Unsupervised hyperspectral band selection in the compressive sensing domain

Bernard Lampe, Adam Bekit, Charles Della Porta, Bai Xue, Chein-I Chang

Bernard Lampe, Adam Bekit, Charles Della Porta, Bai Xue, Chein-I Chang, "Unsupervised hyperspectral band selection in the compressive sensing domain," Proc. SPIE 10986, Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXV, 109860A (14 May 2019); doi: 10.1117/12.2517439

SPIE.

Event: SPIE Defense + Commercial Sensing, 2019, Baltimore, Maryland, United States

Unsupervised Hyperspectral Band Selection in the Compressive Sensing Domain

Bernard Lampe*, Adam Bekit, Charles Della Porta, Bai Xue and Dr. Chein-I Chang

Remote Sensing Signal and Image Processing Laboratory
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County (UMBC), Baltimore, MD 21250

ABSTRACT

Band selection (BS) algorithms are an effective means of reducing the high volume of redundant data produced by the hundreds of contiguous spectral bands of Hyperspectral images (HSI). However, BS is a feature selection optimization problem and can be a computationally intensive to solve. Compressive sensing (CS) is a new minimally lossy data reduction (DR) technique used to acquire sparse signals using global, incoherent, and random projections. This new sampling paradigm can be implemented directly in the sensor acquiring undersampled, sparse images without further compression hardware. In addition, CS can be simulated as a DR technique after an HSI has been collected. This paper proposes a new combination of CS and BS using band clustering in the compressively sensed sample domain (CSSD). The new technique exploits the incoherent CS acquisition to develop BS via a CS transform utilizing inter-band similarity matrices and hierarchical clustering. It is shown that the CS principles of the restricted isometric property (RIP) and restricted conformal property (RCP) can be exploited in the novel algorithm coined compressive sensing band clustering (CSBC) which converges to the results computed using the original data space (ODS) given a sufficient compressive sensing sampling ratio (CSSR). The experimental results show the effectiveness of CSBC over traditional BS algorithms by saving significant computational space and time while maintaining accuracy.

Keywords: Data Reduction (DR), Band Selection (BS), Band Clustering (BC), Compressive Sensing (CS), Restricted Isometric Property (RIP), Restricted Conformal Property (RCP), Compressive Sensing Sampling Rate (CSSR)

1. INTRODUCTION

Hyperspectral sensors collect hundreds of discrete spectral measurements per pixel. Hyperspectral images (HSI) containing tens of thousands of pixels produce large data sets which pose challenges to detection, classification, endmember finding, and umixing algorithms [1]–[3]. However, the narrow spectral range of the contiguous HSI bands results in highly correlated samples due to spatial and spectral locality. Therefore, the information entropy of such images is considerably lower than the data bandwidth required by the Shannon-Nyquist sampling theorem [4], [5]. As a consequence, data reduction (DR) algorithms are effective at compressing HSI. DR algorithms such as principal component analysis (PCA) [6], minimum noise fraction (MNF) [7] or independent component analysis (ICA) [8] seek an efficient lower dimensional projector which captures the salient features of the image required for further image exploitation algorithms. However, these DR approaches linearly combine the spectral and spatial dimensions of HSI which can hinder further processing. Band selection (BS) has an advantage over other DR algorithms because it does not change the semantics of the image samples which allows for HSI data processing directly on the reduced data set [9]. In addition, many exploitation algorithms have minimal performance loss when applied to the selected band subset (SBS) [10], [11]. However, BS is a feature selection optimization algorithm and can be computationally intensive. This study leverages the compressive sensing (CS) incoherent sampling to reduce the computational burden of BS algorithms with minimal effect to the SBS computed [12].

Compressive sensing is a new paradigm in the acquisition of sparse signals which captures global, incoherent, and linear projections of the true signal [13], [14]. Encoding signals using this scheme allows data sampling below the Shannon-Nyquist rate and near the signal information entropy with minimal loss of fidelity. CS can be theoretically implemented in the sensor hardware or used as a DR technique by simulating acquisition of a collected HSI. Ideally, HSI data processing

*Bernard Lampe: E-mail: lampeb1@umbc.edu

Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXV,
edited by Miguel Velez-Reyes, David W. Messinger, Proc. of SPIE Vol. 10986, 109860A
 · doi: 10.1117/12.2517439

algorithms such as BS would take advantage of compressively sensed bands directly in the undersampled form. This avoids the costly non-linear reconstruction process [15]. Prior work has been published proposing that random linear projections of pixels has little effect on target detection, band selection and classification algorithm performance [16], [17]. However, these works were limited to compressing the spectral dimension and do not address the underlying mathematical theory. In addition, the prior art did not address applications of CS to the plethora of BS algorithms or measures.

This study develops a new combination of CS and BS using hierarchical band clustering and inter-band similarity measures based on the Euclidian distance (ED) and spectral angle mapper (SAM) in the compressively sensed sample domain (CSSD) [18]. By reducing the number of samples in within each spectral band using CS, it is possible to exploit the fundamental length and angle preserving properties of CS to maintain the inter-band measurements. This allows band clustering to be performed directly on the reduced band images and compute the same SBS with considerably lower amounts of data thus decreasing computational time.

2. HYPERSPECTRAL BAND SELECTION

As detailed in Chang 2013 [1], unsupervised band selection (BS) algorithms must solve a number of sub-problems. Let the three dimensional HSI be denoted by $\mathbf{I} \in \mathcal{R}^{n_1 x n_2 x L}$ and $N = n_1 x n_2$ as the size of the spatial dimension of N pixels and L spectral bands. Then let each band be $\mathbf{B}_l \in \mathcal{R}^{n_1 x n_2}$ and the full band set of L bands be $\Omega = \{\mathbf{B}_l\}_{l=1}^{L}$ and the selected band set of p bands be $\widetilde{\Omega} = \{\widetilde{\mathbf{B}}_j\}_{j=1}^{p}$. The first problem of BS algorithms is determining the cardinality p of the SBS. The second problem is finding which bands to choose based on an optimality criteria. The final problem to consider is a search strategy over all the band subsets. This study sets the cost of the SBS expressed by the generic objective function $J(\widetilde{\Omega})$. There are L choose p SBS combinations resulting in a exponentially large search space of size $L!/(p!\,(L-p)!)$. In addition, the search space is multi-modal giving rise to more than one optimal solution. The goal of any BS algorithm is to find the optimal band subset which minimizes or maximizes the cost function as in equation (1).

$$\widetilde{\Omega}^* = \arg\left\{\max/\min_{\widetilde{\Omega}\subset\Omega,|\Omega|=p} J(\widetilde{\Omega})\right\} \tag{1}$$

A number of approaches have been used to solve the sub-problems of BS. Firstly virtual dimensionality(VD) has been well studied and used to choose the number of bands to select [19], [20]. VD estimates the number of distinct linearly-separable spectral signatures in the image which is analogous to finding the minimum number of bands to choose to preserve the separability [21]. The second problem of optimizing (1) requires implementing an optimization criteria $J(\widetilde{\Omega})$ to perform band prioritization and/or band decorrelation. Prioritization algorithms evaluate bands individually based on content using criteria such as variance, entropy and channel capacity. In BS it is desirable to avoid redundant band information, therefore band decorrelation uses inter-band measures to find a subset which removes redundant bands [22]. Orthogonal subspace projection as well as mutual information have been explored to solve BS using band decorrelation [23], [24]. Finally, a plethora of search strategies have been put forth in the literature [10], [25], [26].

This study utilizes VD to find the number of bands to choose, then takes advantage of band decorrelation measures which are preserved by CS of length and angle. Firstly the bands are compressively sampled and then inter-band similarity is computed. Finally, bands are clustered into similar groups with the goal of minimizing inter-band similarity.

3. COMPRESSIVE SENSING

3.1 Hyperspectral Compressive Intra-Band Sampling

This research takes advantage of CS in a novel manner compared to other HSI algorithms which only utilize sparsity or focus on image reconstruction optimization algorithms [27]. The HSI bands are considered sparse and the incoherent

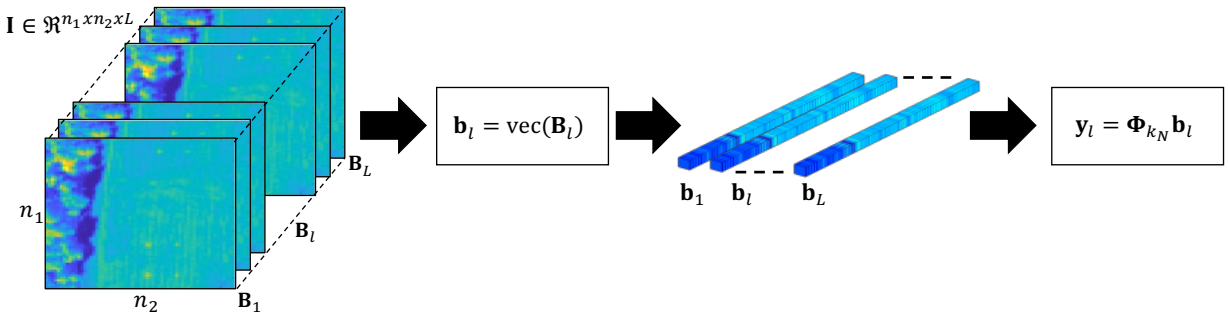


Figure 1. Compressively sensing the set of L HSI bands to the compressively sensed sample domain

sensing is used to reduce the band images while preserving the spectral dimension [15]. Let a band vector (BV) be formed $\mathrm{vec}(\mathbf{B}_l) = \mathbf{b}_l = (b_{l1}, b_{l2}, \ldots, b_{lN})^T \in \mathcal{R}^{Nx1}$ or a selected BV be $\mathrm{vec}(\widetilde{\mathbf{B}}_j) = \tilde{\mathbf{b}}_j$. CS can be applied to the l^{th} BV via linear projection using equation (2) where $\mathbf{\Phi}_{k_N} \in \mathcal{R}^{m_N x N}$ is used reduce the l^{th} band, $\mathbf{b}_l$ of N samples to $k_N < m_N \ll N$ samples [28]. Here k_N is the sparsity of the l^{th} band vector $\mathbf{b}_l$ and the matrix $\mathbf{\Phi}_{k_N}$ can be used to sense any k_N-sparse signal.

$$\mathbf{y}_l = \mathbf{\Phi}_{k_N}\mathrm{vec}(\mathbf{B}_l) = \mathbf{\Phi}_{k_N}\mathbf{b}_l \tag{2}$$

The design of the sampling matrix $\mathbf{\Phi}_{k_N}$ is critical to CS. The first assumption of CS is that BVs are sparse in some domain $\|\mathbf{\Psi}\mathbf{b}_l\|_0 < k_N$ where $\mathbf{\Psi}$ represents the sparse transform [14]. Due to the universality of CS, the optimal domain does not need to be known prior to reconstruction [29], [30]. The second assumption of CS is that $\mathbf{\Phi}_{k_N}$ is incoherent with $\mathbf{\Psi}$. Incoherence is defined in (3) and it is required to be near 1 for pairs of sensing matrix column vectors $\mathbf{\Phi}_k = [\boldsymbol{\phi}_1|\boldsymbol{\phi}_2|\ldots|\boldsymbol{\phi}_N]$ and sparse transform $\mathbf{\Psi} = [\boldsymbol{\psi}_1|\boldsymbol{\psi}_2|\ldots|\boldsymbol{\psi}_N]$.

$$\min_{1<i\neq j<N}\{\boldsymbol{\phi}_i{}^{\mathrm{T}}\boldsymbol{\psi}_j\} \approx 1 \tag{3}$$

Incoherence must be achieved with high probability in the absence of an explicitly chosen $\mathbf{\Psi}$. Therefore, CS generates $\mathbf{\Phi}_{k_N}$ using a random matrix. The Gaussian distribution with normalized columns is widely-used and well-studied for this application [28]. When creating $\mathbf{\Phi}_{k_N}$ the compressive sensing sampling rate (CSSR) defined as m_N/N is a crucial parameter and given a Gaussian matrix must satisfy (4) for all bands.

$$m_N \geq 2k_N \log\left(\frac{N}{k_N}\right) \tag{4}$$

A typical approach to finding m_N hinders on finding k_N [31]. Unfortunately, finding an optimal sparse transform $\mathbf{\Psi}$ has been shown to be a difficult problem and a new approach to finding the CSSR is given in the following section [32]. When m_N is sufficiently high, two properties central to CS are preserved and can be exploited. The first is the restricted isometry property (RIP) in (5) and the second derived from the RIP, is the restricted conformal property (RCP) in (6) [33], [34]. These properties ensure that the lengths of sensed bands $\|\mathbf{b}_l\|_2^2$ are preserved and the angle between sensed bands $\theta = \angle(\mathbf{b}_{l_1}\mathbf{b}_{l_2})$ is approximately preserved under CS where α is the angle between sensed bands.

$$(1-\delta_k)\|\mathbf{b}_l\|_2^2 \leq \left\|\mathbf{\Phi}_{k_N}\mathbf{b}_l\right\|_2^2 \leq (1+\delta_k)\|\mathbf{b}_l\|_2^2 \tag{5}$$

$$\frac{(1-\delta_k)}{(1+\delta_k)}\cos\theta \leq \cos\alpha \leq \frac{(1+\delta_k)}{(1-\delta_k)}\cos\theta \tag{6}$$

The restricted isometric constant (RIC) δ_k goes to zeros as the number of samples increases $\delta_k \to 0$ as $m_N \to N$ [12]. The RIC does not go to zero uniformly but does so exponentially in the number of compressive samples. This demonstrates the data reduction power of CS. The RIC is defined as the minimum value which satisfies the following inequality.

$$\left\|\boldsymbol{\Phi}_{k_N}{}^{\mathrm{T}}\boldsymbol{\Phi}_{k_N} - \mathbf{I}\right\|_2 \leq \delta_k \tag{7}$$

3.2 Compressive Sensing Reconstruction

The non-linear CS reconstruction process is an optimization over the L_1-norm of the possible solutions as denoted in (8). Two ubiquitous methods to solving this problem are basis pursuit (BP) or orthogonal matching pursuit (OMP) [35], [36]. Reconstruction decodes each sensed band one at a time. Here it is reiterated that the sparse transform $\boldsymbol{\Psi}$ is only needed for reconstruction due to universality and will not be specified this paper.

$$(\boldsymbol{\Psi}\mathbf{b}_l)^* = \arg\min_{\boldsymbol{\Psi}\mathbf{b}_l}\|\boldsymbol{\Psi}\mathbf{b}_l\|_1 \text{ subject to } \mathbf{y}_l = \boldsymbol{\Phi}_{k_N}\mathbf{b}_l \tag{8}$$

3.3 Compressive Sensing Sampling Rate Estimation

The compressive sensing sampling rate (CSSR) denoted $0 \leq m_N/N \leq 1$ is a critical design parameter of the sampling matrix $\boldsymbol{\Phi}_{k_N} \in \mathcal{R}^{m_N x N}$ [37]. Setting this rate too low results in poor reconstruction performance and unacceptable data loss. Setting this rate too high will lead to an inefficient implementation. Sequential CS (SCS) applications obtain samples one at a time and then test for sufficiency between each sample [38], [39]. Current research does this by performing reconstruction in line with sampling. However, reconstruction has a computational complexity of $O(N^3)$. It is desirable to reduce this computational complexity and perform SCS with a light-weight and recursive analysis.

Each compressive sample collected is an inner product between the rows of the sampling matrix $\boldsymbol{\Phi}_k = [\boldsymbol{\varphi}_1^T, \boldsymbol{\varphi}_2^T, \dots, \boldsymbol{\varphi}_m^T]$ and the BV such that $y_i = \langle \boldsymbol{\varphi}_i, \mathbf{b}_l \rangle$ for $1 \leq i \leq m$. Breaking the sampling vector $\boldsymbol{\varphi}_i^T = (\phi_{i1}, \phi_{i2}, \dots, \phi_{iN})^T$ into scalars, \ the inner product can be expressed as a summation of Gaussian random variables ϕ and the BV components $\mathbf{b}_l = \left(b_1, b_2, \dots, b_N\right)^T$. Therefore the scalar components of $\mathbf{y}_l = (y_1, y_2, \dots, y_m)^T$ can be cast as Gaussian random variable arising from equation (9) [40].

$$y_i = \sum_{j=1}^{N} \phi_{ij} b_j \quad \text{where } \phi_{ij} \sim \mathrm{N}(0,1) \tag{9}$$

With the $y_i \sim \mathrm{N}(0, \sum_{j=1}^{n} b_j^2)$, the information conveyed in the encoding process is preserved in the variance of the entries of y_i. In order to find the dimensionality m_N of $\mathbf{y} = (y_1, y_2, \dots y_m)^T$, the question becomes, how many samples are needed to estimate the Gaussian distribution of $\mathrm{N}(0, \sum_{j=1}^{n} b_j^2)$? This is a density estimation problem and has been well studied [41]. While there are many techniques to solve density estimation, employing a method that is amenable to the sequential collection and computation is desired. This will minimize the work needed by the encoding process at each iteration. A simple method to determine convergence is to estimate the kurtosis $\hat{\beta}(y_1, y_2, \dots)$ of the collected samples and then determine when the estimate converges value of 3 for excess kurtosis. Figure 2 gives an overview of the estimation method. Here using $\hat{\beta}_i = \hat{\beta}_{i-1} + \delta_i$ is the online kurtosis estimate which relies on the recursive form of the first three moments mean, variance and, skewness . This method has a considerably lower computational complexity than reconstruction and estimates the CSSR $\hat{m}$ efficiently in line with CS encoding.

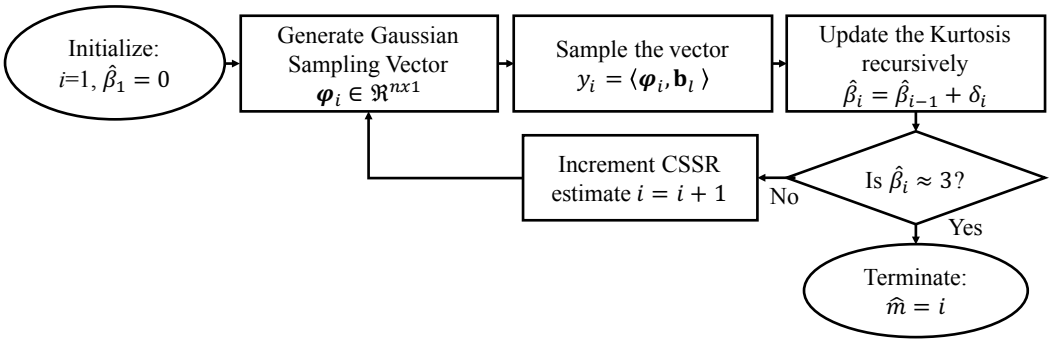


Figure 2. Flow chart of the recursive kurtosis CSSR estimation method

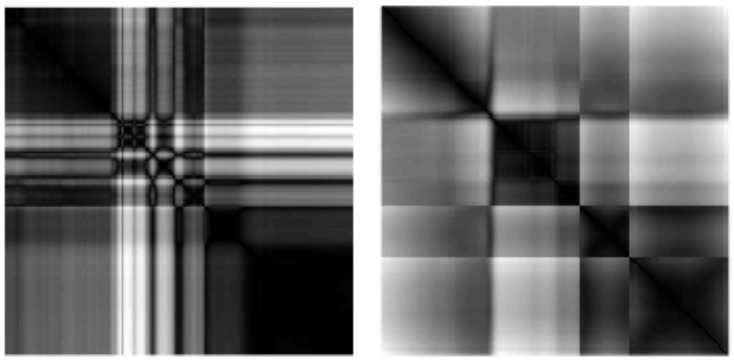
4. HYPERSPECTRAL BAND CLUSTERING

4.1 Similarity Matrix and Hierarchical Clustering

Agglomerative hierarchical clustering begins with each band in a separate cluster $c_l = \{\mathbf{B}_l\}$ for $1 \leq l \leq L$. Exploiting the fact that adjacent bands are highly correlated dramatically reduces the optimization problem in (1). Therefore, clustering can merge adjacent bands into larger clusters by splitting the available band range into two clusters and then recursively splitting on those clusters until the desired number of bands is achieved. The criteria used for this clustering is the average similarity between cluster members in (10) [24]. The distance measures used in this study are the Euclidean distance (ED) and the spectral angle mapper (SAM). These are chosen because they are preserved by (5) and (6) in the CSSD.

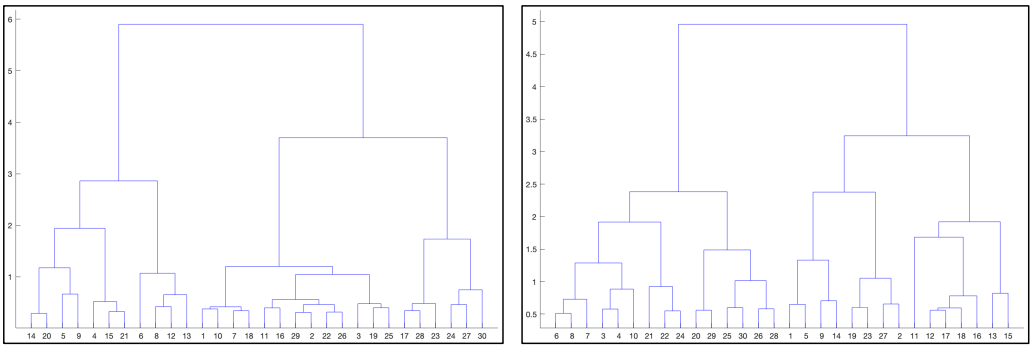
$$d(c_i, c_j) = \frac{1}{|c_i||c_j|} \sum_{\mathbf{B}_1 \in c_i} \sum_{\mathbf{B}_2 \in c_j} dist(\mathbf{B}_1, \mathbf{B}_2) \tag{10}$$

The similarity measure between all bands is depicted in figure 3 rendering all values of $dist(\mathbf{B}_i, \mathbf{B}_j)$ when each band is a separate cluster. These matrices suggest the clustering which exists due to the highly correlated spectral bands. The rows of the similarity matrices themselves are used when clustering as a surrogate for the bands themselves. The recursive splitting of the clusters can be visualized in a dendrogram in figure 4. The final SBS is found as the bands who are closest to the cluster centroids using Euclidean distance. The key to CSBC is that the similarity matrices are preserved due to the RIP and RCP. Therefore, CS can be used to reduce the band images using (2) and then the similarity matrices can be computed using the reduced bands.



(a) (b)

Figure 3. Similarity matrix between bands using (a) ED and using (b) SAM



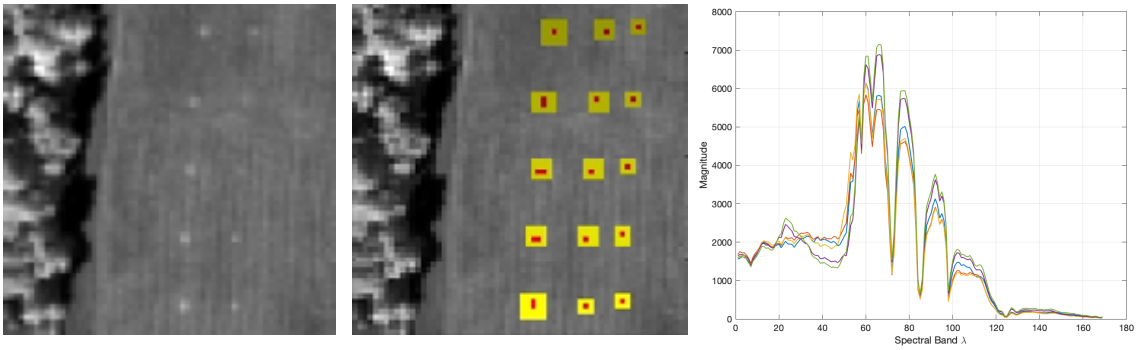
(a) (b)

Figure 4. Dendrogram of the hierarchical clustering found via recursive splits

5. HYPERSPECTRAL EXPERIMENT RESULTS

5.1 Real Hyperspectral Test Image

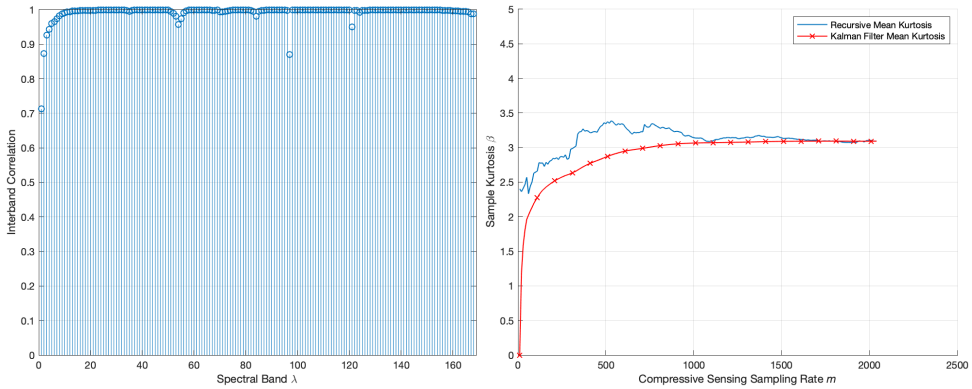
The data set used for experiment is the ubiquitous Hyperspectral Digital Imagery Collection Experiment (HYDICE) image collected using the airborne visible infrared imaging spectrometer (AVIRIS) [42]. The sensor has a spectral range of 400–2500 nm with 224 bands at 10 nm resolution. The scene was collected in August 1995 and consists of 64 x 64 pixels at 1.56 m ground spatial distance. The low SNR bands were removed consisting of 1-3 and 202-210. In addition, water vapor absorption bands 101-112 and 137-153 were removed with 169 bands remaining. The scene has a large grass field, 15 painted target panels, road and a forest left of the field. The panels are arranged in a five by three grid. The ground truth map is in Figure 5(b). The red pixels are the ground truth and the yellow pixels are the surrounding pixels. The red pixels in the first two columns are considered full pixel signatures and the third column is sub-pixel targets.



(a) (b) (c)

Figure. 5. HYDICE image (a) band 80, (b) band 80 with ground truth overlay, and (c) 5 panel signatures

The assumption of many BS algorithms is that the inter-band correlations are high. Plotting the Pearson correlation between bands is done in figure 6(a) and shows that information redundancy is high across the spectrum. VD is used to determine the number of bands used in the experiment and is estimated to be 9 bands using the HYDICE image. The final parameter input to the CSBC algorithm is the CSSR. Using the algorithm outlined in figure 2, the CSSR is estimated to be approximately $m_N = 1024$ which was derived from the curve in figure 6(b). This leads to a CSSR of 0.25 implying only a quarter of the number of compressive samples are required to match the results using the ODS.



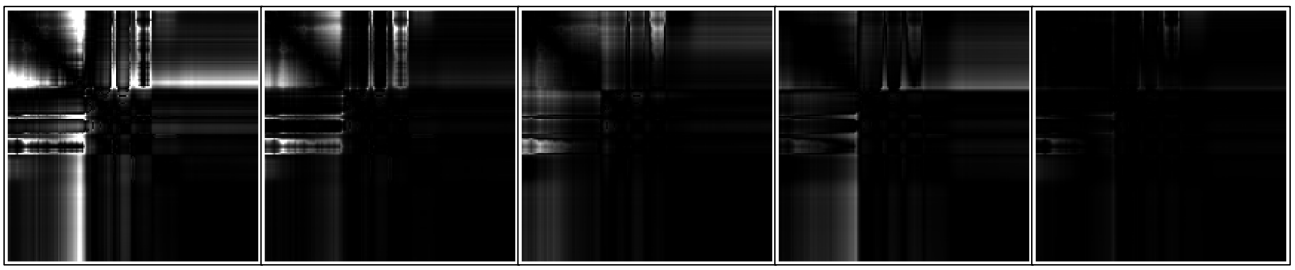
(a) (b)

Figure 6. (a) HYDICE inter-band correlation and (b) the recursive kurtosis CSSR estimate

5.2 Preservation of Band Similarity

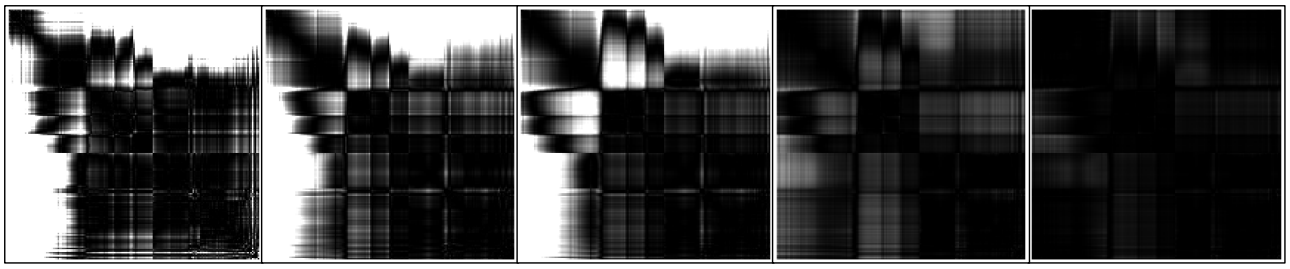
The key insight to CSBC is the preservation of the ED and SAM using the RIP and RCP properties of CS used to compute the distance between clusters in (10). An experiment was conducted to show the accuracy by computing the difference between the similarity matrices using the ODS and the sensed bands. Figures 7 and 8 show the absolute value of the error between the similarity matrices in the ODS and CSSD using equation (11).

$$\epsilon_{SIM}(i,j) = \left| dist\left(\mathbf{b}_i, \mathbf{b}_j\right) - dist\left(\mathbf{\Phi}_{k_N}\mathbf{b}_i, \mathbf{\Phi}_{k_N}\mathbf{b}_j\right), \right| \quad \text{for } 1 \le i \le L, 1 \le j \le L \tag{11}$$



(a) (b) (c) (d) (e)

Figure 7. Absolute error between similarity matrices using ED at increasing CSSR (a) $m_N = 8$ (b) $m_N = 16$, (c) $m_N = 32$, (d) $m_N = 64$, and (e) $m_N = 128$



(a) (b) (c) (d) (e)

Figure 8. Absolute error between similarity matrices using SAM at increasing CSSR (a) $m_N = 32$ (b) $m_N = 128$, (c) $m_N = 256$, (d) $m_N = 512$, and (e) $m_N = 1024$

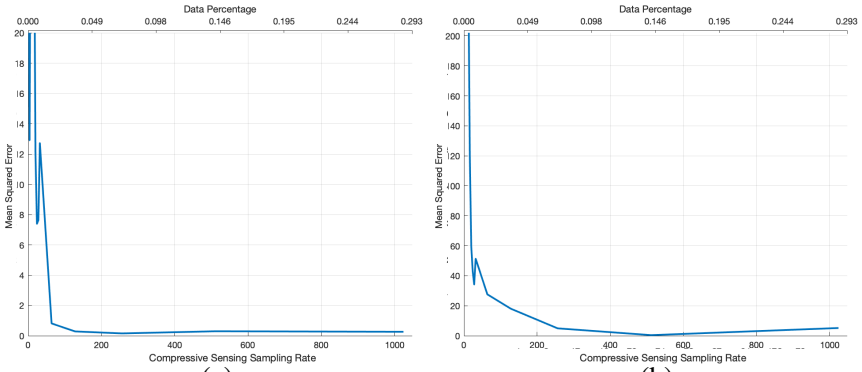


Figure 9. MSE between the similarity matrices at varying CSSR using (a) ED and (b) SAM

To further illustrate the preservation of the distance measures of ED and SAM from Figures 7 and 8, the absolute error was aggregated using the sum in (12) and is presented in Figure 9 as the CSSR increases.

$$\epsilon_{MSE} = \sum_{i=1}^{L}\sum_{j=1}^{L}\left|dist\left(\mathbf{b}_i, \mathbf{b}_j\right) - dist\left(\mathbf{\Phi}_{k_N}\mathbf{b}_i, \mathbf{\Phi}_{k_N}\mathbf{b}_j\right),\right| \tag{12}$$

5.3 Hyperspectral Band Clustering Accuracy

The SBS was computed using the ODS and the data in the CSSD and the coincident band rate (CBR) was computed as shown in equation (13). This error metric is an unnormalized Tanimoto index and used to estimate the accuracy of the SBS when using CS and without [43].

$$\boldsymbol{\epsilon_{CBR}} = \left|\left\{\widetilde{\mathbf{B}}_j\right\}_{j=1}^{p} \cap \left\{\widetilde{\mathbf{B}}_{\mathbf{\Phi}j}\right\}_{j=1}^{p}\right| \tag{13}$$

Table 1 shows the SBS using the Euclidean distance between clusters in (10). When the CSSR is inadequate, the CBR is lower and the SBS is erroneous. As the CSSR is increased, the SBS converges to the SBS computed using the full bands. Table 2 details the SBS using the SAM as a similarity measure between clusters in (10). The same trend appears showing the converges of the SBS in the CSSD to that of the ODS. However, using the angle, the convergence is slower requiring significantly more compressive samples. This is due to the wider error bound of the RCP as compared to the RIP in (6).

Table 1. Selected band subsets results using ED at varying CSSR

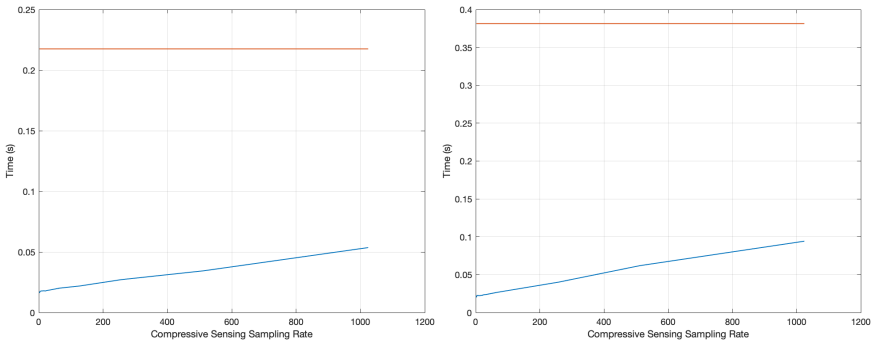
Algorithm	CSSR	CBR	Selected Bands
Full Band	N/A	N/A	8 29 54 64 66 75 79 113 146
$dist_{Euclidean}$	8 / 4096	5 / 9	27 54 61 62 75 79 93 113 146
$dist_{Euclidean}$	16 / 4096	9 / 9	8 29 54 64 66 75 79 113 146
$dist_{Euclidean}$	32 / 4096	7 / 9	8 29 54 64 66 75 79 114 149
$dist_{Euclidean}$	64 / 4096	9 / 9	8 29 54 64 66 75 79 113 146
$dist_{Euclidean}$	128 / 4096	9 / 9	8 29 54 64 66 75 79 113 146

Table 2. Selected band subsets results using SAM at varying CSSR

Algorithm	CSSR	CBR	Selected Bands
Full Band	N/A	N/A	10 29 45 54 58 88 100 145 161
$dist_{SAM}$	32 / 4096	2 / 9	10 35 46 80 88 115 124 137 159
$dist_{SAM}$	128 / 4096	5 / 9	9 19 29 45 58 89 100 139 161
$dist_{SAM}$	256 / 4096	6 / 9	10 29 45 54 58 88 115 137 159
$dist_{SAM}$	512 / 4096	6 / 9	10 29 45 54 58 88 115 137 159
$dist_{SAM}$	1024 / 4096	8 / 9	10 29 45 54 58 88 100 139 161

5.4 Similarity Matrix Compute Time

The advantage of utilizing the compressively sensed bands directly is the dimensionality reduction resulting in faster downstream processing using the reduced data size. Figure 10 shows the linear runtime trend as the CSSR increases versus the runtime required using the ODS plotted as the red curve. The data reduction power of CS is exponential in the number of samples and the runtime is linearly increasing implying considerable data reduction and runtime performance.



(a) (b)

Figure 10. Runtime computing the full inter-band similarity matrices with increasing CSSR. The red line details the runtime using the ODS.

CONCLUSIONS

This study reviewed the compressive sensing (CS) principles relevant to sensing hyperspectral bands. Exploiting the length and angle preservation properties of CS, the inter-band Euclidean distances (ED) and spectral angle mapper (SAM) are preserved due to the restricted isometric property (RIP) and restricted conformal property (RCP). Therefore, the similarity matrices between bands can be estimated using data in the compressive sensing sampling domain (CSSD). Using the reduced data sets, it is shown that the accuracy of the similarity measures is preserved as well as improving the runtime. Finally, a novel method of estimating the compressive sensing sampling rate (CSSR) using sequential compressive sensing (SCS) and a recursive kurtosis formulation is detailed. This method improves upon the current SCS research by replacing the sufficiency criteria of reconstruction with an light-weight recursive kurtosis (RK)formulation. Future research will focus on developing the RK-CSSR method by determining the convergence criteria programmatically.

REFERENCES

[1] C.-I. Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ: Wiley-Interscience, 2013.

[2] C.-I. Chang, *Real-Time Progressive Hyperspectral Image Processing*. New York, NY: Springer Berlin Heidelberg, 2016.

[3] C.-I. Chang, *Real-Time Recursive Hyperspectral Sample and Band Processing*. New York, NY: Springer Science+Business Media, 2017.

[4] C. E. Shannon, "A Mathematical Theory of Communication*," *Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 1948.

[5] Z. Wang, Y. Feng, and Y. Jia, "Spatial-spectral compressive sensing of hyperspectral image," *2013 IEEE 3rd Int. Conf. Inf. Sci. Technol. ICIST 2013*, pp. 1254–1259, 2013.

[6] H. Abdi and L. J. Williams, "Principal Component Analysis," *WIREs Comput. Stat.*, vol. 2, no. 4, pp. 433–459, Jul. 2010.

[7] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, 1988.

[8] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Netw.*, vol. 13, no. 4–5, pp. 411–430, May 2000.

[9] S. De Backer, P. Kempeneers, W. Debruyn, and P. Scheunders, "Band Selection for Hyperspectral Remote Sensing," *Pattern Recognit.*, vol. 2, no. 3, pp. 319–323, 2005.

[10] C. Yu, M. Song, and C.-I. Chang, "Band subset selection for hyperspectral image classification," *Remote Sens.*, vol. 10, no. 1, pp. 1–25, 2018.

[11] L. Wang and C.-I. Chang, "Multiple band selection for anomaly detection in hyperspectral imagery," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, no. 1, pp. 7022–7025.

[12] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. 2013.

[13] D. L. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[14] E. Candes and J. Romberg, "Sparsity and Incoherence in Compressive Sampling," *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, 2007.

[15] D. Sarvaiya and J. Amin, "Compression of Hyperspectral Image using Compressive Sensing," *Int. J. Adv. Res. Innov. Ideas Educ.*, vol. 2, no. 3, pp. 1904–1911, 2016.

[16] Y. Chen, "Effects of linear projections on the performance of target detection and classification in hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 5, no. 1, p. 053563, 2011.

[17] M. A. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Detection and Estimation with Compressive Measurements," *Dept. ECE, Rice Univ. Tech. Rep*, pp. 1–16, 2006.

[18] N. Keshava and P. W. Boettcher, "Relationships between physical phenomena, distance metrics, and best-bands selection in hyperspectral processing," in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, 2003, vol. 4381, no. August 2001, pp. 55–67.

[19] C.-I. Chang, "A Review of Virtual Dimensionality for Hyperspectral Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 4, pp. 1285–1305, 2018.

[20] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, 2004.

[21] C.-I. Chang, L. C. Lee, B. Xue, M. Song, and J. Chen, "Channel Capacity Approach to Hyperspectral Band Subset Selection," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 10, pp. 4630–4644, 2017.

[22] L. G. Chang, C.I.; Du, Q.; Sun, T.; Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, 1999.

[23] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: a comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, 2005.

[24] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based multispectral band selection using mutual information," *Proc. - Int. Conf. Pattern Recognit.*, vol. 2, pp. 760–763, 2006.

[25] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, 2016.

[26] Y. Xu, Q. Du, and N. Younan, "Particle swarm optimization-based band selection for hyperspectral target

detection," *2016 IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, no. 4, pp. 5872–5875, 2016.
[27] Yuan Yuan, Guokang Zhu, and Qi Wang, "Hyperspectral Band Selection by Multitask Sparsity Pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, 2014.
[28] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
[29] E.J. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
[30] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.
[31] M. E. Lopes, "Estimating Unknown Sparsity in Compressed Sensing," *UC Berkeley Dep. Stat.*, 2014.
[32] M. E. Lopes, "Unknown sparsity in compressed sensing: Denoising and inference," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5145–5166, 2016.
[33] J. Haupt and R. Nowak, "A generalized restricted isometry property," *Univ. Wisconsin-Madison, Tech. Rep*, pp. 1–16, 2007.
[34] T. Cheng, "Restricted conformal property of compressive sensing," *2014 11th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2014*, pp. 152–161, 2014.
[35] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 2003.
[36] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via OMP," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
[37] D. M. Malioutov, S. R. Sanghavi, and A. S. Willsky, "Sequential compressed sensing," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 435–444, 2010.
[38] D. M. Malioutov, S. Sanghavi, and a S. Willsky, "Compressed Sensing with Sequential Observations," *Bernoulli*, pp. 3357–3360, 2008.
[39] R. Ward, "Compressed sensing with cross validation," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, 2009.
[40] M. Gray, Robert and D. Davisson, Lee, *An Introduction to Statistical Signal Processing*. Cambridge University Press, 2004.
[41] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[42] P. A. Mitchell, "Hyperspectral Digital Imagery Collection Experiment (HYDICE)," in *Geographic Information Systems, Photogrammetry, and Geological/Geophysical Remote Sensing*, 2004, vol. 2587, pp. 70–95.
[43] T. T. Tanimoto, *An elementary mathematical theory of classification and prediction by T.T. Tanimoto*. International Business Machines Corporation New York, 1958.