A Trace Lasso Regularized L1-norm Graph Cut for Highly Correlated Noisy Hyperspectral Image

Ramanarayan Mohanty
ATDC, IIT Kharagpur,
Email: ramanarayan@iitkgp.ac.in

S L Happy
Dept. of EE, IIT Kharagpur,
Email: happy@iitkgp.ac.in

Nilesh Suthar
Dept. of Biotechnology
IIT Kharagpur,

Aurobinda Routray
Dept. of EE, IIT Kharagpur,
Email: aroutray@iitkgp.ac.in

Abstract—This work proposes an adaptive trace lasso regularized L1-norm based graph cut method for dimensionality reduction of Hyperspectral images, called as 'Trace Lasso-L1 Graph Cut' (TL-L1GC). The underlying idea of this method is to generate the optimal projection matrix by considering both the sparsity as well as the correlation of the data samples. The conventional L2-norm used in the objective function is sensitive to noise and outliers. Therefore, in this work L1-norm is utilized as a robust alternative to L2-norm. Besides, for further improvement of the results, we use a penalty function of trace lasso with the L1GC method. It adaptively balances the L2-norm and L1-norm simultaneously by considering the data correlation along with the sparsity. We obtain the optimal projection matrix by maximizing the ratio of between-class dispersion to within-class dispersion using L1-norm with trace lasso as the penalty. Furthermore, an iterative procedure for this TL-L1GC method is proposed to solve the optimization function. The effectiveness of this proposed method is evaluated on two benchmark HSI datasets.

Index Terms—Correlation, dimensionality reduction, graph cut, greedy method, hyperspectral classification, L1-norm, sparsity, trace lasso.

I. Introduction

Hyperspectral remote sensing images (HSI) capture the inherent properties of the observed scene through hundreds of contiguous spectral bands. These spectral bands are highly correlated [1] and rich with information contents. The captured information of the spectral bands offer great potentials for information retrieval. However, these large number of correlated spectral bands contain redundant information. This high dimensional redundant data is a significant challenge in HSI information extraction for the subsequent classification task. Spectral domain dimensionality reduction (DR) method is widely employed in HSI data analysis to mitigate this challenge. It improves the classification performance by reducing computational complexity as well as storage capacity and explores the intrinsic property of the extracted features.

For HSI analysis, a wide variety of DR methods have drawn attention. There are broadly two types of DR methods, unsupervised and supervised. Among them, principal component analysis (PCA) [2] is widely used in the unsupervised category. Whereas, in supervised category linear discriminant analysis (LDA) [3] is the most popular one, which uses the labeled samples to determine the projection matrix. However, LDA considers the class center for its computation by assuming the data distribution as Gaussian. Hence, it fails to handle the real world HSI data with more complex distribution. The graph-

based scaling cut (SC) method proposed in [4], [5] address the issue mentioned earlier. This graph-based method addresses the issue by constructing a pairwise similarity matrix among the samples of the classes.

The idea behind the graph based SC method is to project the data into a lower dimensional space by maximizing the between-class variance with respect to the total-class variance. However, the SC method by Zhang *et al.* in [5] are sensitive in handling the outliers and noise in the dataset, because the SC method uses the conventional L2-norm to compute the dissimilarity matrix between the samples. The square operation in the L2-norm exaggerates the effects of the outliers and noises in the data.

Generally, L1-norm based DR method is a robust alternative to the L2-norm based method to handle the outliers problem [6], [7], [8], [9], [10], [11]. Li *et.al.* in [12] proposed the 2D version of the LDA (L1-2DLDA) using the L1-norm optimization. Similarly, in [6], Wang *et.al.* proposed LDA-L1 by solving the supervised LDA method iteratively using L1-norm maximization. In [8], Mohanty *et.al.* computed the dissimilarity matrix using L1-norm and proposed the L1-SC method by greedy strategy. However, the L1-norm have an undesired stability problem when the data are strongly correlated with each other [13]. For example, if two variables are correlated, the L1-norm based method will only choose one of them randomly without considering any correlation metric.

The L1-norm based SC method in [8] overlooks the spectral correlation property of HSI data, which adversely affect its performance. The addition of a convex penalty or regularization term (squared L2-norm) [14] to the L1-norm based objective function is the best possible remedy for the instability issue [13]. However, these squared L2-norm based regularization terms are blind to the exact correlation structure of the data. That means this squared regularization term is not uniformly required for every variable or the regularizer is not adaptive. Hence, "trace lasso" was proposed in [13] as an adaptive regularizer to keep the balance between L1 and L2-norm by considering both the sparsity and correlation of the data.

L1-norm based SC has achieved excellent performance for noisy HSI data classification in [8]. The trace lasso (TL) regularizer has proved its worth on feature extraction of image datasets in [15]. However, our HSI data is noisy and has the strong spectral correlation. Furthermore, the high-resolution HSI data exhibit high spatial variability. That lead to

poor spatial pixel correlation resulting undesired classification results. Hence, pixel consistency is essential for maintaining spatial correlation in HSI data. Motivated by the existing literature, we propose a trace lasso regularized L1-norm based graph cut method (TL-L1GC) preceded by a guided filter [16] for DR and classification of noisy and correlated HSI data. In this work, we formulate the GC algorithm into an L1-norm optimization problem by maximizing the ratio of between-class dispersion and within-class dispersion matrix. Then, we solve this TL regularized L1-norm optimization problem by using an iterative algorithm to generate a projection matrix. The projected reduced dimension HSI data are further used for classification by using SVM classifier.

II. RELATED WORKS

The classical LDA method is computed using the class centers of the data, which means it assumes that, the distribution of the data is Gaussian with equal variance. However, the characteristics of real-world HSI data is multimodal, heteroscedastic with more complex distribution than Gaussian [17]. The scaling cut (SC) method is proposed in [5], [17] to overcome this limitation of LDA by constructing the dissimilarity matrix among the data samples.

Let $X = \{x_i, L_i\}|_{i=1}^n \in R^{D \times n}$ is the input training dataset. Here $L_i = \{1, 2, ..., C\}$ is the class label of the corresponding training data with total C classes and n training data samples. The objective of DR methods is to determine a projection matrix W that project the input training data X of dimensions D into reduced dimension d, i.e., $Y = W^T X$ such that $Y \in R^{d \times n}$, $W \in R^{D \times d}$, and $d << D$. The between-class dissimilarity matrix and the within-class dissimilarity matrix of SC are defined as

$$
\begin{aligned}
S_{B_k} &= \sum_{x_i \in U_k} \sum_{x_j \in \bar{U}_k} \frac{1}{n_k n_{\bar{k}}} (x_i - x_j)(x_i - x_j)^T \\
S_{W_k} &= \sum_{x_i \in U_k} \sum_{x_j \in U_k} \frac{1}{n_k n_k} (x_i - x_j)(x_i - x_j)^T
\end{aligned}
\tag{1}
$$

where U_k represents all the samples from kth class and n_k is the total number of elements in U_k. Similarly, $\bar{U}_k$ represents all the data points that does not belong to the kth class and $n_{\bar{k}}$ denotes the total number of elements in $\bar{U}_k$. S_{B_k} represents the dissimilarity matrix between U_k and $\bar{U}_k$, whereas S_{W_k} is the dissimilarity matrix within the U_k class. Based on the S_{W_k} and S_{B_k}, the objective function $J(W)$ of SC can be written as

$$
\begin{aligned}
J(W) &= \frac{\left| \sum_{k=1}^{c} W^T S_{B_k} W \right|}{\left| \sum_{k=1}^{c} (W^T S_{W_k} W + W^T S_{B_k} W) \right|} \\
&= \frac{|W^T S_B W|}{|W^T (S_W + S_B) W|} = \frac{|W^T S_B W|}{|W^T S_T W|}
\end{aligned}
\tag{2}
$$

where, $S_B = \sum_{k=1}^{C} S_{B_k}$ is denoted as the between-class, $S_W = \sum_{k=1}^{C} S_{W_k}$ the within-class, $S_T = (S_B + S_W)$ is the total-class dissimilarity matrix and W is the projection matrix.

III. PROPOSED TRACE LASSO REGULARIZED L1-GC

The proposed method initially use the guided filter [16] to incorporate the spatial correlation with the spectral correlation by maintaining the pixel consistency. The discriminant information in the filtered data is extracted by the proposed TL-L1GC.

The conventional L2-norm based graph cut methods mostly concentrate on the geometric structure and correlation between the data points. However, they are prone to noises and outliers. Although L1-norm is a robust alternative to it, they omit the data correlation, that lead to unstable results. To mitigate the above all issues, we focus on robust trace lasso regularized L1-norm based graph based scaling cut criterion (L1GC), which adaptively balance the noise sensitivity along with the correlation among the data. This formulation replaces the L2-norm optimization by L1-norm optimization. Then to balance the L1 and L2 norm and considering sparsity and correlation simultaneously, it utilizes the trace lasso to regularize the basis vector. This trace lasso regularized L1-norm based SC method is solved by following an iterative algorithm

Inspired by the existing literature on L1-norm based scaling cut method in [8] and trace lasso regularization in [13] and [15], we propose to maximize the SC criterion using L1-norm with trace lasso regularizer. The equation (2) can be simplified to a trace difference [18], [5] problem, which can further be reduced to the Frobenius norm based trace difference problem, given by,

$$
\begin{aligned}
W^* &= \arg \max_{W^T W = I} \frac{Tr(W^T S_B W)}{Tr(W^T S_W W)} \\
&= \arg \max_{W^T W = I} Tr(W^T (S_B - \lambda S_W) W) \\
&= \arg \max_{W^T W = I} (Tr(W^T S_B W) - Tr(W^T S_W W)) \\
&= \max_{W^T W = I} \sum_{\substack{k; x_i \in U_k; \\ x_j \in \bar{U}_k}} \frac{1}{n_k n_{\bar{k}}} Tr\left(W^T (x_i - x_j)(x_i - x_j)^T W\right) \\
&\quad - \sum_{\substack{k; x_i \in U_k; \\ x_j \in U_k}} \frac{1}{n_k n_k} Tr\left(W^T (x_i - x_j)(x_i - x_j)^T W\right) \\
&= \max_{W^T W = I} \sum_{k=1}^{c} \sum_{x_i \in U_k} \sum_{x_j \in \bar{U}_k} \frac{1}{n_k n_{\bar{k}}} \left\| W^T (x_i - x_j) \right\|_F^2 \\
&\quad - \sum_{k=1}^{c} \sum_{x_i \in U_k} \sum_{x_j \in U_k} \frac{1}{n_k n_k} \left\| W^T (x_i - x_j) \right\|_F^2
\end{aligned}
\tag{3}
$$

where, λ is the eigenvalue parameter. As can be observed, the above objective function is the trace difference problem based on Frobenius norm. As it involves the square operations, it is sensitive to noise and outliers. To reduce the sensitivity and to maintain the correlation, we reformulate the objective function in terms of the L1-norm with trace lasso regularizer. The proposed model of the objective function for L1-norm SC with trace lasso regularizer is defined as,

$$v_{opt} = \max_{v^T v=1} \sum_{k=1}^{c} \sum_{x_i \in U_k} \sum_{x_j \in \bar{U}_k} \left\| v^T \frac{1}{n_k n_{\bar{k}}} (x_i - x_j) \right\|_1$$

$$- \sum_{k=1}^{c} \sum_{x_i, x_j \in U_k} \left\| v^T \frac{1}{n_k n_k} (x_i - x_j) \right\|_1 - \delta ||X^T Diag(v)||_*$$

$$= \max_{v^T v=1} \sum_{k=1}^{c} \sum_{x_i \in U_k} \sum_{x_j \in \bar{U}_k} \frac{1}{n_k n_{\bar{k}}} |v^T(x_i - x_j)|$$

$$- \sum_{k=1}^{c} \sum_{x_i, x_j \in U_k} \frac{1}{n_k n_k} |v^T(x_i - x_j)| - \delta ||X^T Diag(v)||_*$$

$$= \arg \max_{v^T v=1} ||(v^T F)||_1 - ||(v^T G)||_1 - \delta ||X^T Diag(v)||_*$$

$$(4)$$

and,

$$F = \sum_{k=1}^{c} \sum_{x_i \in U_k} \sum_{x_j \in \bar{U}_k} \frac{1}{n_k n_{\bar{k}}} |v^T(x_i - x_j)|$$

$$G = \sum_{k=1}^{c} \sum_{x_i \in U_k} \sum_{x_j \in U_k} \frac{1}{n_k n_k} |v^T(x_i - x_j)|$$

$$(5)$$

where, $F = [f_1, f_2, ..., f_n] \in R^{D \times n}$ is the between-class dispersion, $G = [g_1, g_2, ..., g_n] \in R^{D \times n}$ is the within-class dispersion, $||X^T Diag(v)||_*$ is the trace lasso, δ is the regularizing parameter, $|| \bullet ||_1$ denote the L1-norm and $|| \bullet ||_*$ denotes the trace norm of the matrix (trace lasso), i.e. the sum of its singular values. $Diag(\bullet)$ represents the conversion of a vector to a diagonal matrix. As oppose to $Diag(\bullet)$, we will use $diag(\bullet)$ to represent the diagonals of a matrix. The term X in trace lasso ($||X^T Diag(v)||_*$) is the sample data matrix. As per [13], the trace lasso interpolate between L2 and L1 norm to consider both sparsity as well as correlation of the data simultaneously. The optimized projection vector (v_{opt}) in (4) is used to construct the optimal projection matrix $V = \{v_1, v_2, ..., v_d\}$. These projecting vectors are sequentially optimized in d directions. The following iterative algorithm is derived to find the optimal projection vector v. The entire optimization procedure for TL-L1GC method is listed below.

Optimization procedure of TL-L1GC

We used an iterative algorithm to determine the local optimal v_{opt} by maximizing the equation (4). The basic idea of the proposed technique is to iteratively update the v_{opt} until it converges to its local optima.

As per [13] and [15], the trace norm of a matrix $J \in R^{n \times d}$ is computed as

$$||J||_* = \frac{1}{2} \inf_{s \succ \underline{0}} tr(J^T S^{-1} J) + tr(S) \qquad (6)$$

where, $tr(\bullet)$ represent the trace of the matrix and the infimum of the matrix is achieved for $S = (J^T J)^{1/2}$.

By using the equation (6), we can reformulate the objective function in equation (4) as

$$\arg \max_{v,S} ||(v^T F)||_1 - ||(v^T G)||_1$$

$$- \frac{\delta}{2} v^T Diag(diag(X S^{-1} X^T)) v - \frac{\delta}{2} tr(S) \qquad (7)$$

The maximum value of the equation (7) is achieved for $S = (X^T Diag(v)^2 X)^{1/2}$. The variable S and v in the objective function are optimized alternatively: updating S with v fixed, and updating v with S fixed. Here v is updated with fixed S in the following equation

$$\arg \max_{v,S} ||(v^T F)||_1 - ||(v^T G)||_1$$

$$- \frac{\delta}{2} v^T Diag(diag(X S^{-1} X^T)) v \qquad (8)$$

The objective function in (8) is similar to the trace difference formulation of the generalized graph cut in [5]. The regularization term in the objective function makes it difficult to determine the optimal v. Inspired by the current literature [8] and [15], we are using trace lasso regularized L1-norm optimization technique in this work. Thus, we solve the objective function (8) to find the optimal projection vector v^* by the iterative technique. The algorithmic procedure of TL-L1GC is given as follows.

1. The iteration variable t is set to zero ($t = 0$). Then we randomly initialize the d dimensional vector $v(t)$ and normalize it such that $v(t)^T v(t) = 1$.

2. Initially, we compute the inverse of S, i.e. S^{-1}, by using the following equation

$$S^{-1} = U Diag(1/\sqrt{s_i}) V \qquad (9)$$

where, U and V are the left and right eigen vectors and s_i is the ith ($i = 1, 2, ..., n$) eigen value of $X^T Diag(v(0))^2 X$. Here, U and V are orthogonal matrices. The computed S^{-1} is fixed and used to update the $v(t)$ to $v(t+1)$ in the following steps.

3. Two sign functions are defined to compensate the absolute value operation for between-class and within-class dispersion term of (8) . These sign functions are computed as

$$q_i(t) = \begin{cases} 1, & \text{if } v^T(t) f_i > 0 \\ -1, & \text{if } v^T(t) f_i \leq 0 \end{cases} \quad i = 1, 2, ..., n$$

$$\text{and} \qquad (10)$$

$$r_j(t) = \begin{cases} 1, & \text{if } v^T(t) g_j > 0 \\ -1, & \text{if } v^T(t) g_j \leq 0 \end{cases} \quad j = 1, 2, ..., n$$

4. Then the $v(t)$ is updated to $v(t + 1)$ based on the Theorem 1 in [15]. It is given as

$$v(t+1) = (\delta Diag(diag(X^T S^{-1} X) + M(t)))^{-1} N(t) \qquad (11)$$

where δ is the regularizing parameter, $N(t) = \sum_{i=1}^{n} q_i(t) f_i$,

and $M(t) = \sum_{i=1}^{n} \frac{g_i g_i^T}{|z_i(t)|}$, here $z_i(t) = v(t)^T g_i$.

5. Convergence check: If the objective function at $v(t)$ doesn't show significant increment on $v(t + 1)$ or $||J\{v(t+1)\} - J\{v(t)\}|| \leq \epsilon$ or total iteration number is greater then maximum given iteration number, then go to step-7, otherwise go to step-3.

6. Stop iteration and assign $v^* = v(t+1)$ and update train data by

$$X \leftarrow X - v^*(v^{*T})X \qquad (12)$$

7. Append the projection vector v^* in the column space of the projection matrix $V = [V, v*]$.

8. If the iteration $< d$, then goto step 2. Here the updated train data and $v(t+1)$ are used to update the S^{-1} in step 2. Further, the computation of $v(t+2)$ is depended upon the updated S^{-1}. Thus, the algorithm will end up with V such that $V \in R^{D \times d}$.

Using the above procedure, we can form the optimal projection matrix V of size $R^{D \times d}$. The pseudo-code for the complete algorithmic procedure for the projection matrix of TL-L1GC is listed in Algorithm 1

Algorithm 1 Trace lasso regularized L1GC algorithm

Input: The training dataset $\{x_i, L_i\}_{i=1}^n \in R^{D \times n}$; L_i is the label of each training data x_i; Desired dimensionality is d and $d \ll D$.

1: Formulate the trace lasso regularized L1-norm based graph cut objective function in (4) to solve the optimization problem.

2: Compute $F = [f_1, f_2, ..., f_n]$ and $G = [g_1, g_2, ..., g_n]$ using (5).

3: Set $t = 0$, $iter = 0$, projection matrix $V = []$ and initialize $v(0)$ such that $v(0)^T v(0) = 1$

4: **foreach** $\{1...d\}$ **do**

5:   Compute the S^{-1} using Eq. (9)

6:   **while** true **do**

7:     $iter = iter + 1$

8:     Compute the sign function $q_i(t)$ and $r_j(t)$ using Eq. (10)

9:     Compute $Fv = ||v(t)^T F||_1$ and $Gv = ||v(t)^T G||_1$

10:    Update the $v(t + 1)$ by using Eq. (11)

11:    Use $v(t + 1)$ to determine $Fv' = ||v(t+1)^T F||_1$ and $Gv' = ||v(t+1)^T G||_1$

12:    Converge if: $||v(t + 1) - v(t)|| \leq \epsilon$ or $iter > itmax$

13:    $v(t) \leftarrow v(t + 1)$

14:   **endwhile**

15:   Update the input data X by using Eq. (12)

16:   Pad these optimal projection vectors v^* into the optimal matrix by $V = [V, v^*] \in R^{D \times d}$

17: **endfor**

Output: Projection matrix $V = \{v_1, v_2, ..., v_d\} \in R^{D \times d}$, consists of d projection vectors

---

IV. EXPERIMENTS

This section evaluates the performance of the proposed TL-L1GC method on two benchmark HSI datasets[1]: Botswana ($D = 145, C = 14$) and Salinas ($D = 204, C = 16$). Here, we have made a comparative study with the state-of-the-art L2 based methods (such as SC [5], LSC [17]) and L1 based methods (e.g. L1-LDA [6], L1-SC [8]). The conventional L2-norm based methods use PCA as preprocessing step. In these experiments, 10 samples are randomly selected from each class of the dataset for training and the remaining data samples are used for testing. All the results are obtained by taking the average of the 5 iterations. In order to identify the robustness of the proposed algorithm, SVM with a linear kernel is used as the classifier. Empirically, the regularizing parameter δ is tuned to 0.4 at varied dimensions ranges from 4 to 60 with data samples 10.

The proposed TL-L1GC method is compared with other state-of-the-art L2-norm and L1-norm based methods. The statistics of the highest OA along with corresponding average accuracy (AA), kappa coefficient (k) and dimension of the algorithms for Botswana and Salinas dataset are highlighted in Table I. Here the classification results are obtained by performing the average of 5 iterations while using 10 random training samples per class. The classification results of the proposed TL-L1GC method significantly outperforms the other L2-norm as well as L1-norm based methods for both the datasets in all the metrics. In Botswana dataset, the proposed TL-L1GC method outperforms the other methods by a large margin of approximately 10% of OA at the reduced dimension of 58. Similarly, for Salinas data set, it also outperforms others at 54 dimensions. This verifies the effectiveness of the algorithm in finding proper projection direction.
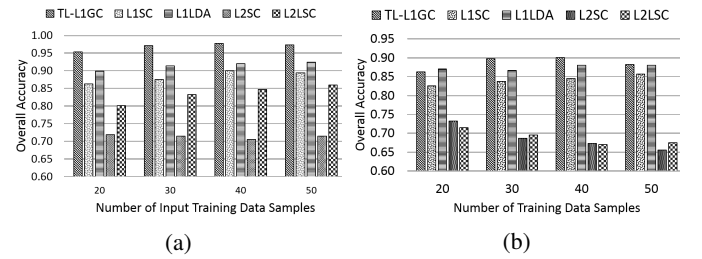

Fig. 1: Illustration of performance of the proposed method with respect to other methods in terms of overall classification accuracies for different input training samples ($20, 30, 40$ and 50) on two HSI datasets Botswana (1a) and Salinas (1b).

Fig. 1 illustrates the behavior of different L2-norm and L1-norm based algorithms in terms of overall classification accuracy with respect to varied number of input data samples. From this figure, it is clearly observed that the proposed TL-L1GC method significantly outperforms the other L1-norm and L2-norm based methods when the input training data sample size is less.

[1]http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote _Sensing_Scenes

TABLE I: Classification performance of proposed approach compared with other L2-norm and L1-norm based approaches

Dataset	Botswana				Salinas			
Methods	OA	AA	k	dim	OA	AA	k	dim
SC [5]	76.81	77.09	59.15	56	81.11	87.57	62.75	58
LSC [17]	78.12	79.25	60.20	52	82.98	88.84	63.12	50
L1-LDA [6]	86.90	88.18	69.02	56	83.66	88.15	61.92	56
L1-SC [8]	84.06	85.53	65.95	42	83.57	86.81	61.70	56
TL-L1GC (Without Filter)	**84.92**	**85.94**	**77.54**	**58**	**84.54**	**87.25**	**70.45**	**54**
TL-L1GC (With Filter)	**94.54**	**94.69**	**75.54**	**58**	**86.42**	**91.67**	**67.53**	**54**

In order to, demonstrate the noise robustness features of the proposed TL-L1GC method with respect to others, we inject white Gaussian noises of different levels to the raw input HSI data and performed the classification on these data using different approaches. The variance of the noise level is varied from 1% to 10% of the variance of the pixel values at fixed dimension 50 and number of training sample 10. As observed in Fig. 2, the proposed TL-L1GC method is more robust to noises and achieve better classification accuracies than other methods.
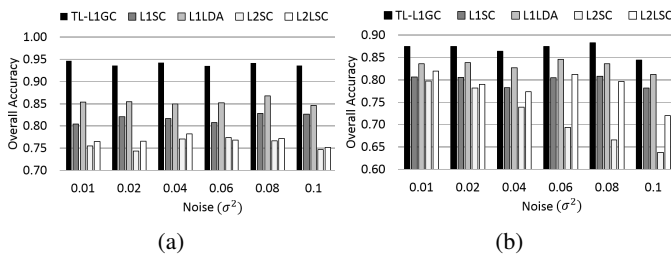

Fig. 2: Illustration of noise robustness of the proposed method with respect to other methods on two HSI datasets Botswana (2a) and Salinas (2b).

V. CONCLUSION

In this study, we have proposed a novel DR method TL-L1GC preceded by a guided filter through computing the trace lasso regularized L1-norm based graph cut. This method determines the projection directions by adaptively considering both the sparsity as well as data correlation in both spectral and spatial domain. It exploits the discriminant structure while preserving the geometrical structure of the data. Our method differs from other state-of-the-art methods in various ways. For instance, it maintains the intrinsic property as well as the distribution of the data, and we believe it handles multimodal and heteroscedastic noisy data with high correlation quite well. We examined the performance of our method and other methods over two real-world HSI datasets. The promising results of TL-L1GC on these two datasets demonstrates its noise robustness and efficiency.

REFERENCES

[1] J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li, and S. Marshall, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4418–4433, 2015.

[2] A. M. Martínez and A. C. Kak, "Pca versus lda," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 2, pp. 228–233, 2001.

[3] J. Ye, R. Janardan, C. H. Park, and H. Park, "An optimization criterion for generalized discriminant analysis on undersampled problems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 8, pp. 982–994, 2004.

[4] X. Zhang, S. Zhou, and L. Jiao, "Local graph cut criterion for supervised dimensionality reduction," in *Sixth International Symposium on Multispectral Image Processing and Pattern Recognition*. International Society for Optics and Photonics, 2009, pp. 74 962I–74 962I.

[5] X. Zhang, Y. He, L. Jiao, R. Liu, J. Feng, and S. Zhou, "Scaling cut criterion-based discriminant analysis for supervised dimension reduction," *Knowledge and information systems*, vol. 43, no. 3, pp. 633–655, 2015.

[6] H. Wang, X. Lu, Z. Hu, and W. Zheng, "Fisher discriminant analysis with l1-norm," *IEEE transactions on cybernetics*, vol. 44, no. 6, pp. 828–842, 2014.

[7] Y. Liu, Q. Gao, S. Miao, X. Gao, F. Nie, and Y. Li, "A non-greedy algorithm for l1-norm lda," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 684–695, 2017.

[8] R. Mohanty, S. Happy, and A. Routray, "Graph scaling cut with l1-norm for classification of hyperspectral images," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 793–797.

[9] Q. Ke and T. Kanade, "Robust l/sub 1/norm factorization in the presence of outliers and missing data by alternative convex programming," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 739–746.

[10] N. Kwak, "Principal component analysis based on l1-norm maximization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 9, pp. 1672–1680, 2008.

[11] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2dpca," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 1170–1175, 2010.

[12] C.-N. Li, Y.-H. Shao, and N.-Y. Deng, "Robust l1-norm two-dimensional linear discriminant analysis," *Neural Networks*, vol. 65, pp. 92–104, 2015.

[13] E. Grave, G. R. Obozinski, and F. R. Bach, "Trace lasso: a trace norm regularization for correlated designs," in *Advances in Neural Information Processing Systems*, 2011, pp. 2187–2195.

[14] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[15] G.-F. Lu, J. Zou, and Y. Wang, "L1-norm and maximum margin criterion based discriminant locality preserving projections via trace lasso," *Pattern Recognition*, vol. 55, pp. 207–214, 2016.

[16] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE transactions on geoscience and remote sensing*, vol. 52, no. 5, pp. 2666–2677, 2014.

[17] X. Zhang, Y. He, N. Zhou, and Y. Zheng, "Semisupervised dimensionality reduction of hyperspectral images via local scaling cut criterion," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1547–1551, 2013.

[18] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 2013.