Article

Relative Weighted Feature Space for Dimensionality Reduction and Classification of Hyperspectral Images

Seyyed A. Ahmadi [1], Nasser Mehrshad [2,*] and Seyyed M. Razavi [3]

[1,2,3] Faculty of Electrical & Computer Engineering, University of Birjand, Birjand 97175, Iran
* Correspondence: nmehrshad@birjand.ac.ir; Tel.: +98-56-3220-2049

Abstract: Containing hundreds of spectral bands (features), hyperspectral images (HSIs) have high ability in discrimination of land cover classes. Traditional HSIs data processing methods consider the same importance for all bands in the original feature space (OFS), while different spectral bands play different roles in identification of samples of different classes. In order to explore the relative importance of each feature, we learn a weighting matrix and obtain the relative weighted feature space (RWFS) as an enriched feature space for HSIs data analysis in this paper. To overcome the difficulty of limited labeled samples which is common case in HSIs data analysis, we extend our method to semisupervised framework. To transfer available knowledge to unlabeled samples, we employ graph based clustering where low rank representation (LRR) is used to define the similarity function for graph. After construction the RWFS, any arbitrary dimension reduction method and classification algorithm can be employed in RWFS. The experimental results on two well-known HSIs data set show that some dimension reduction algorithms have better performance in the new weighted feature space.

Keywords: dimension reduction; feature extraction; hyperspectral image; weighted feature space; low rank representation; spectral clustering

1. Introduction

Hyperspectral imagery has been increasingly utilized in a wide range of fields due to its invaluable source of spectral information at high resolution. Using hundreds of spectral bands makes hyperspectral image data suitable for discriminating different classes [1]. However, it leads to high-dimensionality data and obstacles such as Hughes phenomenon [2] or the curse of dimensionality [3] arise in small sample size situations. Thus, it is necessary to develop advanced data interpretation methods which are able to deal with high-dimensional data sets and limited training samples [4].

Dimensionality reduction (DR) methods are used as a preprocessing step in order to reduce the dimensionality of available data without dropping significant information [3]. DR methods can be realized using feature selection (FS) and feature extraction (FE) has been proposed to alleviate the problems of high dimensionality of data such as HSIs data. FS techniques aim to find the most discriminative subset of features without decreasing the information content [5] based on the evaluation of a fitness function followed by a selection strategy [3,5-7]. On the other hands, in FE methods, a new more powerful subset of features is determined usually by using a projection matrix [3,8].

There are a number of DR methods of HSIs ranging from unsupervised methods to supervised ones according to their use of label information [9]. Principle component analysis (PCA) is the best well-known unsupervised DR method which seeks a transform that can preserve the largest data variance [10,11]. In most HSIs analysis, it has been found that the feature values of the number of the first PCA components account for most of the total original image spectral information [12].

 doi:10.20944/preprints201712.0057.v1

Linear discriminant analysis (LDA) is the most widely used supervised DR method which tends to maximize the between-class scatter while minimizing the within-class scatter. LDA can extract $C-1$ features where C denotes the number of classes [13].

Usually, both PCA and LDA fail to provide acceptable performance in some cases because of their key limitations [14]. Different DR methods have been proposed and used for HSIs analysis either by presentation a new idea or combination of existing methods. Nonparametric weighted FE (NWFE) improves LDA by a nonparametric extension of scatter matrices and can extract more than $C-1$ features [15]. Locality Preserving Projection (LPP) as a popular manifold learning method is used to preserve the locality of sample structure [16], and works in supervised and unsupervised framework [8]. We refer the reader to [3,9] for a complete survey. In order to deal with the nonlinear structure of data, some kernelized DR methods have been proposed which perform in Reproducing Kernel Hilbert Space (RKHS) [17], such as kernel PCA (KPCA) [18], kernel LDA (KLDA) [19], kernel LPP (KLPP) [20] and FKNFLE [21].

As mentioned before, it is fairly expensive and difficult to collect labeled samples in most HSIs analysis applications, so, very limited labeled samples are usually available. On the other hand, unlabeled samples are available in large quantities at very low cost [22-24]. Semisupervised learning (SSL) [23,25,26] and active learning (AL) [27-29] are two strategy which have been proposed to overcome the difficulties of limited labeled samples. However, AL and SSL have different workflows [27]. In recent years, both AL and SSL have been successfully applied in hyperspectral image analysis such as [27,30,31]. The main idea of SSL is to incorporate unlabeled samples for improving the classification and generalization capability on the testing/unseen samples [3,32]. Commonly the unlabeled samples are applied as an extra regularization term to the objective function from traditional supervised learning algorithms [3,33]. Active learning aims to find the most informative additional training inputs through an iterative process [34,35].

Although the key to the design of a powerful classification system lies in extracting pertinent features from high-dimensional data and employing classifiers to exploit those features [36], traditional DR methods and classification algorithms consider the same importance for all spectral bands (features), whereas the abilities of different features are different in identification of classes [8]. So, it is very useful to quantify the relative importance of each feature in identification of classes. Several works have been done to determine the relative importance of each feature in dimension reduction and classification process. In [37] a feature weighting method for band selection has been proposed which is based on the pairwise separability criterion and matrix coefficient analysis. It involves calculation of the criterion values for decorrelation of each class through principal component transformation, sorting of bands for each class, determination of final weights, and remove bands by a correlation threshold. Two approaches have been introduced in [38] to assign weights to support vector machine (SVM) features in order to improve the classification accuracy. One is SVM with compactness and a separation coefficient feature weighting algorithm, and the other is SVM with a similarity entropy feature weighting algorithm. As a weighted feature extraction method, FEWT algorithm in [8] determines the relative importance of different bands in feature extraction process. But, FEWT is a supervised approach; hence extending the weighted feature space to unlabeled and unseen samples (also called out-of-samples) cannot be done with FEWT, because the labels of unlabeled and unseen samples are unknown before the classification.

In this paper, we focus on the construction of a relative weighted feature space (RWFS) as a data pre-processing procedure. In the first step, we determine weighted features for available labeled samples. In order to overcome the difficulties of limited labeled samples as a common case in HSIs data analysis, we extend our procedure to semisupervised framework. To transfer the knowledge from available labeled samples to unlabeled/unseen data samples, it is necessary to establish some relationship between them. In this paper we utilize graph based clustering and apply Low-Rank Representation (LRR) to define the weight matrix for graph. After construction the RWFS, any dimension reduction method and classification algorithm can be used.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

The organization of this paper is as follows: section 2 describes graph based clustering and the Low Rank Representation (LRR). We introduce the RWFS in section 3. The description of data and the experimental results are given in section 4. Finally, section 5 concludes the paper.

2. Graph Based Clustering and LRR

Generally, in order to transfer the knowledge from available labeled samples to unlabeled data samples; it is necessary to establish some relationship between them. Different approaches can be used for this purpose such as LLE [39] and graph-based methods [40]. Graph-based methods have attracted much attention and are widely used in SSL framework because of their good results in many areas. The performance of semisupervised graph-based method highly depended upon the construction of a graph representation which can reveal data structure [10,41,42]. So, constructing a good graph to discover the intrinsic structures of the data is critical for SSL tasks [43].

The vertices of semisupervised graph-based methods are the samples (labeled and unlabeled) and edges represent the similarity among samples [40]. Two widely used criteria in graph based SSL are ϵ-neighborhood and k-nearset neighbor (kNN) [44]. The graph based on ϵ-neighborhood is often sensitive to the selection of parameter ϵ. Although kNN graph performs better than ϵ-neighborhood graph [45], it may be hard to tune the proper number of neighbors [46]. After construction the graph, a weighted edge between connected nodes is set to model their similarity. The weight (similarity) matrix can be obtained using RBF kernel which is a common choice [40] or can be represented as the linear combination of the neighboring vicinity and the combination coefficients [1]. Both of them require setting their parameters while the proper parameter setting still remains an open issue for these applications, and there is no known effective method.

Intuitively, nearby samples are possibly from the same subspace and thus the nearby samples in graph should have similar label. In this paper, we want to utilize the idea of spectral clustering method [47-49] as a relaxation of graph partitioning problems to transfer knowledge from labeled samples to unlabeled samples. Spectral clustering belongs to unsupervised learning algorithms which have been used with success in the field of computer vision for data clustering [50,51]. It has been shown that SC outperforms classical methods such as k-means and hierarchical clustering. Nevertheless, it suffers from several problems which they are still not solved satisfactory. One of these problems appears when an affinity matrix is constructed by the similarity of each pair of samples. Since the Gaussian similarity (RBF kernel) is a popular choice as similarity function, appropriate setting of σ is crucial for obtaining good clustering results [50].

The low rank representation [42,52,53] is a powerful representation tool and a promising weighted graph construction method. LRR is capable of representing the structural information of the union of multiple independent subspaces [42]. Also, it is possible to get the lowest rank representation in a parameter free way, which is very convenient and robust for kinds of data [54]. The graph identified by LRR provides the representation of all the data under a global low-rank constraint, and thus is better at capturing the global structures of data, such as multiple clusters and subspaces [42]. Because of the efficacy of LRR, we want to use it to define the similarity function for constructed graph. In the following, we briefly describe LRR.

2.1. LRR

Given $\mathbf{O} = \{o_i\}_{i=1}^{n} \in \mathbb{R}^D$ where o_i is a spectral vector of ith sample, and R denotes the number of spectral bands (features). In LRR, the objective function is as follows [10] (Equation (1)):

$$\min rank(\mathbf{Z}) + \gamma\|\mathbf{E}\|_{2,0}, s.t. \mathbf{O} = \mathbf{AZ} + \mathbf{E} \quad (1)$$

Where $\mathbf{A}$ is a dictionary that can span the data subspaces. $\|\mathbf{E}\|_{2,0}$ denoted the $\ell_{2,0}$-norm of the outliers and noises $\mathbf{E}$, and γ is a balance parameter, and $\mathbf{Z}$ denotes the coefficient matrix. Since Equation (1) is nonconvex, the rank constraint and the $\ell_{2,0}$-norm are replaced, respectively, by the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

nuclear norm ($\ell_* - norm$) and the $\ell_{2,1}$-norm. Finally, LRR proposes to solve the convex optimization problem in Equation (2):

$$\min_{Z,E} \|Z\|_* + \gamma \|E\|_{2,1}, s.t. \mathbf{O} = \mathbf{AZ} + \mathbf{E} \quad (2)$$

The Augmented Lagrange Multiplier (ALM) [55] can be used to solve Equation (2).

3. RWFS Procedure

Let $\mathbf{O} = \{\mathbf{O_L}, \mathbf{O_U}\} = \{o_1, o_2, \dots, o_l, o_{l+1}, \dots, o_{l+u}\} \in \mathbb{R}^{D \times (l+u)}$, $(n = l + u)$ be the data set in the original feature space (OFS) where the first l and the remaining u columns are the labeled and unlabeled samples, respectively, $\mathbf{F}_\mathrm{L} \in \mathbb{R}^{C \times l}$ is a binary label matrix associated with the samples with $\mathbf{F}_\mathrm{L}(i,j) = 1$ if x_i has label $y_i = j$; $\mathbf{F}_\mathrm{L}(i,j) = 0$, otherwise, R denotes the numbers of original features (dimensions) and C is the number of classes.

In order to quantify the importance of each feature of training labeled samples, we can fix a linear model $f_i = \boldsymbol{V}^T o_i + B^T$ by regressing $\mathbf{O}$ on $\mathbf{F}$ and simultaneously preserve the manifold smoothness in the embedding of both the label and unlabeled set, where $\mathbf{V} \in \mathbb{R}^{\mathrm{R} \times \mathrm{C}}$ is the weighting matrix and $\mathbf{B} \in \mathbb{R}^{l \times C}$ is the bias term. Therefore, we can write the following objective function (Equation (3))

$$\mathrm{G}(\mathbf{V}, \mathbf{B}) = min \sum_{i=1}^{l} \|\mathbf{V}^\mathrm{T} o_i + \mathbf{B}^\mathrm{T} - f_i\|_F^2 + \mu_V \|\mathbf{V}\|_F^2 + \mu_T trace(\mathbf{V^T OLO^T V}) \quad (3)$$

Where $\mathbf{L} = \mathbf{D} - \mathbf{W}^*$ is the graph Laplacian matrix associated with both labeled and unlabeled set, W^* is the weight matrix defined as: $w_{ij}^* = \exp(-\frac{\|o_i - o_j\|^2}{2\sigma^2})$, if o_i is within the k nearest neighbors of o_j or o_j is within the k nearest neighbors of o_i; $w_{ij}^* = 0$, otherwise, σ is the radius parameter, $\mathbf{D} = [d_{ij}]$ is a diagonal matrix with $d_{ii} = \sum_j w_{ij}^*$, the two parameters μ_V and μ_T balance the trade-off between manifold and Tikhonov regularization terms. The closed form solution of Equation (3) is given by Equation (4) [56].

$$\mathbf{V}^* = (\boldsymbol{OH_t O^T} + \mu_V \boldsymbol{I} + \mu_T \boldsymbol{OLO^T})^{-1} \boldsymbol{OH_t F^T} \quad (4)$$

Where $\boldsymbol{H_t} = \frac{Ue^T eU}{eUe^T}$ is used for centring the labeled samples, $e \in \mathbb{R}^{1 \times (l+u)}$ is a unit vector with size $l + u$, $\boldsymbol{U} \in \mathbb{R}^{(l+u) \times (l+u)}$ is a diagonal matrix with the first l and the remaining u diagonal elements as 1 and 0, respectively.

With using the calculated $\mathbf{V}$, we want to determine the weight of the labeled samples in OFS. Each column of $\mathbf{V}$ represents the weight of specific class corresponding to that column. In other words, for all labeled samples with the same class label, the same weight column is used to weight their features such a way that each component of that column is assigned to its corresponding feature as follows: if sample o_i belongs to the class C, then Cth column of matrix $\mathbf{V}$ is used to weight it by Equation (5):

$$p_{ij} = o_{ij} v_{jC}, 1 \le j \le R \quad (5)$$

Where o_{ij} is the jth feature of ith sample, v_{jC} is the weight of jth feature in class C, and p_{ij} is jth feature of ith weighted sample in the new weighted feature space. The obtained weighting matrix $\mathbf{V}$ cannot be used for unlabeled samples, because the labels of them are unknown before the classification, and thus, it is unknown which column of $\mathbf{V}$ should be used for weighting the unlabeled data samples. Now we want to transfer the knowledge from training labeled samples to unlabeled data sample.

We use graph based clustering method for this purpose. We cluster the all data samples $\mathbf{O} = \{\mathbf{O_L}, \mathbf{O_U}\}$ by the following steps. As the first step, we obtain the low rank representation of $\mathbf{O}$ by solving Equation (2) while the symmetry constraint $Z = Z^T$ is conducted on it. Then, we construct an undirected graph by low rank representation where Z, as the optimal solution of Equation (2), defines the affinity matrix of graph. It is mentioned that the dictionary $\mathbf{A}$ is always unknown; therefore we use $\mathbf{O}$ itself as dictionary. Finally, to obtain final clustering results, Normalized Cuts [57] is used where the number of clusters is equal to the number of classes. But, clustering methods

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

cannot predict the class label of each cluster. We assign the label of the maximum number of labeled samples in each cluster as the label of unlabeled samples in that cluster. Then, for all unlabeled samples in each cluster, the same weight column in Equation (4) corresponding to their class label is used to weight their features using Equation (5). Finally, when the RWFS is obtained, every DR method can be applied as the last step. The work flow for construction the RWFS is summarized in Table 1.

Table 1 Proposed procedure for constructing the RWFS

Algorithm 1: Relative Weighted Feature Space (RWFS) procedure	
Input	l labeled samples in original feature space (OFS) $\boldsymbol{O_L} = \{(\boldsymbol{o_i}, \boldsymbol{f_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{l}}$, u unlabeled samples in OFS, $\boldsymbol{O_U} = \{o_j\}_{\boldsymbol{j=l+1}}^{\boldsymbol{l+u}}$, $\boldsymbol{O} = \boldsymbol{O_L} \cup \boldsymbol{O_U}$, σ, μ_V and μ_T Find weighting matrix for labeled samples using Equation (4). Obtain low rank representation of $\boldsymbol{O}$ by solving Equation (2). Construct an undirected graph where LRR defines the affinity matrix of the graph. Use NCut to cluster the nodes of the graph into C clusters. Assign label to unlabeled samples as described in section 3. Weight the features of labeled and unlabeled using Equation (5).
Output	Relative Weighted Feature Space $\mathbf{P} \in \mathbb{R}^{D\times(l+u)}$

4. Data Sets and Experimental Results

We evaluate the final RWFS with two popular HSI data sets. The first data set is the Indian Pines which was collected over the northwestern Indiana by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in June 1992. It contains 145 pixels*145 pixels and 16 classes and comprises 224 spectral bands. In our experiment, we select 200 bands after removing 24 noisy and water absorption bands [1]. Table 2 represents the name and the number of samples in the Indiana Pines data set used in our experiments.

The second is the University of Pavia data set which is a Reflective Optic System Imaging Spectrometer (ROSIS) image from university of Pavia, Italy. It contains of 610 pixels*340 pixels and after removing 12 bands due to noise and water absorption, 103 spectral bands with 9 different classes are used in our experiments [1]. The name and the number of samples in University of Pavia data set is shown in Table 3.

Throughout our experiments, we used two conventional and widely used classifiers, 1-nearest neighbor (1-NN) and support vector machines (SVMs). The SVM classifier with radial basis function (RBF) kernel is used which has two parameters, the penalty factor C and the RBF kernel widths σ. We optimized C within the given set $\{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$, with step size increment 20, and σ within the given set $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ by fivefold cross validation approach. The one-against-one (OAO) strategy for multiclass classification algorithm are used for SVM.

In order to evaluate the new weighted feature space, we selected a number of algorithms used for hyperspectral images analysis. First of all, as the traditional algorithms, LDA [13], NWFE [15] and PCA [11] were selected. LPP [16] was also selected whose idea is used for incorporating the local geometry in the regularization term in most semisupervised frameworks. In addition, we evaluated RWFS on three semisupervised algorithms proposed for feature extraction of hyperspectral images were: SEGL [22], SELD [23] and SLDA [24]. We tested several other algorithms but their performance in RWFS was less than their performance in OFS such as GDA. The binary similarity matrix was used in feature extraction process. It observed that the results of Equation (3).were relatively stable with regard to the variation of k. Thus, k set to 8. The parameter σ for determination of weighting matrix in Equation (3) set to be equal to the average distance of a sample to its 7th nearest neighbors. The appropriate values for μ_V and μ_T determined by experiments. With

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

searching over a large range of nonnegative values for these parameters, we found that a proper value for them can be selected from $[10^3 - 10^9]$ and [0-1], respectively. $\mu_V = 10^6$, and $\mu_T = 10^{-2}$ has been selected for both data set.

Table 2 The name and the number of samples in the Indiana Pines data set

Class	Name	Sample	Class	Name	Sample
C1	Alfalfa	46	C9	Oats	20
C2	Corn-no till	1428	C10	Soybean-no till	972
C3	Corn-mint ill	830	C11	Soybeans-mint ill	2455
C4	Corn	237	C12	Soybeans-clean	593
C5	Grass-pasture	483	C13	Wheat	205
C6	Grass-pasture	730	C14	Woods	1265
C7	Grass-pasta removed	28	C15	Bldg-grass-tree-drives	386
C8	Hay-windrow wed	478	C16	Stone-steel-towers	93

Table 3 The name and the number of samples in University of Pavia data set

Classes	Name	Sample	Class	Name	Sample
C1	Corn-no till	1434	C6	Soybeans-no till	968
C2	Corn-min	834	C7	Soybeans-clean	2468
C3	Grass/pasture	497	C8	Woods	614
C4	Grass/tree	747	C9	Soybeans-min	1294
C5	Hay-windrowed	489	C10	Bldg-grass tree	380

We present the overall classification accuracy measure and average accuracy measure and kappa coefficient to evaluate the benefits of RWFS using selected DR methods. Because of random selection of labeled and unlabeled training samples, each experiment repeated 30 times, and the recorded results averaged over 30 runs for each classifier.

In the RWFS, a relative weight is assigned to each feature of a sample which it determines the relative importance of that feature in predicting the class of the sample. Similar to what has been done in [8], we use the relative weighted features space just in dimension reduction process and the original feature space is used in the classification. The obtained results of the different methods with respect to the reduced dimension numbers for Indian Pine data set using a) ML; b) SVM classifiers; 10 labeled samples and 800 unlabeled samples are represented in Figure 1. The results of selected methods in OFS and RWFS were specified by solid and dashed lines, respectively. It is shown that the performance of selected algorithms have been improved in RWFS in comparison with OFS.

In order to investigate the sensitivity of the RWFS to labeled samples, we repeat the above experiments for Indiana Pine data set with 20 labeled samples which is shown in Figure 2. As shown in Figure 2, the performance of selected algorithms has been improved. Also, Figure 3 shows the average classification accuracy versus the number of extracted features for University of Pavia data set using: a) ML classifier and b) SVM classifier with 10 labeled samples and 1500 unlabeled samples using four selected dimension reduction methods. Also, in this case the solid lines represent the results in OFS and dashed lines represent the results in RWFS. The obtained results show that the performance of DR methods has improved in RWFS in comparison with OFS.

Table 4 represents the results of all selected methods for Indian Pines data set using SVM classifier, 25 labeled samples, 800 unlabeled samples and 8 extracted features. Comparing the obtained results of all the methods in RWFS and original feature space (OFS), it can be observed that the results in RWFS are better than OFS. For example, it can be observed that the SELD-LPP, SLDA,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

SEGL and LPP methods yield 2.8%, 3.3%, 2% and 1.5% higher overall accuracy in RWFS than in OFS for Indiana Pines data set using SVM classifier.

Table 4 The obtained classification results for Indian Pines data set by SVM classifier, 25 labeled samples, 800 unlabeled samples and 8 extracted features.

CLASS	SELD-*LPP*		SDLA		SEGL		LDA	NWFE	PCA	LPP	
	OFS.	RWFS	OFS	RWFS	OFS	RWFS	OFS	OFS	OFS	OFS	RWFS
C1	93.21	94.53	92.53	94.13	94.64	95.32	53.87	93.21	70.46	84.96	98.73
C2	62.03	62.33	62.43	62.29	63.67	64.08	21.34	60.43	48.56	51.24	69.95
C3	47.46	52.54	46.66	51.12	47.86	51.45	22.87	44.96	39.64	42.34	56.32
C4	65.45	65.15	64.06	65.00	64.96	66.54	20.01	67.86	60.43	72.87	61.23
C5	43.55	45.85	43.98	44.15	45.73	46.83	32.76	42.75	40.32	68.12	74.85
C6	59.04	60.64	59.46	61.44	60.35	62.74	44.97	58.94	55.32	62.12	69.87
C7	94.86	95.06	95.06	94.86	95.35	96.05	77.32	95.86	79.54	96.43	97.12
C8	85.43	86.84	86.25	87.80	86.63	88.90	41.65	83.03	56.76	66.76	75.01
C9	100.0	100.0	100.0	100.0	100.0	100.0	79.43	100.0	84.3	100.0	100.0
C10	49.16	55.64	48.87	56.60	49.94	57.98	21.45	48.96	37.54	51.05	56.23
C11	48.73	49.47	48.84	49.23	49.53	51.96	23.74	46.63	23.4	35.32	41.32
C12	56.04	59.85	55.93	58.89	57.84	59.90	24.41	54.96	28.53	51.75	56.45
C13	83.16	85.06	83.56	86.48	84.14	87.89	47.43	82.46	62.54	81.23	84.54
C14	77.72	79.95	78.03	80.12	78.83	81.07	52.74	76.54	69.54	86.79	80.34
C15	28.82	35.94	29.02	34.47	30.25	35.75	32.18	27.85	20.43	34.43	38.32
C16	94.38	96.03	93.25	95.67	95.63	97.07	81.65	93.54	81.32	93.86	94.21
Average Acc (%)	**68.05**	**70.30**	**67.99**	**70.14**	**69.08**	**71.47**	**42.39**	**66.38**	**53.68**	**65.48**	**67.21**
Overall Acc (%)	**79.45**	**81.68**	**79.78**	**82.45**	**80.75**	**82.37**	**52.34**	**75.63**	**59.87**	**75.87**	**77.06**
Kappa coefficient	**0.55**	**0.67**	**0.56**	**0.72**	**0.57**	**0.73**	**0.25**	**0.52**	**0.42**	**0.46**	**0.56**

Table 5 represents the results of selected dimension reduction methods for University of Pavia data set using ML classifier, 10 labeled samples, 1500 unlabeled samples and 8 extracted features. Comparing the obtained results of all the methods in RWFS and OFS, it can be observed that the results in RWFS are better than OFS. For example, it can be observed that the SELD-LPP, SLDA, SEGL and LPP methods yield 1.4%, 1.7%, 3.6% and 1.6% higher overall accuracy in RWFS than in OFS for University of Pavia data set using ML classifier.

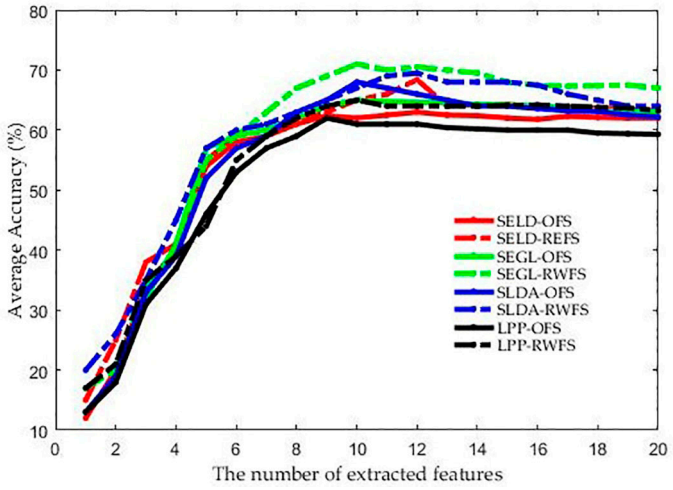
The highest classification accuracies achieved by 15 training labeled samples are represented in Table 6. The numbers in the parentheses show the number of features, which obtain the highest average accuracies in the experiments.

Generally, these results confirm the conclusion that RWFS is better than OFS for use in some traditional dimensionality reduction and classification algorithms for HSI. Although using of RWFS increases the performance of selected algorithms, the experiments showed that the use of RWFS may degrade the classification accuracy of several other feature extraction methods such as GDA. We ignored from these methods and only selected several algorithms that their performances increased.

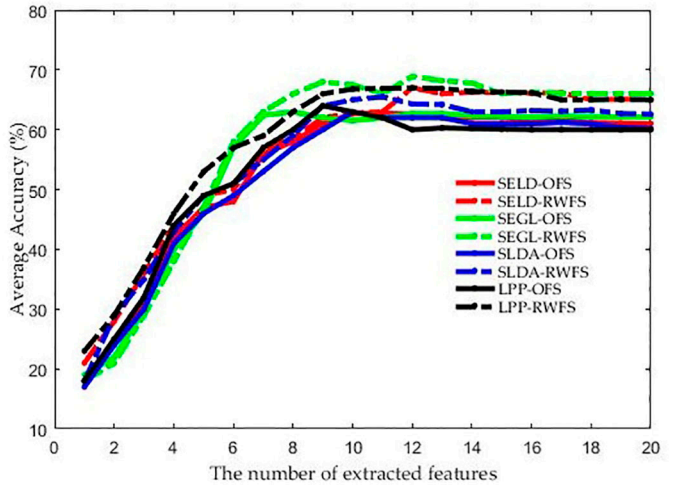
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

272 **Table 5** The obtained classification results for University of Pavia data set by ML classifier, 40 labeled
273 samples, 1500 unlabeled samples and 8 extracted features.

CLASS	SELD-LPP		SDLA		SEGL		PCA	LDA	NWFE	LPP	
	OFS	RWFS	OFS	RWFS	OFS	RWFS	OFS	OFS	OFS	OFS	RWFS
C1	65.27	67.04	66.07	68.34	65.27	68.38	60.54	59.23	64.34	62.21	67.56
C2	64.73	65.87	64.43	66.83	64.73	66.03	59.39	55.43	64.13	63.32	67.06
C3	81.34	83.25	82.04	83.95	81.34	84.05	74.43	70.12	80.34	79.03	82.67
C4	81.99	83.07	82.64	83.84	81.99	84.53	77.23	73.48	81.84	79.11	84.67
C5	94.34	96.64	93.74	97.14	94.34	97.38	81.47	80.32	93.34	91.02	95.12
C6	70.11	69.94	71.09	71.35	70.11	70.43	60.83	54.65	69.51	69.11	73.34
C7	76.93	77.24	77.27	78.13	76.93	78.49	69.73	64.43	75.83	74.23	75.35
C8	55.96	63.56	55.78	65.37	55.96	64.73	52.04	50.19	54.94	53.54	60.38
C9	89.37	90.34	90.09	91.53	89.37	91.26	81.52	79.32	88.45	86.43	91.94
Average Acc (%)	**75.56**	**77.43**	**75.90**	**78.51**	**75.27**	**78.32**	**68.12**	**65.32**	**74.70**	**70.22**	**74.57**
Overall Acc (%)	**79.98**	**81.12**	**81.36**	**82.76**	**79.73**	**82.63**	**72.43**	**69.43**	**79.88**	**72.39**	**77.23**
Kappa coefficient	**0.76**	**0.79**	**0.76**	**0.81**	**0.77**	**0.80**	**0.67**	**0.58**	**0.75**	**0.71**	**0.73**



(a)



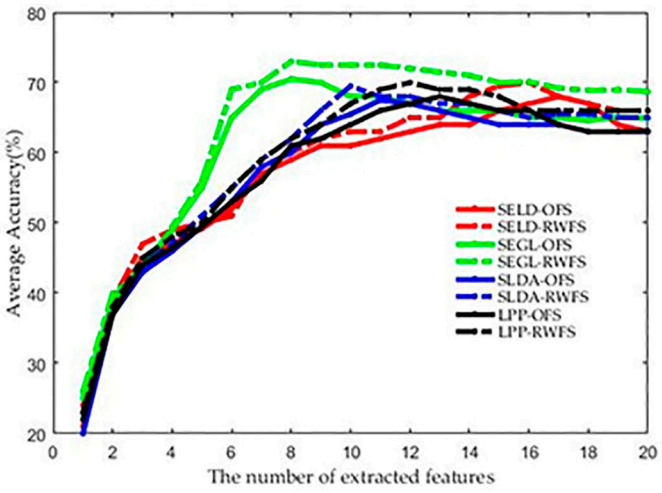
(b)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

Figure 1 The average classification accuracy versus the number of extracted features for Indian Pines data set obtained by **a) ML** classifier**; b) SVM** classifier with 10 labeled samples, 800 unlabeled samples; dashed lines belong to RWFS and solid lines belong to OFS

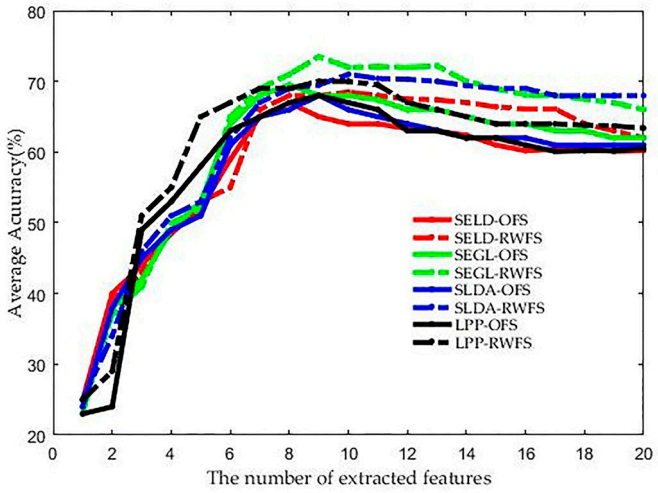
Table 6 Highest average classification accuracies obtained by SVM and ML classification and 15 training samples for two data sets

Dataset	Classifier	SELD-LPP		SDLA		SEGL		LDA	PCA	NWFE	LPP	
		OFS	RWFS	OFS	RWFS	OFS	RWFS	OFS	OFS	OFS	OFS	RWFS
Indian Pines	**SVM**	70.67 (12)	75.60 (11)	71.04 (16)	78.75 (15)	71.11 (11)	78.46 (11)	45.32 (12)	57.32 (6)	69.04 (6)	68.84 (6)	78.65 (7)
	ML	72.69 (11)	74.86 (10)	73.45 (9)	75.94 (9)	72.89 (11)	75.89 (11)	33.54 (5)	58.32 (19)	72.04 (4)	69.87 (5)	71.37 (5)
University of Pavia	**SVM**	79.49 (8)	81.34 (9)	79.43 (7)	81.46 (7)	80.13 (9)	82.97 (10)	49.05 (5)	69.95 (8)	78.34 (9)	78.93 (14)	79.96 (13)
	ML	79.74 (7)	83.02 (7)	79.47 (8)	81.45 (8)	81.21 (9)	83.45 (9)	46.43 (5)	66.74 (11)	78.48 (5)	78.56 (6)	80.85 (6)



(a)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1

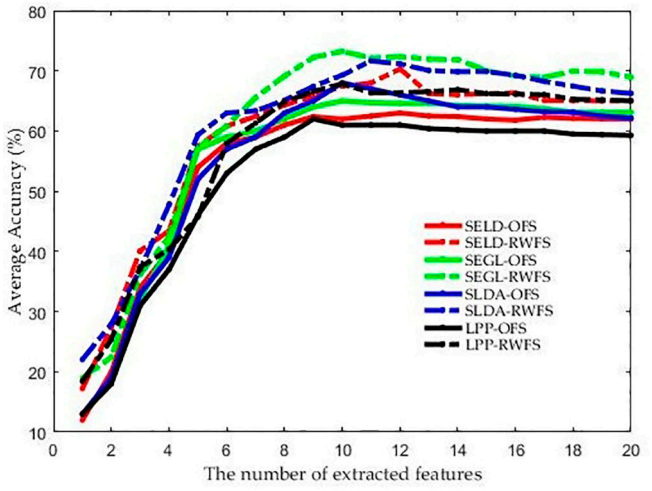


(b)

Figure 2 The average classification accuracy versus the number of extracted features for Indian Pines data set obtained by **a) ML** classifier**; b) SVM** classifier with 20 labeled samples, 800 unlabeled samples; dashed lines belong to RWFS and solid lines belong to OFS

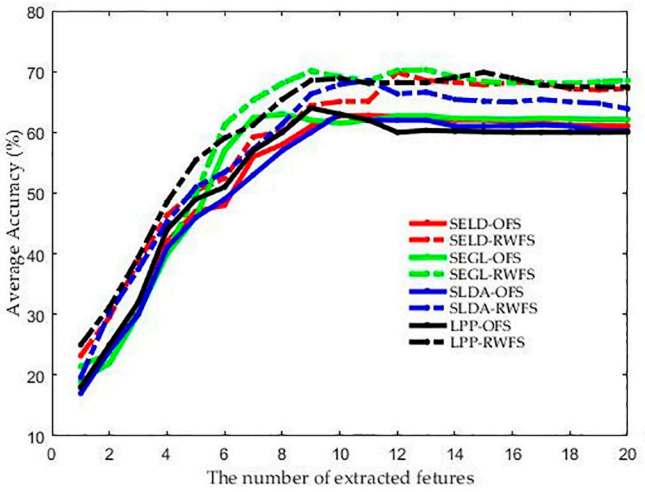
5. Conclusions

It is clear that all spectral bands of a hyperspectral image are not equally important and do not have the same discrimination ability in identification of class labels. In this paper, we proposed a method to determine a relative weighted feature space where every feature has a relative weight in identification of class label. We proposed a method to extend the available knowledge from training labeled samples to unlabeled samples using graph based clustering where low rank representation as a powerful representation tool is used to define the affinity matrix for graph. Any DR method and classification algorithm can be used in RWFS. In fact, the proposed procedure can be applied as a preprocessing phase in any feature extraction method and classification algorithm. The experimental results indicate the ability of weighted feature space in dimension reduction methods in comparison with original feature space. But, our experiments showed that the relative weighted feature space is not suitable for all dimension reduction methods.



(a)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0057.v1



(b)

Figure 3 The average classification accuracy versus the number of extracted features for University of Pavia data set obtained by **a)** ***ML*** *classifier***; b)** ***SVM*** *classifier* with 10 labeled samples, 1500 unlabeled samples; dashed lines belong to RWFS and solid lines belong to OFS

References

1. Dong, Y.; Du, B.; Zhang, L.; Zhang, L. Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 2509-2524.
2. Binol, H.; Bilgin, G.; Dinc, S.; Bal, A. Kernel fukunaga-koontz transform subspaces for classification of hyperspectral images with small sample sizes. *IEEE Geoscience and Remote Sensing Letters* **2015**, *12*, 1287-1291.
3. Dollar, P.; Zhuowen, T.; Hai, T.; Belongie, S. In *Feature mining for image classification*, Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 17-22 June 2007, 2007; pp 1-8.
4. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *Geoscience and Remote Sensing, IEEE Transactions on* **2010**, *48*, 4085-4098.
5. Falco, N.; Benediktsson, J.A.; Bruzzone, L. Spectral and spatial classification of hyperspectral images based on ica and reduced morphological attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing* **2015**, *53*, 6223-6240.
6. Rui, H.; Mingyi, H. Band selection based on feature weighting for classification of hyperspectral data. *Geoscience and Remote Sensing Letters, IEEE* **2005**, *2*, 156-159.
7. Kavzoglu, T.; Mather, P.M. The role of feature selection in artificial neural network applications. *International Journal of Remote Sensing* **2002**, *23*, 2919-2937.
8. Imani, M.; Ghassemian, H. Feature extraction using weighted training samples. *Geoscience and Remote Sensing Letters, IEEE* **2015**, *12*, 1387-1391.
9. Junshi, X.; Chanussot, J.; Peijun, D.; Xiyan, H. (semi-) supervised probabilistic principal component analysis for hyperspectral remote sensing image classification. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **2014**, *7*, 2224-2236.
10. Wang, M.; Yu, J.; Niu, L.; Sun, W. Feature extraction for hyperspectral images using low-rank representation with neighborhood preserving regularization. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 836-840.
11. Schott, J.R. *Remote sensing: The image chain approach*. Oxford University Press: 2007.
12. Tang, F.; Xu, H. Impervious surface information extraction based on hyperspectral remote sensing imagery. *Remote Sensing* **2017**, *9*, 550.
13. Fukunaga, K. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc.: 1990; p 592.
14. Wang, F.; Wang, X.; Zhang, D.; Zhang, C.; Li, T. Marginface: A novel face recognition method by average neighborhood margin maximization. *Pattern Recogn.* **2009**, *42*, 2863-2875.

15. Bor-Chen, K.; Landgrebe, D.A. Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing* **2004**, *42*, 1096-1105.
16. He, X. Locality preserving projections. University of Chicago, 2005.
17. Aronszajn, N. Theory of reproducing kernels. *Transactions of the American Mathematical Society* **1950**, *68*.
18. Schölkopf, B.; Smola, A.; Müller, K.-R. Kernel principal component analysis. In *Artificial neural networks — icann'97: 7th international conference lausanne, switzerland, october 8–10, 1997 proceeedings*, Gerstner, W.; Germond, A.; Hasler, M.; Nicoud, J.-D., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1997; pp 583-588.
19. Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B.; Mullers, K.R. In *Fisher discriminant analysis with kernels*, Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468), Aug 1999, 1999; pp 41-48.
20. Li, J.-B.; Pan, J.-S.; Chu, S.-C. Kernel class-wise locality preserving projection. *Inf. Sci.* **2008**, *178*, 1825-1835.
21. Chen, Y.-N.; Hsieh, C.-T.; Wen, M.-G.; Han, C.-C.; Fan, K.-C. A dimension reduction framework for hsi classification using fuzzy and kernel nfle transformation. *Remote Sensing* **2015**, *7*, 14292.
22. Luo, R.; Liao, W.; Huang, X.; Pi, Y.; Philips, W. Feature extraction of hyperspectral images with semisupervised graph learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2016**, *9*, 4389-4399.
23. Liao, W.; Pizurica, A.; Scheunders, P.; Philips, W.; Pi, Y. Semisupervised local discriminant analysis for feature extraction in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51*, 184-198.
24. Shi, Q.; Zhang, L.; Du, B. Semisupervised discriminative locally enhanced alignment for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51*, 4800-4815.
25. Bor-Chen, K.; Chun-Hao, C.; Tian-Wei, S.; Chih-Cheng, H. In *Feature extractions using labeled and unlabeled data*, Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05., 25-29 July 2005, 2005; pp 1257-1260.
26. Bruzzone, L.; Chi, M.; Marconcini, M. A novel transductive svm for semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing* **2006**, *44*, 3363-3373.
27. Zhang, Z.; Pasolli, E.; Crawford, M.M.; Tilton, J.C. An active learning framework for hyperspectral image classification using hierarchical segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2016**, *9*, 640-654.
28. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing* **2010**, *48*, 4085-4098.
29. Cao, Y.; Xu, L.; Clausi, D. Exploring the potential of active learning for automatic identification of marine oil spills using 10-year (2004–2013) radarsat data. *Remote Sensing* **2017**, *9*, 1041.
30. Crawford, M.M.; Tuia, D.; Yang, H.L. Active learning: Any value for classification of remotely sensed data? *Proceedings of the IEEE* **2013**, *101*, 593-608.
31. Tuia, D.; Volpi, M.; Copa, L.; Kanevski, M.; Munoz-Mari, J. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing* **2011**, *5*, 606-617.
32. Jackson, Q.; Landgrebe, D.A. An adaptive method for combined covariance estimation and classification. *IEEE Transactions on Geoscience and Remote Sensing* **2002**, *40*, 1082-1087.
33. Zhou, D.; Bousquet, O.; Lal, T.N.; Weston, J.; Sch, B.; #246; lkopf. Learning with local and global consistency. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, MIT Press: Whistler, British Columbia, Canada, 2003; pp 321-328.
34. Sun, S.; Zhong, P.; Xiao, H.; Wang, R. Active learning with gaussian process classifier for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2015**, *53*, 1746-1760.
35. Persello, C. Interactive domain adaptation for the classification of remote sensing images using active learning. *IEEE Geoscience and Remote Sensing Letters* **2013**, *10*, 736-740.
36. Samiappan, S.; Prasad, S.; Bruce, L.M. Non-uniform random feature selection and kernel density scoring with svm based ensemble classification for hyperspectral image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2013**, *6*, 792-800.
37. Rui, H.; Mingyi, H. Band selection based on feature weighting for classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters* **2005**, *2*, 156-159.
38. Qi, B.; Zhao, C.; Yin, G. Feature weighting algorithms for classification of hyperspectral images using a support vector machine. *Appl. Opt.* **2014**, *53*, 2839-2846.
39. Chen, J.; Liu, Y. Locally linear embedding: A survey. *Artif. Intell. Rev.* **2011**, *36*, 29-48.

40. Camps-Valls, G.; Marsheva, T.V.B.; Zhou, D. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2007**, *45*, 3044-3054.
41. Feng, Z.; Yang, S.; Wang, S.; Jiao, L. Discriminative spectral–spatial margin-based semisupervised dimensionality reduction of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters* **2015**, *12*, 224-228.
42. Fei, L.; xu, Y.; Fang, X.; Yang, J. *Low rank representation with adaptive distance penalty for semi-supervised subspace classification*. 2017; Vol. 67.
43. Nie, F.; Wang, X.; Huang, H. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM: New York, New York, USA, 2014; pp 977-986.
44. Zhuang, L.; Zhou, Z.; Gao, S.; Yin, J.; Lin, Z.; Ma, Y. Label information guided graph construction for semi-supervised learning. *IEEE Transactions on Image Processing* **2017**, *26*, 4182-4192.
45. Jebara, T.; Wang, J.; Chang, S.-F. Graph construction and b-matching for semi-supervised learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM: Montreal, Quebec, Canada, 2009; pp 441-448.
46. Dong, W.; Moses, C.; Li, K. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th international conference on World wide web*, ACM: Hyderabad, India, 2011; pp 577-586.
47. Jianbo, S.; Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2000**, *22*, 888-905.
48. Fowlkes, C.; Belongie, S.; Chung, F.; Malik, J. Spectral grouping using the nystrom method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2004**, *26*, 214-225.
49. Nie, F.; Zeng, Z.; Tsang, I.W.; Xu, D.; Zhang, C. Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering. *IEEE Transactions on Neural Networks* **2011**, *22*, 1796-1808.
50. Alzate, C.; Suykens, J.A.K. Multiway spectral clustering with out-of-sample extensions through weighted kernel pca. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2010**, *32*, 335-347.
51. Zhang, X.; Jiao, L.; Liu, F.; Bo, L.; Gong, M. Spectral clustering ensemble applied to sar image segmentation. *IEEE Transactions on Geoscience and Remote Sensing* **2008**, *46*, 2126-2136.
52. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2013**, *35*, 171-184.
53. Li, B.; Liu, R.; Cao, J.; Zhang, J.; Lai, Y.K.; Liu, X. Online low-rank representation learning for joint multi-subspace recovery and clustering. *IEEE Transactions on Image Processing* **2018**, *27*, 335-348.
54. Zu, B.; Xia, K.; Dai, S.; Aslam, N. Low-rank kernel-based semisupervised discriminant analysis. *Applied Computational Intelligence and Soft Computing* **2016**, *2016*, 9.
55. Lin, Z.; Chen, M.; Ma, Y. *The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices*. 2010; Vol. 9.
56. Zhao, M.; Zhang, H.; Zhang, Z. In *Learning from local and global discriminative information for semi-supervised dimensionality reduction*, The 2013 International Joint Conference on Neural Networks (IJCNN), 4-9 Aug. 2013, 2013; pp 1-8.
57. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888-905.