \mathbf{w}\|_2^2 \quad (8)$$

where Γ is known as the Tikhonov matrix and λ is the regularization parameter; this strategy for MH prediction was initially proposed in [21]. The Γ term allows the imposition of prior knowledge on the solution; in some contexts, it may make sense to use a high-pass operator or a difference operator to enforce smoothness on the solution, or, in other cases, to set $\Gamma = \mathbf{I}$ to impose an energy constraint on the solution. In our case, we take the approach proposed in [21] that hypotheses which are the most dissimilar from the original pixel vector should be given less weight than hypotheses which are most similar. Specifically, a diagonal Γ takes the form of

$$\Gamma = \begin{bmatrix} \|\tilde{\mathbf{y}} - \mathbf{P}^T \mathbf{h}_1\|_2 & & 0 \\ & \ddots & \\ 0 & & \|\tilde{\mathbf{y}} - \mathbf{P}^T \mathbf{h}_{K_H}\|_2 \end{bmatrix} \quad (9)$$

where $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{K_H}$ are the columns of \mathbf{H} . With this structure, Γ penalizes weights of large magnitude assigned to hypotheses which have a significant distance from $\tilde{\mathbf{y}}$ when projected into the measurement domain. For each pixel vector, then, $\hat{\mathbf{w}}$ can be calculated directly by the usual Tikhonov solution

$$\hat{\mathbf{w}} = \left((\mathbf{P}^T \mathbf{H})^T (\mathbf{P}^T \mathbf{H}) + \lambda^2 \Gamma^T \Gamma \right)^{-1} (\mathbf{P}^T \mathbf{H})^T \tilde{\mathbf{y}}. \quad (10)$$

For the reference data set \mathbf{X}_{ref} , an initial reconstruction (i.e., using MT-BCS or CPPCA without MH prediction) of $\hat{\mathbf{X}}$ is used.¹ Furthermore, once we have reconstructed the data set via MH prediction and residual reconstruction, we can use the current reconstruction as the reference data set \mathbf{X}_{ref} for a subsequent MH prediction and residual reconstruction, further improving the quality of the reconstructed data in an iterative

¹Alternatively, some other hyperspectral data set(s) could be used as a dictionary for \mathbf{X}_{ref} depending on availability.

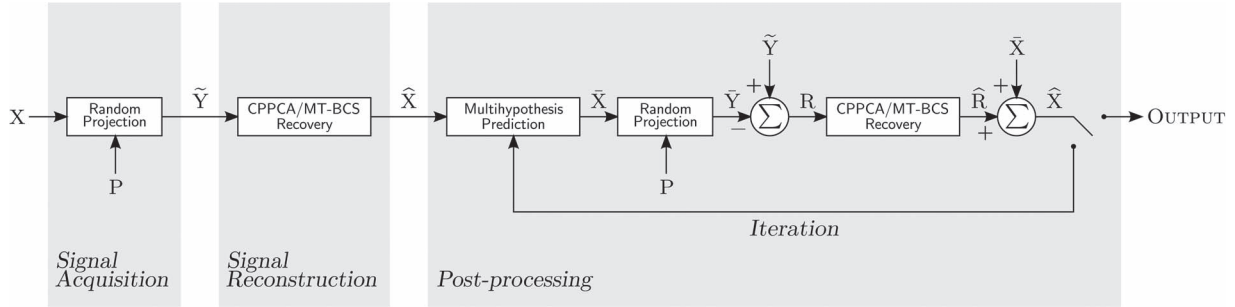


Fig. 1. MH prediction and residual reconstruction as post-processing for HSI reconstruction.

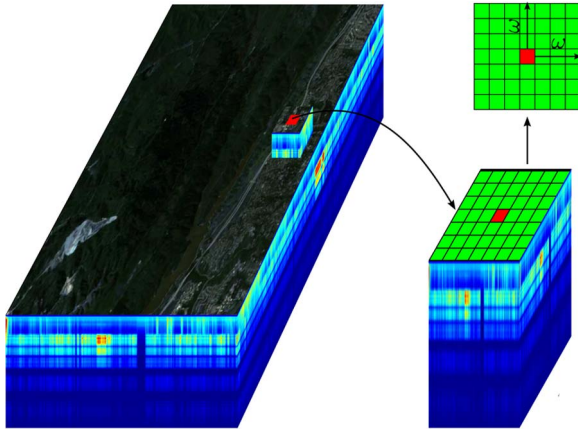


Fig. 2. Generation of multiple hypotheses from a search window with window size ω . Red indicates the current pixel vector of interest; green indicates a possible hypothesis pixel vector.

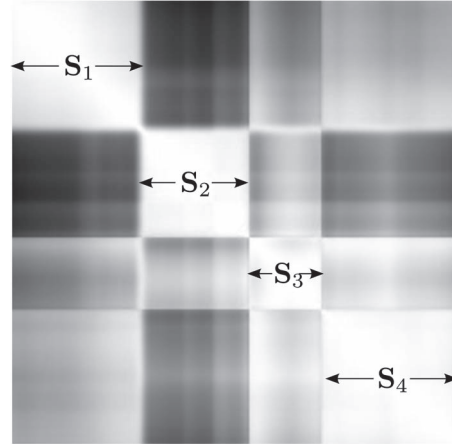


Fig. 3. 191×191 matrix of cross-band correlation coefficients of the Washington DC Mall HYDICE data set. White = ± 1 . Black = 0. S_1 , S_2 , S_3 , and S_4 are the resulting four partitions.

fashion. A general diagram of this iterative post-processing algorithm is shown in Fig. 1.

We note that others have suggested an alternative to the Tikhonov regularization used in (8); specifically, [22], [23] consider the CS reconstruction of video using multihypothesis frame-to-frame predictions, effectively assuming that the prediction weights w are sparse and consequently imposing an ℓ_1 regularization on w . However, we argue in [5], [6] that such an assumption of sparsity is needlessly restrictive for video prediction since we actually anticipate high correlation among the candidate predictions (i.e., non-sparse weights). The same is true for the hyperspectral prediction we consider here due to the spatial coherence of the hypothesis-generation process that we describe next.

B. Hypothesis Generation

Hyperspectral data usually includes a large number of homogeneous regions. For each sample, its neighboring pixel vectors will likely share similar spectral characteristics, and such contextual information has been employed previously for HSI classification [24], [25]. For the present approach, multiple hypotheses are generated for the current sample considering all the neighboring pixel vectors in a search window with size ω —this hypothesis-search procedure is illustrated in Fig. 2. Next, all hypotheses are placed as columns of the hypothesis matrix \mathbf{H} after the following partitioning procedure.

That is, since the spectral bands of a hyperspectral image are correlated, they can be partitioned into several groups based on the correlation coefficients between bands such that the bands in each group are highly correlated with one another [26]. For example, Fig. 3 illustrates the matrix of correlation coefficients between bands of the Washington DC Mall data set as well as the resulting four partitions of spectral bands. This spectral-band partitioning based on correlation coefficients is denoted as non-uniform (NU) partitioning. An alternative spectral-band partitioning is to simply partition the bands uniformly—i.e., each group has the same number of bands—this is denoted as uniform (U) partitioning. With either of the two partitioning methods, each hypothesis pixel vector can be divided into several partitions. For example, assume a N -dimensional pixel vector is divided into four partitions using uniform partitioning; then, four distinct hypothesis vectors are created by keeping only one of the four partitions while replacing the other partitions with zeros (i.e., “zero padding”) to form a N -dimensional vector. This process for hypothesis generation based on spectral-band partitioning is illustrated in Fig. 4.

If K_H hypotheses are drawn from the search window, and α partitions are used for spectral-band partitioning, then the total number of hypotheses in \mathbf{H} is αK_H . The motivation for this hypothesis generation from the partitioned spectral bands is such that the weights calculated for the hypotheses become adjustable for different spectral partitions. The details are illustrated in Fig. 5 with $\alpha = 4$ spectral-band partitions. One

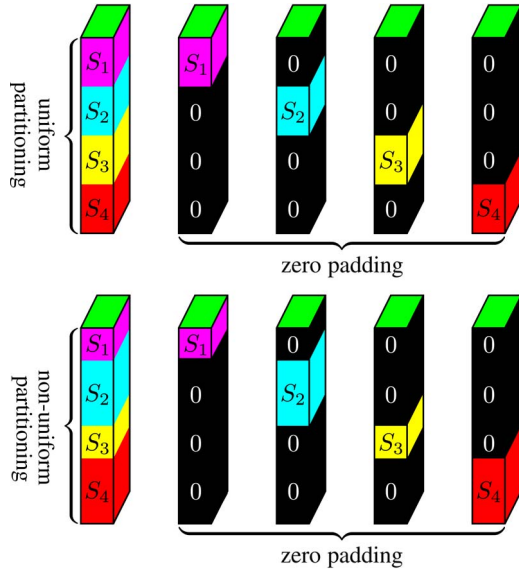


Fig. 4. Generation of hypotheses via non-uniform or uniform spectral-band partitioning and zero padding.

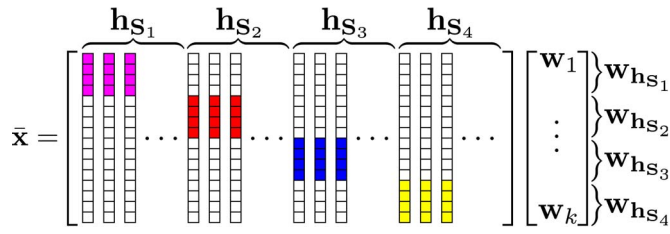


Fig. 5. Formation of a prediction using multiple hypotheses and the corresponding weights. \mathbf{h}_{S_j} ($j = 1, 2, 3,$ and 4) is the hypothesis set generated using the j th partition of the hypotheses with zero padding. $\mathbf{w}_{\mathbf{h}_{S_j}}$ are the corresponding set of weights for hypothesis set \mathbf{h}_{S_j} .

issue is the obtaining of the correlation coefficients (Fig. 3) of the original data set to perform NU spectral-band partitioning since we have only the reconstructed data set at the reconstruction side. However, we assume that the correlation coefficients between the spectral bands are fixed and known *a priori* for a given hyperspectral sensor since all images taken by this sensor would have similar correlation coefficients. Other alternatives would be to estimate the correlation coefficients from the initial reconstructed data set or simply use uniform spectral-band partitioning exclusively.

Spectral-band partitioning based on the correlation coefficients may not accurately capture the true correlation between bands since the data set at hand is the reconstructed data set, not the original. To analyze the accuracy of spectral-band partitioning based on the correlation coefficients, we take the Washington DC Mall data set as an example. According to the correlation-coefficient map of the original data set in Fig. 3, the spectral bands are divided into four partitions with band indices 1–55, 56–102, 103–132, and 133–191. For each pixel vector in the image, we generate its hypothesis matrix \mathbf{H} following the procedures described in Figs. 2 and 4 with non-uniform partitioning, storing them in the order as shown in Fig. 5.

For the m th pixel vector in the data set, we can calculate the Euclidean norms between the projection of the original

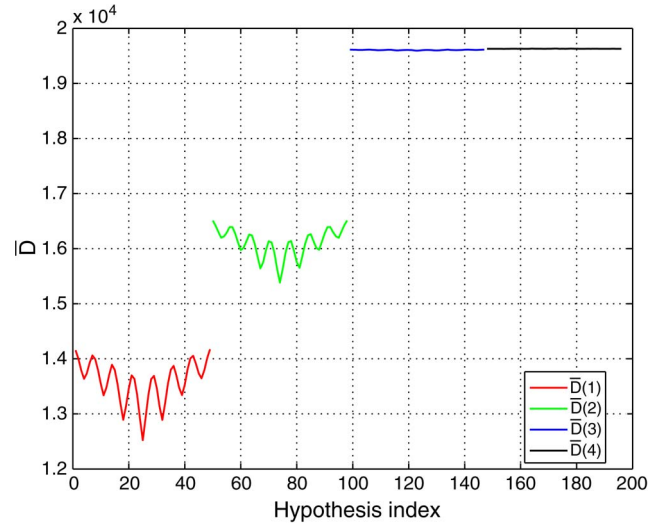


Fig. 6. Mean distance vector for all pixel vectors in the CPPCA-reconstructed Washington DC Mall data set at $K/N = 0.2$. Search window size $\omega = 3$; $\alpha = 4$ partitions in spectral-band partitioning are used for hypothesis generation.

pixel vector and the projections of all hypothesis vectors, which are the diagonal terms of Γ in (9). For projected vector $\tilde{\mathbf{y}}_m$, we define a distance vector $\mathbf{D}_m = [\|\tilde{\mathbf{y}}_m - \mathbf{P}^T \mathbf{h}_1\|_2 \dots \|\tilde{\mathbf{y}}_m - \mathbf{P}^T \mathbf{h}_{\alpha K_H}\|_2]$, where $\mathbf{h}_1, \dots, \mathbf{h}_{\alpha K_H}$ are the αK_H columns of the \mathbf{H} matrix for $\tilde{\mathbf{y}}_m$. In other words, the j th spectral-band partition generates a set of K_H hypotheses in \mathbf{H} ; this subset has a distance vector $\mathbf{D}_m(j) = [\|\tilde{\mathbf{y}} - \mathbf{P}^T \mathbf{h}_{(j-1) \times K_H + 1}\|_2 \dots \|\tilde{\mathbf{y}} - \mathbf{P}^T \mathbf{h}_{j \times K_H}\|_2]$, where $\mathbf{h}_{(j-1) \times K_H + 1}, \dots, \mathbf{h}_{j \times K_H}$ are K_H columns in \mathbf{H} . As a consequence, each pixel vector is associated with a distance vector calculated using its corresponding projection and hypothesis matrix. For a hyperspectral image with M pixel vectors, the global mean distance vector, $\bar{\mathbf{D}}$, is calculated as

$$\begin{aligned} \bar{\mathbf{D}} &= \frac{1}{M} \sum_{m=1}^M \mathbf{D}_m = \frac{1}{M} \sum_{m=1}^M [\mathbf{D}_m(1) \dots \mathbf{D}_m(\alpha)] \\ &= [\bar{\mathbf{D}}(1) \dots \bar{\mathbf{D}}(\alpha)] \end{aligned} \quad (11)$$

where $\bar{\mathbf{D}}(j)$ is the mean calculated over the data set for spectral-band partition j ($j = 1, \dots, \alpha$). For example, the mean distance vector $\bar{\mathbf{D}}$ for all pixel vectors in the initial CPPCA-reconstructed Washington DC Mall image is shown in Fig. 6 for $\alpha = 4$ spectral-band partitions.

We argue that the last two spectral-band partitions depicted in Fig. 6 can be merged together since the distance between the original pixel vector and the hypothesis vector in the projection domain is very close for hypotheses formed with the third and fourth partitions, respectively. To make this merging decision automatically, we calculate the (scalar) mean of the $\bar{\mathbf{D}}(j)$ vector as μ_j which we then normalize such that $\mu_j \in [0, 1]$. Two consecutive partitions are merged together if $|\mu_j - \mu_{j+1}| \leq \sigma$, where σ is a threshold (we use $\sigma = 0.1$ in subsequent experiments).

Consequently, we develop a two-phase spectral-band partitioning process. In the first phase, the spectral-band partitioning is based on the correlation coefficients; in the second phase,

TABLE I
SPECTRAL-BAND PARTITIONS FOR NON-UNIFORM PARTITIONING

Dataset	1 st phase	2 nd phase
Washington DC Mall (307 × 307 × 191)	1–55, 56–102 103–132, 133–191	1–55, 56–102 103–191
University of Pavia (610 × 340 × 103)	1–40, 41–68 69–76, 77–103	1–68, 69–76 77–103
Low Altitude (512 × 512 × 224)	1–40, 41–112, 113–152 153–166, 167–224	1–40, 41–112 113–224
Indian Pines (145 × 145 × 220)	1–37, 38–102, 103–150 151–162, 162–220	1–37, 38–102 103–220

the corresponding partitions of the hypotheses which share similar distances in the projection domain are merged together. The resulting spectral-band partitions for different hyperspectral images are tabulated in Table I. On the other hand, for uniform partitioning, we simply choose four partitions for the first phase and two partitions for the second phase to compare with non-uniform partitioning. We note that both the NU and U partitioning schemes we consider here are rather simple; a more sophisticated process—such as [27] which uses edge detection on the correlation-coefficient map—could equally be employed.

C. Stopping Criterion

The aforementioned multihypothesis–prediction and residual-reconstruction procedures are iterated to achieve high-quality reconstruction. As a criterion to stop these iterations, we apply cross validation [28] to predict the performance. Specifically, the orthonormal random projection $\mathbf{P} = [\mathbf{P}_R, \mathbf{P}_H]$, where $\mathbf{P} \in \mathbb{R}^{N \times K}$, $\mathbf{P}_R \in \mathbb{R}^{N \times (K-L_0)}$, $\mathbf{P}_H \in \mathbb{R}^{N \times L_0}$, and $L_0 \ll K$. For a hyperspectral image of M pixel vectors $\mathbf{X} \in \mathbb{R}^{N \times M}$, \mathbf{P}_R is used to generate random projections $\hat{\mathbf{Y}}_R \in \mathbb{R}^{(K-L_0) \times M}$ for reconstruction, and \mathbf{P}_H is used to generate random projections $\hat{\mathbf{Y}}_H \in \mathbb{R}^{L_0 \times M}$ as a holdout set for the performance test. In other words, $\hat{\mathbf{Y}} = [\hat{\mathbf{Y}}_R; \hat{\mathbf{Y}}_H]$. The residual calculated in the projected domain using the holdout set is

$$\mathcal{R} = \left\| \mathbf{P}_H^T \mathbf{X} - \mathbf{P}_H^T \hat{\mathbf{X}} \right\|_2 = \left\| \hat{\mathbf{Y}}_H - \mathbf{P}_H^T \hat{\mathbf{X}} \right\|_2. \quad (12)$$

This means that, if $\hat{\mathbf{X}}$ is close to \mathbf{X} , then \mathcal{R} should be small.

For each iteration, we obtain a residual, \mathcal{R} , for the performance test. If \mathcal{R} for the current iteration is smaller than that in the previous iteration, we can predict that the quality of the reconstructed image in the current iteration is better and thus continue the iteration. We can set a threshold τ that measures the difference in residual \mathcal{R} between two successive iterations to terminate the iteration. Algorithm 1 details the MH-CPPCA algorithm using this stopping criterion.

Algorithm 1 MH-CPPCA

Input: $\hat{\mathbf{Y}} = [\hat{\mathbf{Y}}_R; \hat{\mathbf{Y}}_H]$, $\mathbf{P} = [\mathbf{P}_R, \mathbf{P}_H]$, $\hat{\mathbf{X}}$, L (number of Ritz vectors used in CPPCA reconstruction), ω (search window size), $\mathbf{S} = \{\mathbf{S}^1, \mathbf{S}^2\}$ (spectral-band partitions in the first and second phase of spectral-band partitioning, respectively), $MaxIter$ (maximum number of iterations), $MaxPhase = 2$ (maximum number of phases), τ (a positive threshold).

Output: $\hat{\mathbf{X}}$.

Initialization: $\mathcal{R}_0 = +\infty$.

for $j = 1 \rightarrow MaxPhase$

 Set $i = 1$

while $i \leq MaxIter$ **do**

 (1) $\bar{\mathbf{X}}_i = \text{MH_Prediction}(\hat{\mathbf{X}}, \hat{\mathbf{Y}}_R, \mathbf{P}_R, \mathbf{S}^j, \omega)$

 (2) $\mathbf{R}_i = \hat{\mathbf{Y}}_R - \mathbf{P}_R^T \bar{\mathbf{X}}_i$

 (3) $\hat{\mathbf{R}}_i = \text{CPPCA}(\mathbf{R}_i, \mathbf{P}_R, L)$

 (4) Update $\hat{\mathbf{X}} \leftarrow \bar{\mathbf{X}}_i + \hat{\mathbf{R}}_i$

 (5) Calculate $\mathcal{R}_i = \hat{\mathbf{Y}}_H - \mathbf{P}_H^T \hat{\mathbf{X}}$

if $\mathcal{R}_i > \mathcal{R}_{i-1}$ **or** $\mathcal{R}_{i-1} - \mathcal{R}_i < \tau$ **then**

Break

else

$i \leftarrow i + 1$

end if

end while

end for

One drawback of using a holdout set for performance testing is that we sacrifice L_0 projections in reconstruction, which leads to a slightly degraded reconstruction as compared to that obtained in the absence of a holdout set. Experimentally, we have found that 5 to 7 iterations in each spectral-band-partitioning phase are sufficient to achieve an appropriate tradeoff between the quality of the reconstructed image and the runtime of the algorithm. In the experimental results to follow, we use a fixed 6 iterations in each phase of spectral-band partitioning to simplify the algorithm.

IV. EXPERIMENTAL RESULTS

We now validate our proposed approach with several popular HIS data sets, presenting experimental results that demonstrate the benefits of MH prediction for the reconstruction of HSI from random projections. Seven algorithms are compared, including the original CPPCA [3], C-CPPCA [10], CS reconstruction in the form of MT-BCS [9], and MH versions of both CPPCA and C-CPPCA. For the MH versions of CPPCA, we consider both non-uniform (NU) and uniform (U) partitioning for both CPPCA and C-CPPCA, resulting in variants which we call MH(NU)-CPPCA, MH(U)-CPPCA, MH(NU)-C-CPPCA, and MH(U)-C-CPPCA. All the algorithms are investigated under different substrates. Since C-CPPCA involves classification, we use K -means clustering with 6 classes for each hyperspectral data set.

For CPPCA, we use the implementation available from the CPPCA website.² For MT-BCS, we use the same random projections as used for CPPCA, a Daubechies length-4 wavelet as the sparsity basis, and the MT-BCS implementation from its authors.³ We note that recent work in dictionary learning (DL) (e.g., [29], [30]) has indicated that CS reconstruction using a learned dictionary can outperform that using a fixed basis (e.g., [29]). However, such DL-based reconstruction typically requires that the dictionary be learned on separate training data

²<http://www.ece.msstate.edu/~fowler/CPPCA/>

³<http://people.ee.duke.edu/~lcarin/BCS.html>

in its full dimensionality, precluding its use for the data sets in question for which we lack suitable training data. That said, we have made comparisons between DL-based reconstruction and CPPCA elsewhere [31] in cases for which training data was available and found that, while DL-based reconstruction can indeed outperform fixed-basis MT-BCS for hyperspectral imagery, it was inferior to even the original CPPCA. Likewise, even though there exist a myriad of techniques to perform CS reconstruction of individual hyperspectral pixel vectors from random projections, our previous results [4] indicate that such “single-task” reconstruction is vastly inferior to “multitask,” multiple-vector reconstruction. As a consequence, we focus on fixed-basis MT-BCS as the benchmark for reconstruction based on CS rather than CPPCA.

A. Experimental Hyperspectral Data

The first experimental HSI data set is Washington DC Mall taken by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor on August 23, 1995 [32]. A total of 210 bands were collected in the 0.4- to 2.40- μm region. The water-absorption bands were removed resulting an image with 191 bands. In our experiment, we crop the original image to spatial dimensions 307×307 .

The second data set used in our experiments, University of Pavia, is a urban image acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) [33]. The ROSIS sensor generates 115 spectral bands ranging from 0.43 to 0.86 μm and has a spatial resolution of 1.3 m per pixel. The image has 103 spectral bands with the 12 noisiest bands removed and a spatial coverage of 610×340 pixels.

The next data set employed was acquired using the National Aeronautics and Space Administration’s Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor and was collected over northwest Indiana’s Indian Pines test site in June 1992.⁴ The image has 145×145 pixels and 220 bands in the 0.4- to 2.45- μm region of the visible and infrared spectrum with a spatial resolution of 20 m.

The final data set, also from the AVIRIS sensor, is the Low Altitude data set. This image has 512×512 pixels and 224 bands. False-color images of all four data sets are shown in Fig. 7.

B. Optimizing Parameters

An important parameter involved in MH prediction is the search-window size, ω , used in hypothesis generation. In this section, we analyze the effect of the search-window size in terms of the quality of the final reconstructed HSI as well as the execution time of the algorithm.

To measure the quality of the reconstructed image $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1 \dots \hat{\mathbf{x}}_M]$, we use both signal-to-noise ratio (SNR) and spectral-angle distortion, as was done in [10]. We use a vector-based SNR measured in dB; i.e.,

$$\text{SNR}(\mathbf{x}_m, \hat{\mathbf{x}}_m) = 10 \log_{10} \frac{\text{var}(\mathbf{x}_m)}{\text{MSE}(\mathbf{x}_m, \hat{\mathbf{x}}_m)} \quad (13)$$

⁴<ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/>

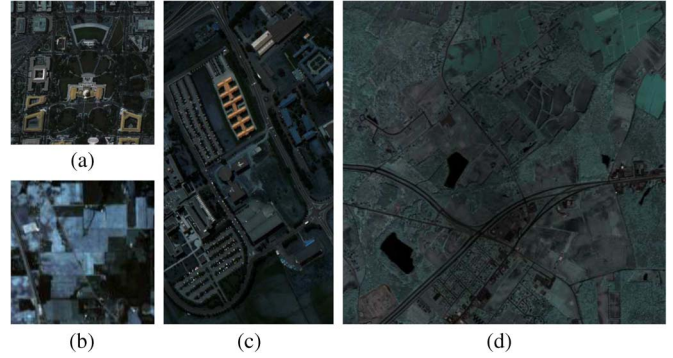


Fig. 7. False color images: (a) Washington DC Mall, using bands 30, 40, and 50 for red, green, and blue, respectively; (b) Indian Pines, using bands 10, 20, and 30 for red, green, and blue, respectively; (c) University of Pavia, using bands 20, 40, and 60 for red, green, and blue, respectively; and (d) Low Altitude, using bands 40, 50, and 60 for red, green, and blue, respectively.

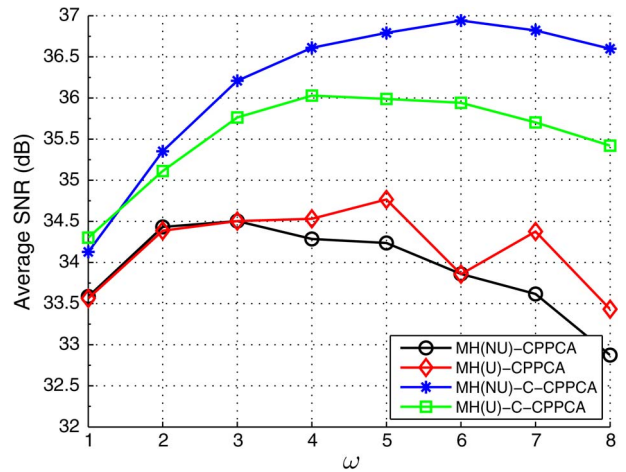


Fig. 8. Average SNR for substrate $K/N = 0.2$ for Indian Pines for various window-search sizes.

where $\text{var}(\mathbf{x}_m)$ is the variance of the components of vector \mathbf{x}_m , and the mean squared error (MSE) is

$$\text{MSE}(\mathbf{x}_m, \hat{\mathbf{x}}_m) = \frac{1}{M} \|\mathbf{x}_m - \hat{\mathbf{x}}_m\|_2^2. \quad (14)$$

The average SNR is then the vector-based SNR of (14) averaged over all vectors of the data set. Alternatively, we can define an average spectral-angle distortion by averaging the spectral angles in degrees between the reconstructed hyperspectral pixel vectors and their corresponding original pixel vectors of the data set, i.e., $\bar{\xi} = \text{mean}(\xi_m)$ where

$$\xi_m = \angle(\mathbf{x}_m - \hat{\mathbf{x}}_m). \quad (15)$$

A set of window sizes, $\omega \in \{1, \dots, 8\}$, is used for testing. Figs. 8–10 show the reconstruction performance of four MH-based algorithms at various search-window sizes for hypothesis generation. From these figures, we can conclude that larger search-window size does not necessary lead to higher reconstruction quality since some data sets may contain complex and mixed materials. In such a case, hypothesis pixel vectors drawn from a large search window could have spectral signatures rather different from the pixel vector of interest. We also find that using $\omega = 4$ takes more than twice the execution

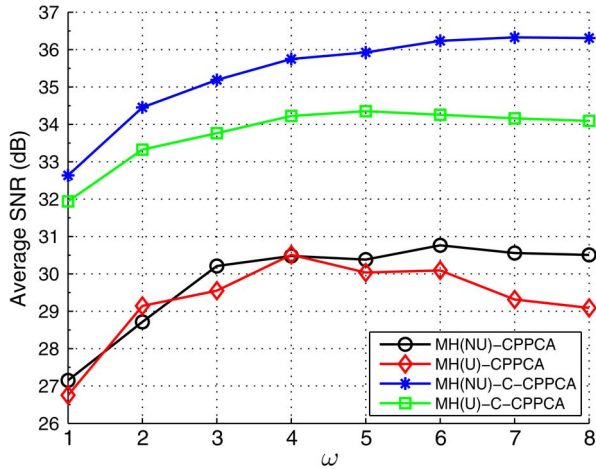


Fig. 9. Average SNR for subrate $K/N = 0.2$ for Washington DC Mall for various window-search sizes.

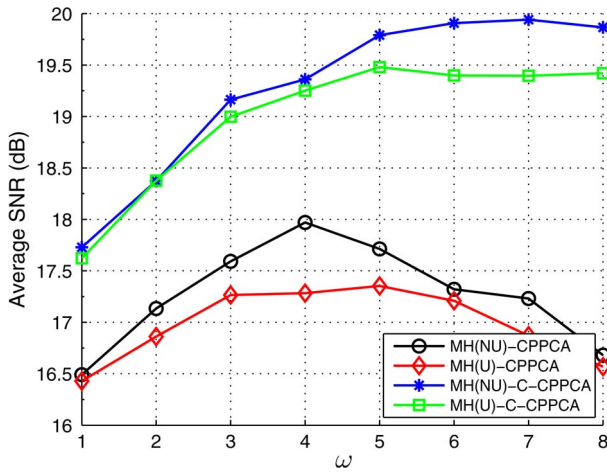


Fig. 10. Average SNR for subrate $K/N = 0.2$ for University of Pavia for various window-search sizes.

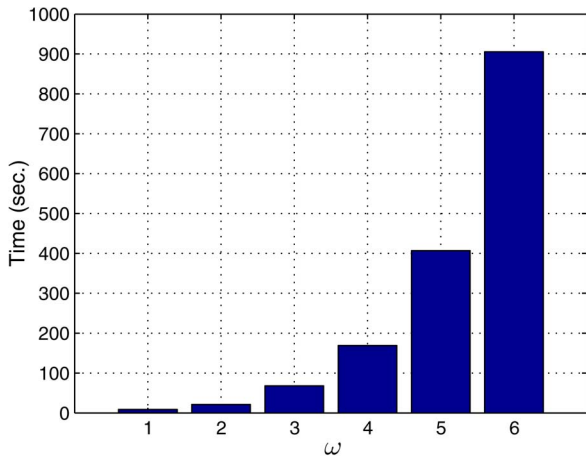


Fig. 11. Execution time for one iteration of MH prediction in MH(NU)-CPPCA for subrate $K/N = 0.3$ for Indian Pines for various search-window sizes. A quadcore 2.67-GHz machine is used.

time of $\omega = 3$, yet does not yield very much performance gain in terms of SNR. Specifically, Fig. 11 shows the execution time of one iteration of MH prediction in MH(NU)-CPPCA for

TABLE II
RECONSTRUCTION TIME FOR INDIAN PINES FOR SUBRATE $K/N = 0.3$

Algorithm	Time (sec.)
CPPCA	3.15
C-CPPCA	32.78
MH(NU)-CPPCA	612.98
MH(U)-CPPCA	391.05
MH(NU)-C-CPPCA	562.53
MH(U)-C-CPPCA	373.49
MT-BCS	3590.15

TABLE III
AVERAGE SNR IN dB

Algorithm		K/N			
		0.2	0.3	0.4	0.5
Washington DC Mall					
CPPCA	Original CPPCA	23.69	30.37	33.94	36.34
	C-CPPCA	25.47	33.41	35.86	37.24
	MH(NU)-CPPCA	30.21	39.26	43.07	46.98
	MH(U)-CPPCA	29.55	37.85	41.82	45.74
	MH(NU)-C-CPPCA	35.19	40.18	43.38	47.29
	MH(U)-C-CPPCA	33.76	38.77	42.55	46.58
	MT-BCS	20.22	24.84	28.64	31.99
Indian Pines					
CPPCA	Original CPPCA	29.34	34.02	36.42	37.62
	C-CPPCA	31.05	35.20	36.59	37.74
	MH(NU)-CPPCA	34.50	38.60	40.98	42.97
	MH(U)-CPPCA	34.51	38.65	40.95	43.17
	MH(NU)-C-CPPCA	36.21	38.78	41.05	42.93
	MH(U)-C-CPPCA	35.76	38.99	41.05	43.21
	MT-BCS	17.70	21.49	25.64	29.43
University of Pavia					
CPPCA	Original CPPCA	13.95	16.80	20.11	22.65
	C-CPPCA	16.55	19.65	21.55	23.62
	MH(NU)-CPPCA	17.59	22.03	25.11	27.55
	MH(U)-CPPCA	17.26	21.52	24.53	26.92
	MH(NU)-C-CPPCA	19.16	23.13	25.44	27.95
	MH(U)-C-CPPCA	18.99	22.48	24.94	27.34
	MT-BCS	18.54	22.90	26.05	29.02
Low Altitude					
CPPCA	Original CPPCA	26.55	30.72	33.87	36.17
	C-CPPCA	26.64	30.90	33.58	35.89
	MH(NU)-CPPCA	33.55	40.01	43.36	45.72
	MH(U)-CPPCA	31.49	39.00	43.24	45.90
	MH(NU)-C-CPPCA	35.15	40.39	43.32	45.67
	MH(U)-C-CPPCA	33.64	38.95	43.12	45.87
	MT-BCS	18.04	22.88	27.93	31.55

various search-window sizes. To balance between the runtime and the reconstruction quality, we fix $\omega = 3$ in all subsequent experiments.

Another important parameter is λ which controls the relative effect of the Tikhonov regularization term in the optimization of (8). Many approaches have been presented in the literature—such as the L-curve [34], discrepancy principle, and generalized cross-validation (GCV)—for finding an optimal value for such regularization parameters. We found that, in practice, over all the test data sets, a value of $\lambda \in [0.001, 0.05]$ provided the best results. In our experiments, we use $\lambda = \{0.01, 0.007, 0.003, 0.002\}$ for sampling subrate $K/N = \{0.2, 0.3, 0.4, 0.5\}$, respectively.

C. SNR and Spectral-Angle Performance

We now measure the quality of the reconstructed hyperspectral data sets in terms of SNR and spectral-angle distortion.

TABLE IV
AVERAGE SPECTRAL-ANGLE DISTORTION

Algorithm		K/N			
		0.2	0.3	0.4	0.5
Washington DC Mall					
CPPCA	Original CPPCA	2.691°	1.278°	0.854°	0.652°
	C-CPPCA	5.714°	0.938°	0.703°	0.601°
	MH(NU)-CPPCA	1.302°	0.475°	0.302°	0.195°
	MH(U)-CPPCA	1.413°	0.561°	0.354°	0.228°
	MH(NU)-C-CPPCA	0.749°	0.421°	0.289°	0.187°
	MH(U)-C-CPPCA	0.894°	0.499°	0.322°	0.205°
	MT-BCS	4.081°	2.372°	1.535°	1.046°
Indian Pines					
CPPCA	Original CPPCA	1.033°	0.605°	0.461°	0.401°
	C-CPPCA	0.922°	0.537°	0.457°	0.398°
	MH(NU)-CPPCA	0.570°	0.357°	0.272°	0.217°
	MH(U)-CPPCA	0.569°	0.355°	0.273°	0.212°
	MH(NU)-C-CPPCA	0.470°	0.349°	0.270°	0.218°
	MH(U)-C-CPPCA	0.498°	0.342°	0.270°	0.211°
	MT-BCS	3.926°	2.555°	1.595°	1.040°
University of Pavia					
CPPCA	Original CPPCA	3.951°	2.881°	1.948°	1.468°
	C-CPPCA	3.047°	2.076°	1.662°	1.303°
	MH(NU)-CPPCA	2.630°	1.635°	1.133°	0.871°
	MH(U)-CPPCA	2.736°	1.728°	1.198°	0.923°
	MH(NU)-C-CPPCA	2.173°	1.394°	1.071°	0.813°
	MH(U)-C-CPPCA	2.221°	1.509°	1.129°	0.866°
	MT-BCS	2.353°	1.410°	0.982°	0.700°
Low Altitude					
CPPCA	Original CPPCA	2.356°	1.473°	1.013°	0.781°
	C-CPPCA	2.351°	1.422°	1.100°	0.833°
	MH(NU)-CPPCA	0.976°	0.456°	0.301°	0.227°
	MH(U)-CPPCA	1.283°	0.521°	0.309°	0.224°
	MH(NU)-C-CPPCA	0.823°	0.435°	0.302°	0.230°
	MH(U)-C-CPPCA	0.998°	0.522°	0.314°	0.229°
	MT-BCS	5.380°	3.092°	1.740°	1.154°

The reconstruction performance of various algorithms under consideration is presented in Tables III and IV. In all cases, applying MH prediction achieves significant SNR gain over both the original CPPCA as well as C-CPPCA. In most cases, MH-C-CPPCA outperforms MH-CPPCA largely due to the fact that the initial reconstructed data set using C-CPPCA has higher SNR than that using CPPCA. We also note that, although MT-BCS achieves the highest SNR for substrates 0.4 and 0.5 for the University of Pavia data set, such high substrates are of limited practical interest. As can be seen in Table II, in terms of execution time, reconstruction with MH-CPPCA and MH-C-CPPCA is, as expected, slower than CPPCA and C-CPPCA due to iterated MH prediction, but both are much faster than MT-BCS.

Finally, we note that, as discussed in Section III-A, the MH technique that we propose here can be used in conjunction with any suitable multiple-vector reconstruction. The results in Tables III and IV focus on MH versions of CPPCA since the fast execution speed of CPPCA is more amenable to the iterative reconstruction we use. Nonetheless, it is possible to use an alternative reconstruction, such as MT-BCS. Since MT-BCS usually requires a large amount of time to reconstruct a single hyperspectral data set, we apply this MH-MT-BCS reconstruction on only the Indian Pines data set which has the smallest dimensionality among all the experimental HSI data sets under consideration. The reconstruction performance of MH-MT-BCS is shown in Table V. We see that, as compared to the corresponding CPPCA-based results for this data set shown

TABLE V
RECONSTRUCTION PERFORMANCE OF MH-MT-BCS FOR INDIAN PINES
IN SNR AND SPECTRAL-ANGLE DISTORTION

	K/N			
	0.2	0.3	0.4	0.5
MT-BCS	17.70 dB 3.926°	21.50 dB 2.555°	25.64 dB 1.595°	29.43 dB 1.040°
MH(NU)-MT-BCS	30.92 dB 0.860°	37.75 dB 0.394°	41.34 dB 0.261°	44.21 dB 0.187°
MH(U)-MT-BCS	31.72 dB 0.785°	38.58 dB 0.358°	41.74 dB 0.249°	43.68 dB 0.199°

in Table III, MH-MT-BCS achieves slightly higher SNR for the higher (and less practically relevant) substrates 0.4 and 0.5, but the MH variants of C-CPPCA perform better at the lower substrates.

V. CONCLUSION

In this paper, we proposed exploitation of the spatial correlation of neighboring pixels in the reconstruction from random projections of hyperspectral data. For each pixel vector in a hyperspectral data set, multiple predictions were drawn from the spatial surrounding pixel vectors. To take advantage of the fact that the spectral bands of a hyperspectral image are correlated, we proposed a two-phase hypothesis-generation procedure based on partitioning and merging of the spectral bands according to the correlation coefficients between bands. We formed MH prediction using a distance-weighted Tikhonov regularization to find the best linear combination of hypotheses. The MH prediction was then used to create a measurement-domain residual of the signal to be recovered—such a residual is typically more compressible than the original signal making it more amenable to the reconstruction. The reconstructed data sets using MH prediction showed significant gain in SNR and spectral-angle distortion over several non-predicted reconstruction techniques.

REFERENCES

- [1] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Exp.*, vol. 15, no. 21, pp. 14013–14027, Oct. 2007.
- [2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [3] J. E. Fowler, "Compressive-projection principal component analysis," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2230–2242, Oct. 2009.
- [4] J. E. Fowler and Q. Du, "Reconstructions from compressive random projections of hyperspectral imagery," in *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, S. Prasad, L. M. Bruce, and J. Chanussot, Eds. New York, NY, USA: Springer-Verlag, 2011, ch. 3, pp. 31–48.
- [5] J. E. Fowler, S. Mun, and E. W. Tramel, "Block-based compressed sensing of images and video," *Found. Trends Signal Process.*, vol. 4, no. 4, pp. 297–416, Mar. 2012.
- [6] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-sensing recovery of images and video using multihypothesis predictions," in *Proc. 45th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2011, pp. 1193–1198.
- [7] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Minneapolis, MN, USA, Apr. 1993, vol. 5, pp. 437–440.
- [8] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC, USA: V. H. Winston & Sons, 1977.
- [9] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 92–106, Jan. 2009.

- [10] W. Li, S. Prasad, and J. E. Fowler, "Classification and reconstruction from random projections for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 833–843, Feb. 2013.
- [11] M. Fornasier and H. Rauhut, "Recovery algorithms for vector-valued data with joint sparsity constraints," *SIAM J. Numer. Anal.*, vol. 46, no. 2, pp. 577–613, Mar. 2008.
- [12] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
- [13] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.
- [14] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [15] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, Mar. 2006.
- [16] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Philadelphia, PA, USA: SIAM, 1998.
- [17] W. Li and J. E. Fowler, "Decoder-side dimensionality determination for compressive-projection principal component analysis of hyperspectral data," in *Proc. Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 321–324.
- [18] S. Mun and J. E. Fowler, "Residual reconstruction for block-based compressed sensing of video," in *Proc. Data Compress. Conf.*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, USA, Mar. 2011, pp. 183–192.
- [19] H. Jung and J. C. Ye, "Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques," *Int. J. Imag. Syst. Technol.*, vol. 20, no. 2, pp. 81–98, Jun. 2010.
- [20] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemp. Math.*, vol. 26, pp. 189–206, 1984.
- [21] E. W. Tramel and J. E. Fowler, "Video compressed sensing with multihypothesis," in *Proc. Data Compress. Conf.*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, USA, Mar. 2011, pp. 193–202.
- [22] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proc. Int. Conf. Image Process.*, Cairo, Egypt, Nov. 2009, pp. 1393–1396.
- [23] J. Prades-Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. Picture Coding Symp.*, Chicago, IL, USA, May 2009, pp. 1–4.
- [24] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
- [25] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [26] Q. Du, W. Zhu, H. Yang, and J. E. Fowler, "Segmented principal component analysis for parallel compression of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 713–717, Oct. 2009.
- [27] Y.-Q. Zhao, L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 747–756, Feb. 2011.
- [28] R. Ward, "Compressed sensing with cross validation," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, Dec. 2009.
- [29] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.
- [30] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral images," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 33–56, Jan. 2012.
- [31] N. H. Ly, Q. Du, and J. E. Fowler, "Reconstruction from random projections of hyperspectral imagery with spectral and spatial partitioning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 466–472, Apr. 2013.
- [32] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley-Interscience, 2003.
- [33] P. Gamba, "A collection of data for urban area characterization," in *Proc. Int. Geosci. Remote Sens. Symp.*, Anchorage, AK, USA, Sep. 2004, vol. 1, pp. 69–72.
- [34] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993.



Chen Chen (S'10) received the B.E. degree in automation from Beijing Forestry University, Beijing, China, in 2009 and the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, in 2012. He is currently working toward the Ph.D. degree in the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX.

His research interests include compressed sensing, signal and image processing, pattern recognition, and computer vision.



Wei Li (S'11–M'13) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, in 2012.

Subsequently, he spent one year as a Postdoctoral Researcher at the University of California, Davis, CA, and will soon join the College of Informa-

tion Science and Technology at Beijing University of Chemical Technology, Beijing, China. His research interests include statistical pattern recognition, hyperspectral image analysis, and data compression.

Dr. Li is an active reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.



Eric W. Tramel (S'08–M'13) received the B.S. and Ph.D. degrees in computer engineering from Mississippi State University, Starkville, MS, in 2007 and 2012, respectively.

In 2011, he was a Research Intern at Canon USA, Inc. from May to August. He also served as a Research Associate for the Geosystems Research Institute (GRI) at Mississippi State, MS, from 2009 to 2012. His research interests include compressed sensing, image and video coding, image and video multiview systems, data compression, and pattern

recognition.



James E. Fowler (S'91–M'96–SM'02) received the B.S. degree in computer and information science engineering and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus in 1990, 1992, and 1996, respectively.

In 1995, he was an Intern Researcher at AT&T Labs, Holmdel, NJ, and, in 1997, he held an NSF-sponsored postdoctoral assignment at the Université de Nice-Sophia Antipolis, France. In 2004, he was a Visiting Professor in the Département TSI at École Nationale Supérieure des Télécommunications

(ENST), Paris, France. He is currently Billie J. Ball Professor and Graduate Program Director of the Department of Electrical and Computer Engineering at Mississippi State University in Starkville, MS. He is also a Researcher in the Geosystems Research Institute (GRI) at Mississippi State, MS.

Dr. Fowler is an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the *EURASIP Journal on Image and Video Processing*. He formerly served as an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA and IEEE SIGNAL PROCESSING LETTERS. He is the Chair of the Image, Video, and Multidimensional Signal Processing Technical Committee of the IEEE Signal Processing Society and a member of the Strategic Planning Committee of the IEEE Publication Services and Products Board. He is General Co-Chair of the 2014 IEEE International Conference on Image Processing, Paris, France, as well as the Publicity Chair of the program committee for the Data Compression Conference.