UNSUPERVISED SPATIAL-SPECTRAL HYPERSPECTRAL IMAGE RECONSTRUCTION AND CLUSTERING WITH DIFFUSION GEOMETRY

Kangning Cui [1,2], *Ruoning Li* [1], *Sam L. Polk* [3],
James M. Murphy [3], *Robert J. Plemmons* [4], *Raymond H. Chan* [1,2*]

[1] Department of Mathematics, City University of Hong Kong
[2] Hong Kong Centre for Cerebro-Cardiovascular Health Engineering
[3] Department of Mathematics, Tufts University
[4] Departments of Mathematics and Computer Science, Wake Forest University

ABSTRACT

Hyperspectral images (HSIs), which store a hundred or more bands of spectral bands of reflectance, have become an important data source in natural and social sciences. HSIs are often generated in large quantities at a relatively coarse spatial resolution. As such, unsupervised machine learning algorithms incorporating known structure of HSIs are needed to analyze these images automatically. This work introduces the Spatial-Spectral Image Reconstruction and Clustering with Diffusion Geometry (DSIRC) algorithm for partitioning highly mixed hyperspectral images. DSIRC reduces measurement noise through a shape-adaptive reconstruction procedure. In particular, for each HSI pixel, DSIRC locates spectrally-correlated pixels within a data-adaptive spatial neighborhood and reconstructs that pixel's spectral signature using those of its neighbors. DSIRC then locates high-density, high-purity pixels that are far in diffusion distance (a data-dependent distance metric) from other high-density, high-purity pixels and treats these as cluster exemplars. Non-modal points are assigned according to the label of their diffusion distance-nearest neighbor of higher density and purity that is already labeled. Strong numerical results indicate that incorporating spatial information through image reconstruction substantially improves the performance of pixel-wise clustering.

Index Terms— Clustering, Diffusion Geometry, Hyperspectral Imagery, Image Reconstruction, Spectral Unmixing, Unsupervised Learning.

1. INTRODUCTION

Hyperspectral images (HSIs)—often remotely sensed by airborne or orbital spectrometers—are high-dimensional images that encode rich spectral and spatial structure [1] that has enabled the detection of material structure in a scene using machine learning algorithms [2, 3]. However, due to the large volume of HSI data continuously generated by remote sensors, expert annotations (often required for supervised algorithms) are usually difficult to obtain. Moreover, there is an inherent trade-off between the spatial and spectral resolution of HSI data [1, 3]. HSIs are often created at a coarse spatial resolution due to this trade-off, meaning that some pixels in an HSI correspond to spatial regions in the scene containing many different materials [1, 3]. Thus, it is crucial to develop unsupervised approaches that capture the underlying geometric structure of an HSI while incorporating spectral mixing.

This work introduces the Spatial-Spectral Image Reconstruction and Clustering with Diffusion Geometry (DSIRC) algorithm for unsupervised material discrimination in hyperspectral images. DSIRC is a variant of the unsupervised Diffusion and Volume maximization-based Image Clustering (D-VIC) algorithm [3]. DSIRC improves D-VIC by incorporating spatial information through a shape-adaptive reconstruction (SaR) before cluster analysis. Since HSI pixels that are spatially close tend to come from the same cluster, DSIRC is shown to substantially outperform D-VIC (its spatially-agnostic counterpart) in extensive numerical results on real-world HSI data.

This article is organized as follows. Section 2 contains background on HSI analysis (e.g., clustering, reconstruction, and spectral unmixing), diffusion geometry, and D-VIC. Section 3 introduces DSIRC. Section 4 contains numerical comparisons of DSIRC against classical and state-of-the-art algorithms. Section 5 concludes and discusses future work.

2. RELATED WORKS

2.1. Hyperspectral Image Clustering

Algorithms for HSI clustering segment HSI pixels $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^B$ (interpreted as a point cloud of N pixels'

*Corresponding Author; email: raymond.chan@cityu.edu.hk
This work was supported in part by HKRGC Grants No. CUHK14301718, CityU11301120, C1013-21GF, CityU Grant 9380101.

spectral signatures, where B is the number of spectral bands) into *clusters* of pixels $\{X_k\}_{k=1}^K$ [4]. Ideally, any two pixels from the same cluster should share key commonalities (e.g., common materials [3]). HSI clustering algorithms are *unsupervised*; i.e., the partition $\{X_k\}_{k=1}^K$ is recovered without the aid of ground truth (GT) labels [3].

2.2. Hyperspectral Image Reconstruction

Spatially close HSI pixels tend to come from the same cluster, but intra-cluster spectral reflectances may vary substantially due to the inherently coarse spatial resolution of HSIs. This section reviews *HSI reconstruction*, which efficiently denoises hyperspectral data by reconstructing HSI pixels using the spectra of spatial nearest neighbors. HSI reconstruction has been successfully used as a preprocessing step for semi-supervised learning [5, 6, 7] and is expected to be useful for unsupervised learning [3].

HSI reconstruction denoises the spectra of pixels through a linear combination of spectral signatures of its spatial nearest neighbors confined by local spatial windows. While simple spatial squares have been used to efficiently incorporate spatial information into unsupervised and semi-supervised algorithms, these generally require the tuning of the spatial radius of the window used, yielding an additional parameter [5, 6, 7, 8, 9, 10]. Shape-adaptive (SA) regions, in contrast, may be used for parameter-free HSI reconstruction [5].

Here we introduce the Shape-adaptive Reconstruction (SaR) algorithm. First, SA region of each pixel is found by the local polynomial approximation (LPA) filtering and intersection of confidence intervals (ICI) rule, implemented on the first principal component (PC) $\mathbf{Z} \in \mathbb{R}^2$ of the HSI [11, 12]. The LPA filtering designs linear filters by fitting polynomials in defined sliding windows to denoise images so that the spatial information within the neighborhood of pixels is well explored. The ICI rule constructs the adaptive window by analyzing the estimation error between the value of ideal image and denoised image using confidence intervals. The LPA filtering and ICI rule are often composed to select SA regions for HSI pixels [12]. We follows a reconstruction step that denoise the spectral of HSI pixels in SaR [5].

Denote the spatial coordinate as $\xi_x = (\xi_1, \xi_2)^T$ for $x \in X$, and assume the first PC $\mathbf{Z}(\xi_x) = \mathbf{I}(\xi_x) + \varepsilon_x$, where $\mathbf{I}(\xi_x)$ and ε_x are the underlying signal and noise of pixel x in its first PC. LPA estimates true signal of each length candidate $l_j \in L_{sar} = \{l_1 \leq \cdots \leq l_j \leq \cdots \leq l_J\}$ and direction θ_m with $m = 1, 2, \ldots, 8$ by $\hat{\mathbf{I}}_{l_j,\theta_m}(\xi_x) = \sum_s g_{l_j,\theta_m}(u_s)\mathbf{Z}(\mathbf{R}_{\theta_m}\xi_s)$, where $g_{l_j,\theta_m}(u_s)$ is the directional LPA kernel for each θ_m and l_j, and $u_s = \mathbf{R}_{\theta_m}(\xi_x - \xi_s)$ is rotated coordinate differences of ξ_x and ξ_s and $\mathbf{R}_{\theta_m} = \begin{bmatrix} cos(\theta_m) & sin(\theta_m) \\ -sin(\theta_m) & cos(\theta_m) \end{bmatrix}$.

Then, ICI is applied to find the optimal length $l^+_{\theta_m} \in L_{sar}$ for each direction θ_m. The confidence interval for $\hat{\mathbf{I}}_{l,\theta_m}(\xi_x)$: $CI(\hat{\mathbf{I}}_{l,\theta_m}(\xi_x)) = [\alpha_{l_j,\theta_m}(\xi_x), \beta_{l_j,\theta_m}(\xi_x)]$ with

$$\begin{cases} \alpha_{l_j,\theta_m} = \hat{\mathbf{I}}_{l_j,\theta_m} - \tau\sigma(\sum_s g^2_{l_j,\theta_m}(u_s))), \ \tilde{\alpha}_{l_j,\theta_m} = \max\limits_{1,\ldots,j} \alpha_{l_j,\theta_m}, \\ \beta_{l_j,\theta_m} = \hat{\mathbf{I}}_{l_j,\theta_m} + \tau\sigma(\sum_s g^2_{l_j,\theta_m}(u_s))), \ \tilde{\beta}_{l_j,\theta_m} = \min\limits_{1,\ldots,j} \beta_{l_j,\theta_m}, \end{cases}$$

where τ is a threshold. The optimal length $l^+_{\theta_m}$ is determined by $\arg\max_{l_j} \tilde{\alpha}_{l_j,\theta_m} \leq \tilde{\beta}_{l_j,\theta_m}$. Finally, a SA region for the target pixel will be constructed after attaining $l^+_{\theta_m}$ for each m.

Let $Z(x)$ be the indices of the pixels contained in the calculated SA region for $x \in X$. Then, the reconstructed spectrum $\tilde{x} = \sum_{y \in Z(x)} p(x,y)y / \sum_{y \in Z(x)} p(x,y)$, where $p(x,y)$ is the Pearson correlation coefficient between x and and y.

Algorithm 1: Shape-adaptive Reconstruction

Input: X (dataset), L_{sar} (length candidates)
Output: $\tilde{X}$ (reconstructed dataset)

1. Apply Principal Component Analysis (PCA) on X to get the first PC $\mathbf{Z} \in \mathbb{R}^{H \times W}$;
2. For each $l_j \in L_{sar}$ and θ_m, LPA estimates the true signal $\hat{\mathbf{I}}_{l_j,\theta_m}(\xi_x) = \sum_{s=1}^n g_{l_j,\theta_m}(u_s)\mathbf{Z}(\mathbf{R}_{\theta_m}\xi_s)$ on the first PC $\mathbf{Z}$ of pixel x, where s is the index of pixels in x's neighborhood [11];
3. For each x and θ_m, computes the optimal length by ICI: $l^+_{\theta_m} = \arg\max_{l_j} \tilde{\alpha}_{l_j,\theta_m} \leq \tilde{\beta}_{l_j,\theta_m}$;
4. Construct the SA region of x by these optimal lengths in all directions, and denote the indices of pixels in its SA region as $Z(x)$;
5. Compute the Pearson Correlation coefficients $p(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\cdot\mathrm{Var}(y)}}$ between x and $y \in Z(x)$;
6. Reconstruct the spectrum of each pixel $x \in X$ by: $\tilde{x} = \frac{\sum_{y \in Z(x)} p(x,y)y}{\sum_{y \in Z(x)} p(x,y)}$, and compose the set $\tilde{X}$;

2.3. Blind Spectral Unmixing

Due to an inherent tension between spatial and spectral resolution, HSIs are often high in spatial resolution [1]. As such, a single pixel may correspond to a spatial region containing multiple materials [3, 13]. To account for mixing in HSIs, linear spectral unmixing algorithms decompose pixel spectra into a linear combination of endmembers: the spectral signatures intrinsic to the materials in the scene. Mathematically, if p is the number of materials in the scene, a blind spectral unmixing algorithm locates a matrix $\mathbf{E} \in \mathbb{R}^{p \times B}$ (with rows containing endmembers) and abundance vectors $a_i \in \mathbb{R}^p$ such that $x_i \approx a_i^\top \mathbf{E}$ [1, 14]. The *purity* of a pixel x_i—defined by $\eta(x_i) = \max_{1 \leq j \leq p} (a_i)_j$—quantifies the level of mixture in the pixel x_i. Indeed, $\eta(x)$ will be large only if it corresponds to a spatial region containing predominantly just one material [2, 3, 14].

2.4. Diffusion Geometry

Graph-based clustering methods efficiently extract latent nonlinear structure in HSIs by interpreting pixels as nodes in an undirected, weighted graph [15]. Edges between nodes are encoded in an adjacency matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$; $\mathbf{W}_{ij} = 1$ if x_j is one of the k_n ℓ^2-nearest neighbors of x_i in X, and $\mathbf{W}_{ij} = 0$ otherwise. Let $\mathbf{D}$ be the $N \times N$ diagonal degree matrix with $\mathbf{D}_{ii} = \sum_{j=1}^{N} \mathbf{W}_{ij}$. Then, $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ may be interpreted as the transition matrix for a Markov diffusion process on HSI pixels. If the graph underlying $\mathbf{P}$ is irreducible and aperiodic, then $\mathbf{P}$ has a unique stationary distribution $\pi \in \mathbb{R}^{1\times N}$ satisfying $\pi\mathbf{P} = \pi$.

Diffusion distances enable comparisons between data points in the context of the diffusion process encoded in $\mathbf{P}$. Define the diffusion distance at time $t \geq 0$ between $x_i, x_j \in X$ by $D_t(x_i, x_j) = \sqrt{\sum_{k=1}^{N}[(\mathbf{P}^t)_{ik} - (\mathbf{P}^t)_{jk}]^2/\pi_k}$ [15]. The diffusion time parameter t controls the scale of structure considered by diffusion distances; smaller t corresponds to the recovery of local structure and larger t corresponds to the recovery of global structure [10, 16]. Diffusion distances may be efficiently computed using the eigendecomposition of $\mathbf{P}$. Indeed, if $\{(\lambda_k, \psi_k)\}_{k=1}^{N}$ are the (right) eigenvalue-eigenvector pairs of the transition matrix $\mathbf{P}$, then $D_t(x_i, x_j) = \sqrt{\sum_{k=1}^{N} |\lambda_k|^{2t}[(\psi_k)_i - (\psi_k)_j]^2}$ for any $t \geq 0$ and $x_i, x_j \in X$. Importantly, for t sufficiently large, diffusion distances therefore can be accurately approximated by using just the eigenvectors ψ_k with sufficiently large $|\lambda_k|$.

2.5. Diffusion and Volume Maximization-based Image Clustering

The unsupervised D-VIC algorithm incorporates spectral mixture into a diffusion-based clustering framework [3]. In its spectral unmixing step, D-VIC estimates the endmember set $E \in \mathbb{R}^{p\times B}$ using alternating volume maximization [17], where p is calculated using HySime. Pixel purity is then estimated from abundances, calculated using a nonnegative least-squares solver [18]. Empirical density of pixels is estimated using $f(x) = \sum_{y\in NN_{k_n}(x)} \exp(-\|x - y\|_2^2/\sigma_0^2)$, where $NN_{k_n}(x)$ is the set of k_n nearest neighbors of x in X and $\sigma_0 > 0$ is the scaling factor controlling the interaction radius between pixels in density calculations. Denoting $\zeta(x)$ as the harmonic mean of normalized purity $\hat{\eta}(x) = \frac{\eta(x)}{\max_{1\leq i\leq N} \eta(x_i)}$ and density $\hat{f}(x) = \frac{f(x)}{\max_{1\leq i\leq N} f(x_i)}$, the following function is constructed to aid the estimation of modal pixels:

$$d_t(x) = \begin{cases} \max\limits_{y\in X} D_t(x, y) & x = \arg\min\limits_{y\in X} \zeta(y), \\ \min\limits_{y\in X}\{D_t(x, y)|\zeta(y) \geq \zeta(x)\} & \text{otherwise.} \end{cases}$$

Hence, the K maximizers of $\mathcal{D}_t(x) = \zeta(x)d_t(x)$ are pixels high in density and purity, but far in diffusion distance from other pixels high in density and purity. These pixels are selected as class modes and given unique labels. Non-modal pixels are (in order of non-increasing $\zeta(x)$) assigned the label of their labeled D_t-nearest neighbor of higher ζ-value that is already labeled. D-VIC is equivalent to the Learning by Unsupervised Nonlinear Diffusion (LUND) algorithm if pixel purity is set to be constant across all pixels in X [3, 19].

Algorithm 2: Spatial-Spectral Image Reconstruction and Clustering with Diffusion Geometry

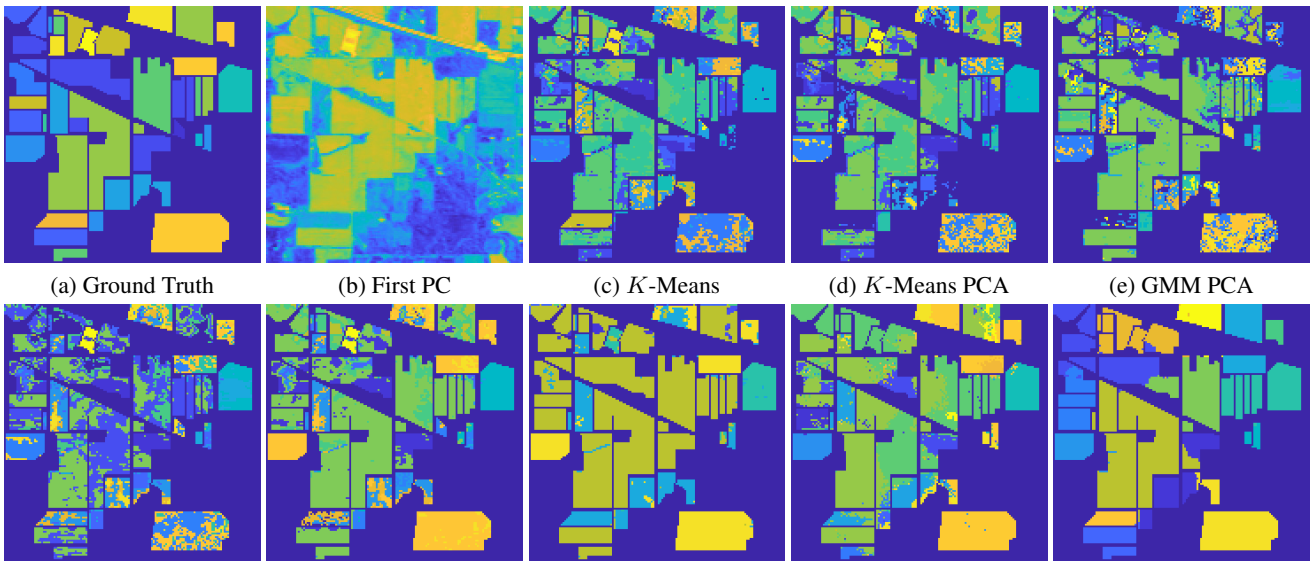
Input: X (dataset), k_n (# nearest neighbors), σ_0 (KDE bandwidth), t (diffusion time parameter), K (# classes), L_{sar}
Output: $\hat{\mathcal{C}}$ (HSI clustering)

1 For each $x \in X$, compute $\eta(x)$ and $f(x)$;
2 For each $x \in X$, compute $\zeta(x) = \frac{2\hat{f}(x)\hat{\eta}(x)}{\hat{f}(x)+\hat{\eta}(x)}$;
3 Compute the reconstructed data $\tilde{X} = \mathrm{SaR}(X)$;
4 Let $\{\tilde{x}_{m_k}\}_{k=1}^{N}$ be a sorting of $\tilde{X}$ by $\mathcal{D}_t(\tilde{x}) = \zeta(x)d_t(\tilde{x})$ in non-increasing order. Label $\hat{\mathcal{C}}(\tilde{x}_{m_k}) = k$ for $1 \leq k \leq K$;
5 In order of non-increasing ζ-value, assign non-modal pixels $\tilde{x} \in \tilde{X}$ the label $\hat{\mathcal{C}}(\tilde{x}) = \hat{\mathcal{C}}(\tilde{x}^*)$, where $\tilde{x}^* = \arg\min\limits_{y\in X}\{D_t(\tilde{x}, y)|\zeta(y) \geq \zeta(\tilde{x}) \wedge \hat{\mathcal{C}}(y) > 0\}$;

3. SPATIAL-SPECTRAL IMAGE RECONSTRUCTION AND CLUSTERING WITH DIFFUSION GEOMETRY

HSI pixels generally satisfy spatial regularity; i.e., pixels that are spatially close are more likely to contain the same materials. This can be observed from real HSI datasets, and has been applied in many studies to improve the pixel-wise clustering and classification algorithms [5, 7, 8, 9, 10].

This section introduces the DSIRC clustering algorithm, which explicitly incorporates HSI reconstruction into the D-VIC algorithm. In its first stage, DSIRC computes $\zeta(x)$ using the original pixel spectra as described in Section 2.5 [3]. Then, the SaR algorithm (Section 2.2) is used to find the SA region of each pixel $x \in X$, and produce the smoothed spectrum $\tilde{x}$ by weighting the spectra of pixels in x's SA region with normalized Pearson correlation coefficients as in Algorithm 1. The K maximizers of $\mathcal{D}_t(\tilde{x}) = \zeta(x)d_t(\tilde{x})$ are considered cluster modes and assigned unique labels. Remaining pixels are labeled in order of non-increasing ζ-value according to their D_t-nearest neighbor of higher ζ-value that is already labeled. Notably, the main difference between DSIRC and D-VIC is that DSIRC incorporates spatial information by reconstructing the spectra of pixels as an intermediate step before labeling.



(a) Ground Truth (b) First PC (c) K-Means (d) K-Means PCA (e) GMM PCA

(f) SC (g) D-VIC (h) SC-I (i) DLSS (j) DSIRC

Fig. 1: Comparison of clusterings results of classical algorithms (Panels (c)-(f)), state-of-the-art algorithms (Panels (g)-(i)) and DSIRC (Panel (j)) on the Indian Pines HSI (Panel (a)-(b)). Outliers are incorrectly clustered by pixel-wise algorithms based on Euclidean distances, leading to noisy outputs, see Panels (c)-(f). By introducing the diffusion geometry and labeling pixels with non-increasing order of $\zeta(x)$ (D-VIC), the mistakes in predictions are diminished, see Panels (g). Enforcing spatial regularity further removed isolated falsely labeled pixels, see Panels (h)-(j).

4. EXPERIMENTAL RESULTS

This section contains comparisons of DSIRC against related HSI clustering algorithms on the Indian Pines HSI. Indian Pines—collected by the NASA AVIRIS sensor in northwest Indiana, USA—encodes $B = 200$ bands of reflectance across 145×145 pixels. The Indian Pines scene consists of $K = 16$ GT classes, which are visualized in Fig.1(a). Fig.1(b) visualizes the first PC of Indian Pines. Clusterings were evaluated using overall accuracy (OA)—the fraction of correctly labeled pixels—and Cohen's κ-coefficient $\kappa = \frac{\text{OA}-p_e}{1-p_e}$, ($p_e$ is the probability of random agreement).

The classical algorithms we compared against include K-Means and the Gaussian Mixture Model (GMM) [4]. K-Means locates the clustering that minimizes within-cluster Euclidean distances to cluster centroids. GMM fits a mixture of Gaussian distributions to the dataset using the expectation-maximization algorithm. As HSIs have hundreds of spectral bands, PCA dimensionality reduction is often implemented before cluster analysis using K-Means and GMM [4].

Two spatially-agnostic graph-based clustering algorithms, namely spectral clustering (SC) [20] and D-VIC (see Section 2.5) are compared to show the clustering performance when spatial dependency of HSIs is disregarded. SC implements K-Means on the nonlinearly-mapped dataset $x_i \mapsto [(\psi_1)_i\ (\psi_2)_i \dots (\psi_K)_i]$, usually after a normalization step [20].

We also compared DSIRC against several graph-based clustering algorithms that use spatial information. First, we compare against improved spectral clustering (SC-I), which modifies the graph underlying $\mathbf{P}$ in SC to incorporate spatial information into edge weights [21]. We analyze spectral-spatial diffusion learning (DLSS) algorithm as well, which incorporates spatial information into LUND by restricting edges between pixels to spatial nearest neighbors in a $(2R+1) \times (2R+1)$ spatial square centered at those pixels, where $R = 1, 2, \dots, 9$ is a spatial window input parameter [8]. We optimized for OA across the same hyperparameter grid for all graph-based algorithms. The set of length candidates of DSIRC is set as $L_{sar} = 1, 2, 3, 5, 7, 9$.

	OA	$\boldsymbol{\kappa}$		**OA**	$\boldsymbol{\kappa}$
GMM PCA	0.3581	0.2821	**SC-I**	0.4696	0.3493
SC	0.3784	0.3029	**D-VIC**	0.4756	0.3848
***K*-Means**	0.3817	0.3080	**DLSS**	0.4886	0.4074
***K*-Means PCA**	0.3837	0.3085	**DSIRC**	**0.6195**	**0.6123**

Table 1: Performances of DSIRC and comparison methods on the Indian Pines dataset. The bold values indicate the highest performance. The performance of K-Means, K-Means PCA, GMM PCA, SC, D-VIC, and DSIRC were averaged across ten trials.

Table 1 compares the performance of DSIRC against the methods described above and Figure 1 visualizes the Indian Pines dataset and optimal clusterings. DSIRC outperforms DLSS (the closest competitor) by 0.13 in OA, and 0.21 in κ. The differences between these two algorithms are that

DSRIC incorporates pixel purity in mode selection and utlizes the spatial regularity of HSI before the unsupervised diffusion-based labeling process. Futhermore, DSIRC relies on a spatially-adaptive window with automatically determined shape, whereas DLSS requires the user to input the spatial window size R [8]. Image reconstruction in DSIRC appears to efficiently remove "spatial noise" observed in the D-VIC clustering, as is visualized in Figure 1. Thus, enforcing spatial regularity appears to improve the quality of a diffusion-based clustering quite substantially.

5. CONCLUSIONS AND FUTURE WORKS

We conclude that incorporating spatial information through image reconstruction appears to substantially improve the performance of pixel-wise HSI clustering algorithms. Thus, incorporating a shape-adaptive reconstruction akin to that which was used in DSIRC may be useful before the labeling of HSI pixels. Future work includes extending DSIRC to the active learning domain, wherein the labels of a few carefully-selected points are queried and propagated to the rest of the image [2, 8]. We also expect that DSIRC may be extended to the unsupervised multiscale clustering setting [10, 16]. The resulting unsupervised and active learning algorithms are likely to be successful in a number of applications; e.g., identifying changes of mining ponds in multispectral images over time, possibly reflecting the occurrence of artisanal and small-scale gold mining activities [22].

6. REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci Remote Sens Mag*, vol. 1, no. 2, pp. 6–36, 2013.

[2] S. L. Polk, K. Cui, R. J. Plemmons, and Murphy J. M., "Active diffusion and VCA-assisted image segmentation of hyperspectral images," *arXiv preprint arXiv:2204.06298*, 2022.

[3] S. L. Polk, K. Cui, R. J. Plemmons, and J. M. Murphy, "Diffusion and volume maximization-based clustering of highly mixed hyperspectral images," *arXiv preprint arXiv:2203.09992*, 2022.

[4] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, Springer, 2009.

[5] R Li, K Cui, RH Chan, and RJ Plemmons, "Classification of hyperspectral images using SVM with shape-adaptive reconstruction and smoothed total variation," *arXiv preprint arXiv:2203.15619*, 2022.

[6] J. Ren, R. Wang, G. Liu, Y. Wang, and W. Wu, "An SVM-based nested sliding window approach for spectral–spatial classification of hyperspectral images," *Remote Sens*, vol. 13, no. 1, pp. 114, 2021.

[7] R. H. Chan and R. Li, "A 3-stage spectral-spatial method for hyperspectral images classification," *arXiv preprint arXiv:2204.09294*, 2022.

[8] J. M. Murphy and M. Maggioni, "Unsupervised clustering and active learning of hyperspectral images with nonlinear diffusion," *IEEE Trans Geosci Remote Sens*, vol. 57, no. 3, pp. 1829–1845, 2018.

[9] J. M. Murphy and M. Maggioni, "Spectral–spatial diffusion geometry for hyperspectral image clustering," *IEEE Geosci Remote Sens Lett*, vol. 17, no. 7, pp. 1243–1247, 2019.

[10] S. L. Polk and J. M. Murphy, "Multiscale clustering of hyperspectral images through spectral-spatial diffusion geometry," in *Proc IEEE Int Geosci Remote Sens Symp*, 2021, pp. 4688–4691.

[11] V. Katkovnik, K. Egiazarian, and J. Astola, *Local approximation techniques in signal and image processing*, SPIE Press, 2006.

[12] W. Fu, S. Li, L. Fang, X. Kang, and J. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 9, no. 2, pp. 556–567, 2015.

[13] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans Geosci Remote Sens*, vol. 46, no. 8, pp. 2435–2445, 2008.

[14] K. Cui and R. J. Plemmons, "Unsupervised classification of AVIRIS-NG hyperspectral images," in *Proc WHISPERS*. IEEE, 2021, pp. 1–5.

[15] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl Comput Harm Anal*, vol. 21, no. 1, pp. 5–30, 2006.

[16] J. M. Murphy and S. L. Polk, "A multiscale environment for learning by diffusion," *Appl Comput Harmon Anal*, vol. 57, pp. 58–100, 2022.

[17] T. Chan, W. Ma, A. Ambikapathi, and C. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans Geosci Remote Sens*, vol. 49, no. 11, pp. 4177–4193, 2011.

[18] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *J Chemom*, vol. 11, no. 5, pp. 393–401, 1997.

[19] M. Maggioni and J. M. Murphy, "Learning by unsupervised nonlinear diffusion.," *J Mach Learn Res*, vol. 20, no. 160, pp. 1–56, 2019.

[20] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Proc Adv Neural Inf Process Syst*, vol. 14, 2001.

[21] Y. Zhao, Y. Yuan, and Q. Wang, "Fast spectral clustering for unsupervised hyperspectral image classification," *Remote Sens*, vol. 11, no. 4, pp. 399, 2019.

[22] S. Camalan et al., "Change detection of amazonian alluvial gold mining using deep learning and sentinel-2 imagery," *Remote Sens.*, vol. 14, no. 7, 2022.