Stochastic Perturbations on Low-Rank Hyperspectral Data for Image Classification

Alex Sumarsono[a,1,*], Farnaz Ganjeizadeh[b,2], Ryan Tomasi[c,3]

a California State University East Bay, 25800 Carlos Bee Blvd., Hayward, CA 94542, USA
b California State University East Bay, 25800 Carlos Bee Blvd., Hayward, CA 94542, USA
c California State University East Bay, 25800 Carlos Bee Blvd., Hayward, CA 94542, USA
1 alex.sumarsono@csueastbay.edu; *2 farnaz.ganjeizadeh@csueastbay.edu; 3 rtomasi@horizon.csueastbay.edu*
* corresponding author

ARTICLE INFO	ABSTRACT
	Hyperspectral imagery (HSI) contains hundreds of narrow contiguous bands of spectral signals. These signals, which form spectral signatures, provide a wealth of information that can be used to characterize material substances. In recent years machine learning has been used extensively to classify HSI data. While many excellent HSI classifiers have been proposed and deployed, the focus has been more on the design of the algorithms. This paper presents a novel data preprocessing method (LRSP) to improve classification accuracy by applying stochastic perturbations to the low-rank constituent of the dataset. The proposed architecture is composed of a low-rank and sparse decomposition, a degradation function and a constraint least squares filter. Experimental results confirm that popular state-of-the-art HSI classifiers can produce better classification results if supplied by LRSP-altered datasets rather than the original HSI datasets.

I. Introduction

Hyperspectral images (HSI) of an object or a scene are typically acquired by electro-optical remote sensors. This is achievable because the materials in various objects in a scene inherently reflect, absorb and emit electromagnetic radiation. Modern hyperspectral sensors can capture a portion of the electromagnetic spectrum from the visible region (0.4 – 0.7 um) through the near-infrared region (approximately 2.4 um) in hundreds of narrow contiguous bands of about 10 nm wide calibrated to within 1 nm [1,2]. The radiation detected by these sensors over a sufficiently broad spectral band creates a distinct spectral signature that can be used to characterize and identify a certain kind of material. An example is shown in Figure 1. The image on the left is the image of an urban area in the Washington DC mall. It consists of three bands out of 210 bands collected by the sensor. The image on the right is the spectral plots of three data pixels as a function wavelength by spectral band number, each of which shows the corresponding material (vegetation, roof and water).
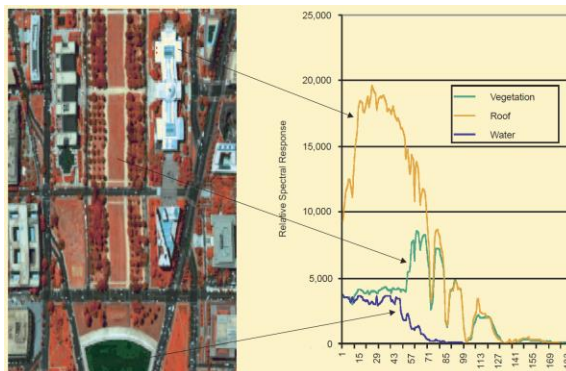

Fig. 1. Image of data pixels of three materials as a function of wavelength

The spatial and spectral information of an HSI dataset is typically organized as a data cube, similar to what is depicted in Figure 2. The face and the depth display the spatial coordinates and the spectral bands, respectively. Each band belongs to a narrowband image of the surface covered by the field of view of the sensor [3]. Therefore, every image pixel is a vector pixel where each element is a member of a spectral band. The vector provides a spectral signature characterizing the materials within the pixel.
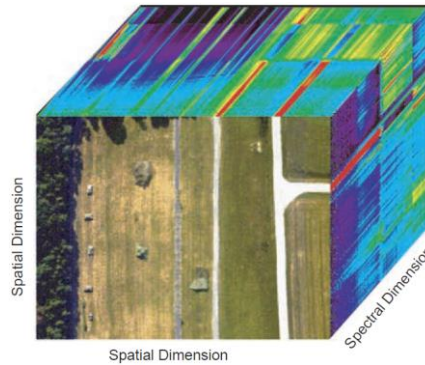

Fig. 2. Hyperspectrometry of 3D data cube

Applications of hyperspectral imagery span across many disciplines including but not limited to resource management, agriculture, mineral exploration and environmental monitoring [4-7]. In recent years machine learning has been used to classify and exploit HSI datasets. Many excellent HSI classifiers have been proposed and deployed, such as Class Feature Weighted Hyperspectral Image Classification (CFW-HSIC), Contextual Information Subspace Based Hyperspectral Image Classification (CIS-HSI), Semi-supervised HSI Classification using Self Training and Data Editing (RDE_self-training), Multiple Features and Nearest Regularized Subspace (MFNRS), Discriminative Marginalized Least Squares Regression (DMLSR) and Iterative Support Vector Machine (ISVM). CFW-HSIC [8] deals with class variability by computing the inter-class feature probabilities as class weights to generalize commonly used overall accuracy from the extracted features from classes of interest. CIS-HSI [9] uses decision fusion at a super-pixel level to generate the classification result. A super-pixel is formed by exploiting the contextual information among the spatial features in the data. RDE_self-training [10] is a semi-supervised learning algorithm with data editing. It revises the labels of mislabeled samples based on a nearest neighbor voting rule. The training set is then enlarged to include some unlabeled samples selected from a pool of candidates drawn from some ordered probability of the revised samples. MFNRS [11] proposes a residual fusion strategy with multiple features that include local binary patterns, Gabor features and the original spectral signatures to represent the test pixel from different perspectives to enhance the classifier's discriminative ability. DMLSR [12] enhances the discriminative power by taking the class separability and data reconstruction ability simultaneously. It employs Fisher criterion and imposes data reconstruction constraints to avoid overfitting and to improve classification performance. ISVM [13] is an iterative version of SVM applied to HSI classification. It extracts spatial information iteratively via feedback loops. A Gaussian filter is used to obtain the spatial information of the classification map to combine with the currently processed hyperspectral cube for the next iteration.

These approaches tend to focus on the design of the classifiers to improve the performance. While performance improvement is also our motivation, we are taking a different approach by appropriately applying preprocessed data rather than the original data to the algorithm. In this paper we propose a novel data preprocessing method to improve classification accuracy by perturbing the statistical nature of the HSI data. Perturbations are achieved by adding stochasticity to the low-rank component of the data, hence the name Low-Rank Stochastic Perturbation (LRSP). The new dataset will then be presented to an HSI classifier. We will show that off-the-shelf state-of-the art HSI classifiers can produce higher classification results by using LRSP-generated datasets rather than the original HSI datasets.

II. Proposed Method

Due to the nature of the dataset, many of the problems in HSI image classification can be attributed to the gradients that are trapped in local optima. The idea behind the LRSP method is to slightly perturb the convexity of the data by adding some stochasticity. It should be sufficiently meaningful that it can push the gradients out of a local minimum as conceptually depicted in Figure 3. At the same time, it should be relatively small so it does not change the underlying structure of the dataset.
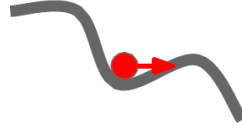

Fig. 3.Example of a figure caption.

LRSP consists of three parts: LRS decomposition, degradation function and CLS filtering. Figure 4 shows the design flow of the proposed method. The original HSI dataset is first decomposed into its low-rank and sparse components using the low-rank and sparse (LRS) decomposition method. The low rank data matrix is the *support* matrix that represents the union of multiple subspaces where each subspace is occupied by most of the spatial-spectral energy of the associated endmember. The sparse data matrix contains the *fine* approximation providing detailed spatial-spectral resolution to the endmember data.
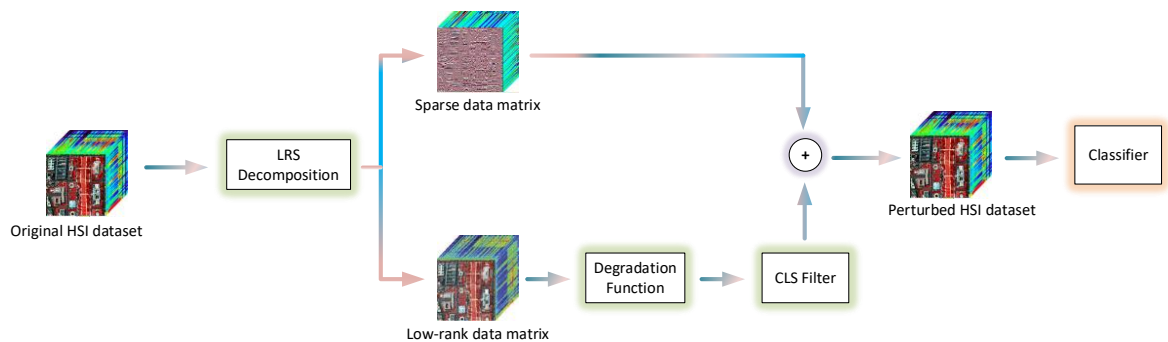

Fig. 4.Flow of LRSP method

An image degradation function based on an atmospheric turbulence model is applied to the low rank matrix. This function adds controlled randomness to the HSI data. The constrained least squares (CLS) filter is used to recover the signal information. This filter performs a deconvolution procedure to remove any corruption caused by the atmospheric turbulence while simultaneously preserving the stochasticity that has been imputed to the data. The last step is to add the sparse data back to produce the final perturbed HSI dataset. This dataset has the appropriate amount of perturbations for classification accuracy improvement. We will now briefly describe the mathematical formulations for each of the steps.

A. *Low-Rank and Sparse Decomposition*

Since most signal information in high-dimensional data generally has low intrinsic dimensionality, it is desirable to model real-world data as a mixture of several low-rank subspaces [14]. Principal Component Analysis (PCA) can be cast as a constraint optimization problem where a data matrix is assumed to be composed of a low-rank matrix and a small perturbation (error) matrix whose entries are *i.i.d.* Gaussian random variables. However, the result of PCA may be quite inaccurate if the entries in the perturbation matrix are arbitrarily large. To address this problem, an approach known as Robust Principal Component Analysis (RPCA) was developed [11]. The Lagrangian formulation is given by:

$$\text{minimize } rank(\mathbf{L}) + \lambda \|\mathbf{E}\|_0 \text{ subject to } \mathbf{X} = \mathbf{L} + \mathbf{E} \tag{1}$$

where λ is a regularization parameter, $\mathbf{X} \in \mathbb{R}^{d \times n}$ is the original high-dimensional data, $\mathbf{L} \in \mathbb{R}^{d \times n}$ is the low-rank matrix, $\mathbf{E} \in \mathbb{R}^{d \times n}$ is the sparse error matrix, and $\|\mathbf{E}\|_0$ is the l_0-norm of matrix $\mathbf{E}$.

Although this optimization problem is highly non-convex, it has been shown that the nuclear norm, which is the sum of the singular values, is an effective surrogate for the rank of a matrix and the l_1-norm for the l_0-norm. Moreover, data drawn from multiple subspaces should be represented as a union of the subspaces $\cup_{i=1}^k S_i = \{y: y \in S_j, 1 \le j \le k\}$. In other words, $\mathbf{X}$ is a set of n d-dimensional vectors strictly drawn from a union of k subspaces [12, 13]. The objective of subspace recovery for data matrix $\mathbf{X} = \mathbf{L} + \mathbf{E}$ is to recover the row space of $\mathbf{L}$. It is still a convex optimization problem given by:

$$\text{minimize } \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \text{ subject to } \mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{E} \tag{0}$$

where $\mathbf{A}$ is a dictionary that linearly spans the data space, $\|\mathbf{Z}\|_*$ is the nuclear norm of matrix $\mathbf{Z}$ and $\|\mathbf{E}\|_{2,1}$ is the $l_{2,1}$-norm of matrix $\mathbf{E}$. The $l_{2,1}$-norm of a data matrix is the sum of the Euclidean norm of each column vector. The minimizer $\mathbf{Z}^*$ is the lowest rank representation with respect to dictionary $\mathbf{A}$. Since $rank(\mathbf{A}\mathbf{Z}^*) \le rank(\mathbf{X})$, $\mathbf{A}\mathbf{Z}^*$ also represents a low-rank recovery of the original data, which is equivalent to matrix $\mathbf{L}$ in (1).

Since off-the-shelf solvers are not sufficiently powerful to solve such a problem, a number of excellent algorithms have been developed over the years. The one used for the experiments in this paper is the Inexact Augmented Lagrange Multiplier [18]. It has been demonstrated that this algorithm is quite stable and highly accurate with fast convergence time.

B. Degradation Function

An image can fluctuate in an unpredictable manner both in intensity and in position as it degrades through a turbulence in the atmosphere. The exact level of degradation varies randomly with time and field angle [19]. Degradation modeling has been used for many years as a solution that allows successful image restoration. The model frames environmental conditions and physical characteristics of atmospheric turbulence as a transfer function $h(x, y)$ where (x, y) is the 2D spatial coordinates [20].

Assuming the degradation function is linear and position invariant, the degraded image in the spatial domain is given by:

$$g(x, y) = h(x, y) * f(x, y) + \eta(x, y) \tag{3}$$

where is $f(x, y)$ the original image and is $\eta(x, y)$ Gaussian noise. Convolution is denoted by $*$. Applying the spatial Fourier transform, (3) can be expressed in the frequency domain as:

$$G(u, v) = H(u, v).F(u, v) + N(u, v) \tag{4}$$

The form of the transfer function $H(u, v)$ is similar to that of a Gaussian low pass filter and is given by:

$$H(u, v) = \exp\left[\left(-k(u^2 + v^2)\right)^{5/6}\right] \tag{5}$$

where k is a hyper-parameter for the blurring effect associated with the nature of the turbulence.

To be compatible with the degradation function, the HSI dataset must be reshaped from a 3D data matrix to a 2D data matrix. The original dataset is composed of $(M \times N)$ pixel vectors of dimension K, which is the number of spectral bands. The new data matrix has R vectors of dimension K where $R = M \times N$.

C. Constrained Least Squares Filter

The CLS algorithm does not require the power spectra of an undegraded image be known in advance. Using it as a filter will restore the image and minimize any sensitivity caused by the degradation function. This algorithm can be expressed as a constraint optimization problem [15].

$$\text{minimize } \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\nabla^2 f(x, y)]^2 \text{ subject to } \left\|\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}\right\|^2 = \|\mathbf{\eta}\|^2 \tag{6}$$

The constraint in (6) is essentially the algebraic manipulation of (4) in vector-matrix form where $\hat{\mathbf{f}}$ is the estimate of the undegraded image instead of the original image. The Laplacian of the image, ∇^2, is added to alleviate a noise sensitivity problem that inherently exists in the transfer function $\mathbf{H}$. ∇^2 for two variables is defined as:

$$\nabla^2 f(x,y) = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y) \qquad (7)$$

The frequency domain solution to (6) is given by:

$$\hat{F}(u,v) = [H^*(u,v)/(|H(u,v)|^2 + \gamma|P(u,v)|^2)]. G(u,v) \qquad (8)$$

where γ is a parameter that must be adjusted to satisfy the constraint in (6). $P(u,v)$ is the Fourier transform of (7). In matrix form, it can be expressed as:

$$p(x,y) = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \qquad (9)$$

The best estimate of the undegraded image can be achieved by simply taking the inverse Fourier transform of (8).

III. Experiments

The datasets for our experiments consist of a synthetic dataset and two real-world HSI datasets (Salinas-A and Indian Pines) [21]. We employed two commonly used multi-class classifiers for HSI data: Subspace Projection Multinomial Logistic Regression (MLRsub) and Linear Discriminant Analysis (LDA). MLRsub [22] is a subspace projection-based classifier that represents hyperspectral data as a linear mixture model. The spectral response of a pixel is a linear combination of the spectral signatures of the endmembers weighted by the corresponding abundance fractions. This approach works well for cases with prevalent nonlinear class separability where projection onto lower subspaces is required to discover the manifold. LDA [23] is a multi-variate generative classifier. It attempts to maximize the posterior probabilities governed by the Bayes rule from the data statistical properties learned during training. This method is best used when linear decision boundaries can easily be applied for class separation in the input space.

All the experiments were conducted using MATLAB simulations running on the ThinkPad X1 Yoga Gen 4 with 1.6 GHz i7 Processor and 16 GB of RAM.

A. Datasets

The synthetic dataset has six endmembers. Six pixel vectors are taken from the Salinas-A scene at randomly chosen locations to form an $80 \times 40 \times 204$ HSI cube. Small Gaussian random noise was added. The spectral signatures are shown in Figure 5.
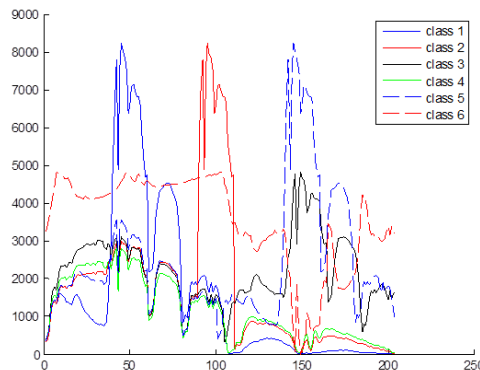

Fig. 5. Synthetic Dataset Spectral Signatures

The Salinas-A dataset, shown in Figure 6, is a subset of the Salinas image collected by the AVIRIS sensor over Salinas Valley in California. It is characterized by high spatial resolution (3.7 m pixel

size) and six classes. 20 water absorption bands are removed (108 – 112, 154 – 167, 224) forming a total of 204 spectral bands for analysis.

(a). Spectral Band 150

(b). Ground Truth

Fig. 6. Salinas-A Spectral Band 75 and Ground Truth. *(figure caption)*

The Indian Pines scene, shown in Figure 7, is a subset of a larger scene over the Indian Pines test site in northwestern Indiana collected by the AVIRIS sensor. It has 16 classes and 224 spectral reflectance bands acquired in the 0.4 – 2.5 um region. The dataset is composed of 145×145 pixels with a ground resolution of 17 m. Two-third of the scene is agriculture. The remaining one-third is forest and other natural perennial vegetation. Two major dual lane highways, a rail line and some low-density housing are also part of the scene. The bands covering the region of water absorption are removed reducing the number of bands to 200.
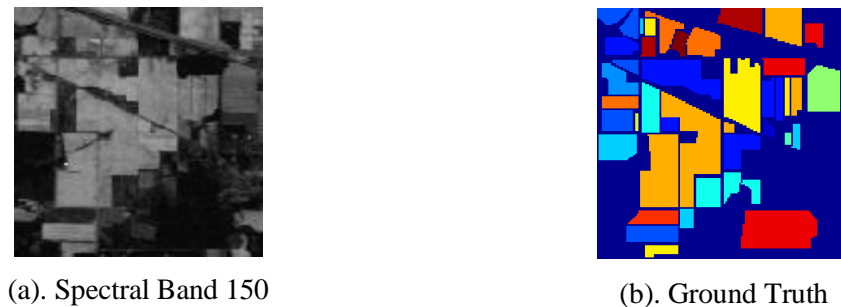

(a). Spectral Band 150

(b). Ground Truth

Fig. 7. Indian Pines Spectral Band 150 and Ground Truth

B. Results

The effect of adding randomness in the dataset is quantified in terms of the dissimilarity measure. It is computed based on the Frobenius norm of the difference between the original and the perturbed datasets.

$$\delta = \left\| \mathbf{X} - \widehat{\mathbf{X}} \right\|_2 / \|\mathbf{X}\|_2 \tag{10}$$

where $\mathbf{X}$ and $\widehat{\mathbf{X}}$ are the original and the perturbed datasets, respectively. The denominator is included as a normalizing factor.

As depicted in Table 1, the blurring effect k has a much greater influence on the dissimilarity measure than the variances of the zero-mean Gaussian noise. This is expected since (4) and (5) show that changing the value of k causes non-linear and exponential perturbations on the data while altering the statistics of the Gaussian noise only produces an additive effect. However, the influence of the noise components should not be underestimated either. Larger variance means more outliers in the data. Although it may not significantly change the statistical characteristics of the dataset, it may cause larger classification errors.

Table 1. Dissimilarity Measures ($\times 10^{-3}$)

σ^2	Synthetic				Salinas-A				Indian Pines			
	k=1e-4	1e-5	1e-6	1e-7	1e-4	1e-5	1e-6	1e-7	1e-4	1e-5	1e-6	1e-7
0.1	83.78	41.46	7.20	5.12	106.4	55.3	22.9	32.1	74.35	45.59	26.59	6.88
0.5	83.78	41.46	7.20	5.12	106.4	55.3	22.9	32.1	74.35	45.59	26.59	6.88
1.0	83.78	41.46	7.20	5.12	106.5	55.3	22.9	32.1	74.37	45.59	26.59	6.88
2.0	83.79	41.46	7.20	5.12	106.5	55.3	22.9	32.1	74.44	45.59	26.59	6.88
5.0	83.80	41.47	7.20	5.12	106.5	55.3	22.9	32.2	74.53	45.60	26.59	6.89
10.0	83.82	41.47	7.21	5.13	106.6	55.3	23.0	32.4	74.84	45.62	26.59	6.89
15.0	83.86	41.48	7.22	5.14	106.8	55.4	23.0	32.8	75.29	45.67	26.60	6.89

Table 2 shows the overall classification accuracy results if the Gaussian noise variance is fixed at 0.1. It is generally accepted that the different classifiers may produce different results. Nevertheless, if the same classifier is used, consistent results are observed. In other words, if the right amount of stochasticity is added to the original dataset, the LRSP method will always produce higher classification accuracy for the same classifier. The best results are indicated in bold associated with k=1e-5. The highest improvement corresponds to the Indian Pines dataset using the MLRsub classifier from 0.5910 to 0.8850 (49.8%). There is hardly any improvement for Salinas-A with LDA (0.4%) because the accuracy is already very high with the original dataset. Further observations show that Salinas-A and Indian Pines datasets contain classes that are linearly separable whereas the synthetic dataset does not. This is the reason why LDA performs better on the former and MLRsub on the latter.

Table 2. Overall Accuracy with Variance = 0.1

	Synthetic		Salinas-A		Indian Pines	
	MLRsub	**LDA**	**MLRsub**	**LDA**	**MLRsub**	**LDA**
Original dataset	0.6588	0.6658	0.9217	0.9921	0.5910	0.7407
k = 1e-4	0.6855	0.6817	0.9471	0.9841	0.8228	0.8340
k = 1e-5	**0.7306**	**0.7037**	**0.9676**	**0.9963**	**0.8850**	**0.9284**
k = 1e-6	0.6941	0.6963	0.9160	0.9933	0.7680	0.8614
k = 1e-7	0.6814	0.6871	0.9248	0.9935	0.6473	0.7723
k = 1e-10	0.6805	0.6850	0.9241	0.9933	0.6000	0.7474

Figures 8, 9 and 10 show classification accuracies for different values of blurring effect as a function of variance. The graphs show that Gaussian noise can have little effect on the results (Figures 8a, 8b and 9b), moderate effect (Figure 10a) and dramatic degradation (Figures 9a and 10b). But in all cases for all the datasets, the right value of k (k = 1e-5) combined with small variances ($\sigma^2 < 1$) always produces better classification accuracy for the perturbed datasets compared to the original datasets. These values define the operating limit of LRSP.
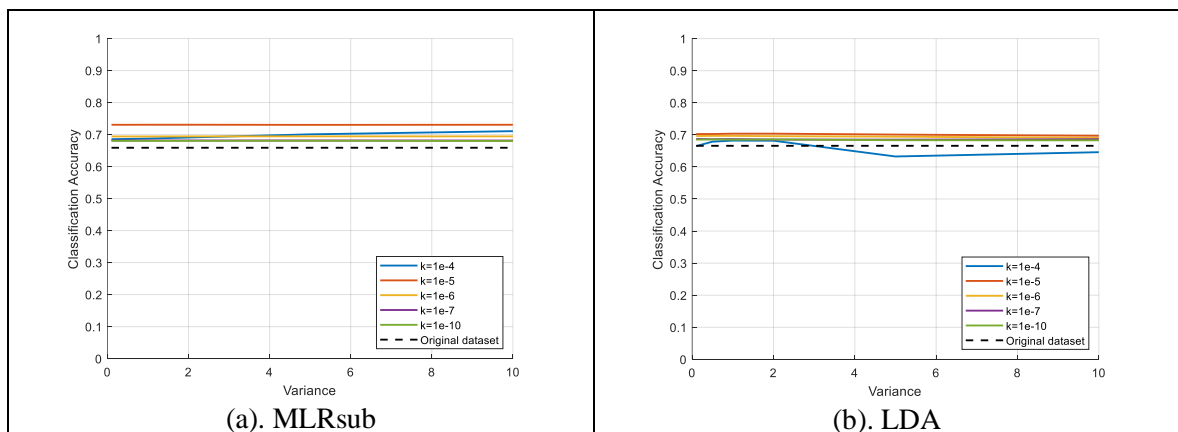
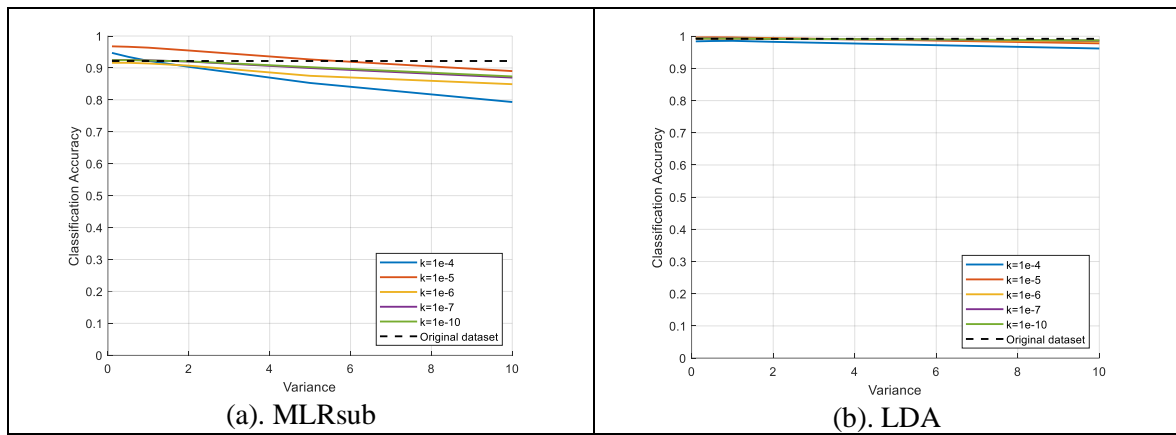

Fig. 8.Classification Accuracy for Synthetic Dataset
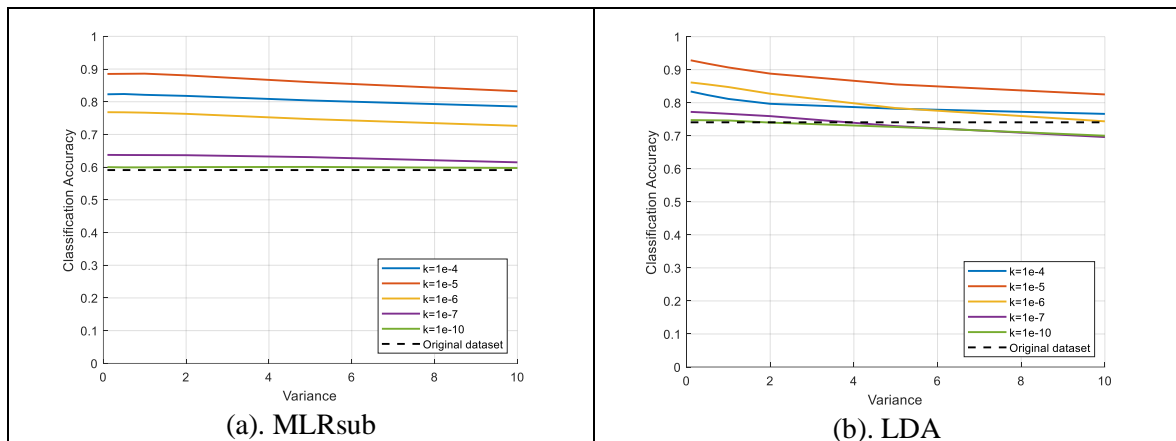
Fig. 9.Classification Accuracy for Salinas-A

Fig. 10. Classification Accuracy for Indian Pines

Classification maps for the two real-world HSI datasets are shown in Figures 11 and 12. It can be easily concluded by visual inspection that the maps in Figures 11b, 11d, 12b and 12d are cleaner than those in Figures 11a, 11c, 12a and 12c, respectively, since they are associated with higher overall classification accuracies.

MLRsub	LDA
(a). Original dataset	(c). Original dataset
(b). Perturbed dataset	(d). Perturbed dataset
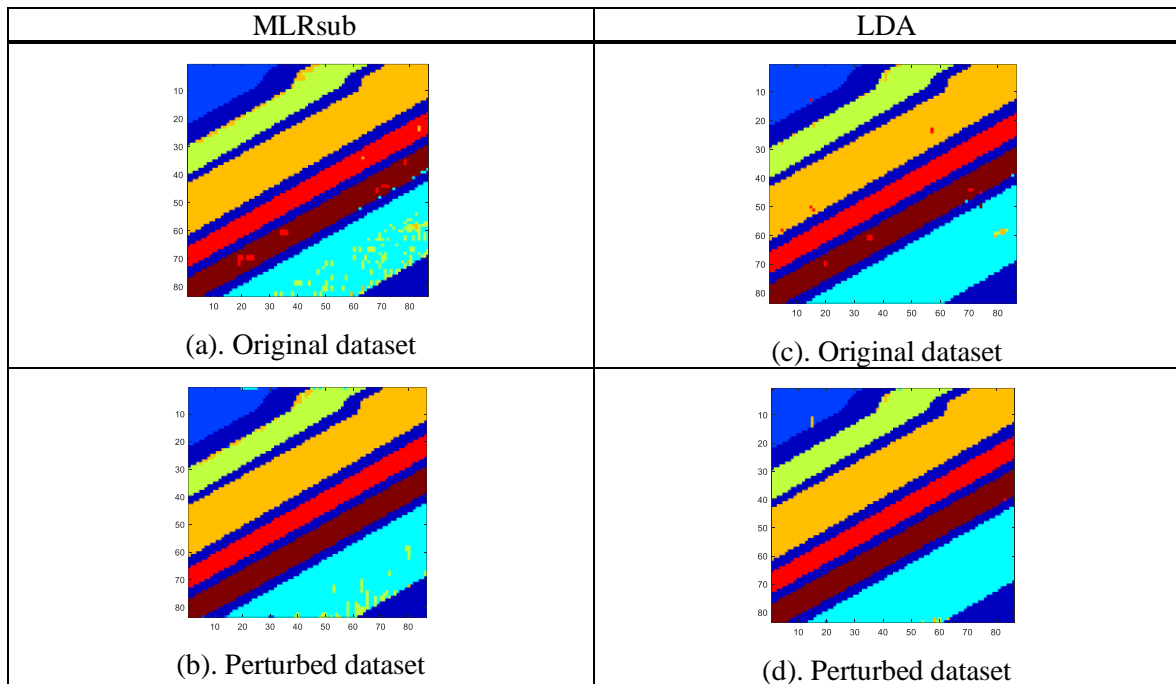
Fig. 11. Classification Maps for Salinas-A for *k=1e-5*, $\sigma^2 = 0.1$

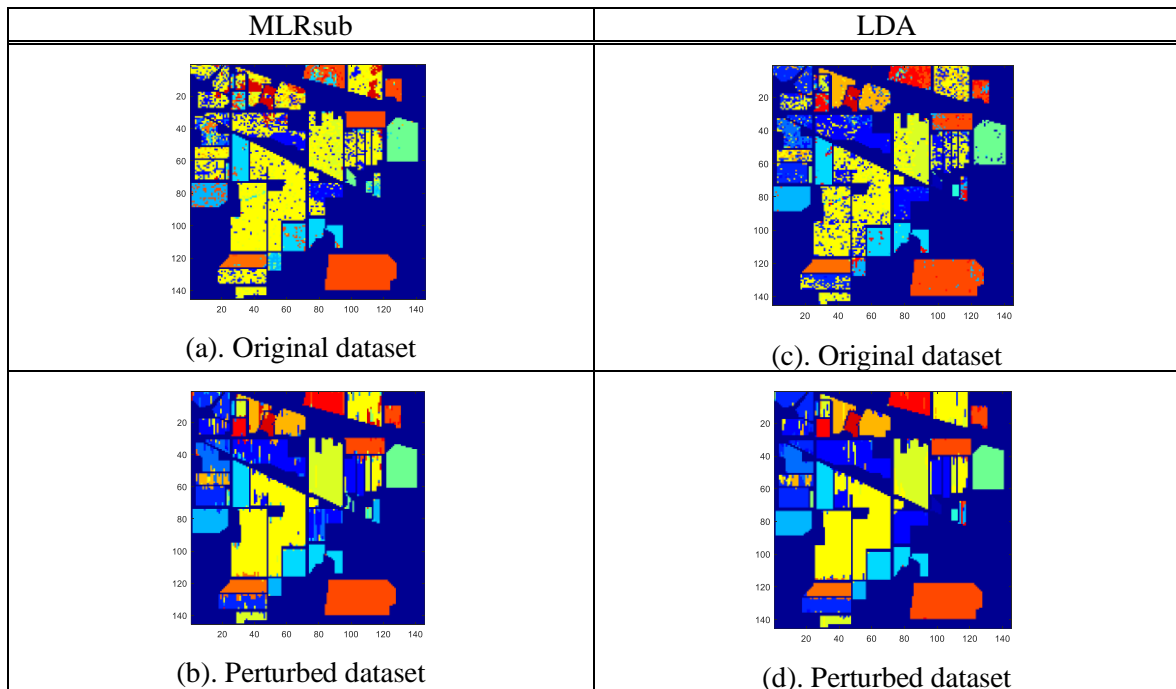| MLRsub	LDA
(a). Original dataset	(c). Original dataset
(b). Perturbed dataset	(d). Perturbed dataset

Fig. 12. Classification Maps for Indian Pines for *k=1e-5*, $\sigma^2 = 0.1$

As listed in tables 3, 4 and 5, the per-class classification accuracies for the perturbed datasets are higher than those of the original datasets. This is again consistent with the overall accuracy results.

Table 3. Per-class Classification Accuracy for Synthetic Dataset

Class	MLRsub		LDA	
	Original dataset	Perturbed dataset	Original dataset	Perturbed dataset
1	1.0000	1.0000	1.0000	0.9730
2	0.6472	0.9806	0.7737	0.8063
3	0.8000	0.6056	0.5072	0.5697
4	0.2972	0.1611	0.5910	0.6058
5	0.1750	0.8556	0.4855	0.5777
6	0.9833	0.9778	0.9460	0.7787
Average	0.6588	0.7306	0.6658	0.7025

Table 4. Per-class Classification Accuracy for SalinasA

Class	MLRsub		LDA	
	Original dataset	Perturbed dataset	Original dataset	Perturbed dataset
1	1.0000	0.9858	0.9977	0.9857
2	0.8673	0.9386	0.9953	0.9977
3	0.9050	0.9283	0.9583	0.9800
4	1.0000	1.0000	0.9968	0.9996
5	0.9950	1.0000	0.9976	0.9996
6	0.9556	0.9985	0.9845	0.9984
Average	0.9217	0.9676	0.9921	0.9963

Table 5. Per-class Classification Accuracy for IndianPines

Class	MLRsub		LDA	
	Original dataset	Perturbed dataset	Original dataset	Perturbed dataset
1	0.0222	0.5111	0.5543	0.8761
2	0.3838	0.7661	0.7025	0.9172
3	0.0830	0.7550	0.6234	0.8808
4	0.1759	0.8843	0.5203	0.8857
5	0.5718	0.9375	0.8749	0.9586
6	0.9087	0.9893	0.9252	0.9903
7	0.1481	0.1852	0.6250	0.9964
8	0.9769	1.0000	0.9759	0.9992
9	0.2278	0.5556	0.2250	0.8400
10	0.2772	0.9450	0.6767	0.9103
11	0.8360	0.9074	0.6660	0.9318
12	0.0998	0.7797	0.7201	0.8801
13	0.9365	0.9894	0.9732	0.9912
14	0.9939	0.9668	0.9182	0.9905
15	0.0399	0.9117	0.6246	0.9251
16	0.8519	0.9012	0.8677	0.9710
Average	0.5910	0.8850	0.7407	0.9284

IV. Conclusion

In this paper we proposed a novel low-rank stochastic perturbation method to improve the classification accuracy for hyperspectral images. Applying an atmospheric turbulence degradation function followed by constraint least squares filtering to the low-rank component of the hyperspectral data causes the statistical nature of the data to be perturbed in a way that is conducive to producing a higher classification accuracy. The experimental results confirm the effectiveness of this method for synthetic and real-world datasets. Popular state-of-the-art HSI classifiers produce better results if the LRSP method is used to preprocess the data. Choosing the appropriate the hyperparameter value for the blurring effect and keeping the variances small can yield quite an improvement of up to 49.8%.

Acknowledgment

References

[1] G. Shaw and D. Manolakis, "Signal Processing for Hyperspectral Image Exploitation," *IEEE Signal Processing Magazine*, vol. 19, no 1, 2002.

[2] J. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*, Prentice Hall, 2007.

[3] D. Landgrebe, "Hyperspectral Image Data Analysis," IEEE Signal Processing Magazine, vol. 19, no. 1, 2002.

[4] H. Yu, X. Shang, X. Zhang, L. Gao, "Hyperspectral Image Classifcation Based on Adjacent Constraint Representation," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[5] X. Shang, M. Song, C. Yu, "Hyperspectral Image Classification with Background," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 2019.

[6] B. Rasti, P. Ghamsi, J. Chanussot, "Mixed Noise Reduction in Hyperspectral Imagery," in 10th *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2019.

[7] A. Sumarsono, Q. Du and N. Younan, "Hyperspectral Image Segmentation with Low-Rank Representation and Spectral Clustering," in *Proceedings of 7th IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing,* 2015.

[8] S. Zhong, C. Chang, J. Li, X. Shang, S. Chen and M. Song, "Class Feature Weighted Hyperspectral Image Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, 2019.

[9] S. Xu, J. Li, "Combining Contextual Information for Subspace Based Hyperspectral Image Classification," in 9th *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2018.

[10] J. Wang, N. Jiang, G. Zhang, B. Hu and Y. Li, "Automatic Framework for Semi-supervised Hyperspectral Image Classification using Self-training with Data Editing," in 7th *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2015.

[11] B. Peng, X. Xie, Q. Du, "Hyperspectral Image Classification Using Multiple Features and Nearest Regularized Subspace," in 7th *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2015.

[12] Y. Zhang, W. Li, Q. Du, X. Sun, "Discriminative Marginalized Least Squares Regression for Hyperspectral Image Classification," in 10th *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2019.

[13] S. Zhong, C. Chang, Y. Zhang, "Iterative Support Vector Machine for Hyperspectral Image Classification," in Proceedings of 25th IEEE International Conference on Image Processing, 2018.

[14] A. Sumarsono, Q. Du and N. Younan, "Low-Rank Subspace Representation for Supervised and Unsupervised Classification of Hyperspectral Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, 2016.

[15] E. Candes, X. Li, Y. Ma and J. Wright, "Robust Principal Component Analysis?," *Journal of the ACM*, vol. 58, no. 3, 2009.

[16] J. Wright, Y. Peng, Y. Ma, A. Ganesh and S. Rao, "Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization," in *Neural Information Processing Systems, NIPS*, 2009.

[17] G. Liu, Z. Lin, S. Yang, J. Sun, Y. Yu and Y. Ma, "Robust Recovery of Subspace Structures by Low-Rank Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, 2013.

[18] Z. Lin, A. Ganesh, L. Wu, M. Chen and Y. Ma, "Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix," UIUC Technical Report UILU-ENG-09-2215, 2009.

[19] R. Gonzales and R. Woods, *Digital Image Processing*, Prentice Hall, 2017.

[20] R. Hufnagel and N. Stanley, "Modulation Transfer Function Associated with Image Transmission through Turbulent Media," *Journal of the Optical Society of America*, vol. 54, no. 1, 1964.

[21] Hyperspcctral Remote Sensing Scenes, March 2013. [Online]. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[22] J. Li, J. Bioucas-Dias and A. Plaza, "Spectral-Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, 2012.

[23] R. Duda and P. Hart, *Pattern Classification*, Wiley-Interscience, 2001.

[24] P. Sidike, C. Chen, V. Asari, Y. Xu and W. Li, "Classification of Hyperspectral Image Using Multiscale Spatial Texture Features," in *Proceedings of 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2016.

[25] L. Sun, T. Zhan, Z. Wu, B. Jeon, "A Novel 3D Anistropic Total Variation Regularized Low Rank Method for Hyperspectral Image Mixed Denoising," International Journal of Geo-Information, vol. 7, 2018.