



Article Pairwise Elastic Net Representation-Based Classification for Hyperspectral Image Classification

Hao Li ¹, Yuanshu Zhang ², Yong Ma ^{2,*}, Xiaoguang Mei ², Shan Zeng ¹ and Yaqin Li ¹

- ¹ School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan 430023, China; lihao@whpu.edu.cn (H.L.); zengshan1981@whpu.edu.cn (S.Z.); leeyaqin@whpu.edu.cn (Y.L.)
- ² Electronic Information School, Wuhan University, Wuhan 430072, China; 2018202120089@whu.edu.cn (Y.Z.); meixiaoguang@whu.edu.cn (X.M.)
- * Correspondence: mayong@whu.edu.cn

Abstract: The representation-based algorithm has raised a great interest in hyperspectral image (HSI) classification. l_1 -minimization-based sparse representation (SR) attempts to select a few atoms and cannot fully reflect within-class information, while l_2 -minimization-based collaborative representation (CR) tries to use all of the atoms leading to mixed-class information. Considering the above problems, we propose the pairwise elastic net representation-based classification (PENRC) method. PENRC combines the l_1 -norm and l_2 -norm penalties and introduces a new penalty term, including a similar matrix between dictionary atoms. This similar matrix enables the automatic grouping selection of highly correlated data to estimate more robust weight coefficients for better classification performance. To reduce computation cost and further improve classification accuracy, we use part of the atoms as a local adaptive dictionary rather than the entire training atoms. Furthermore, we consider the neighbor information of each pixel and propose a joint pairwise elastic net representation-based classification (J-PENRC) method. Experimental results on chosen hyperspectral data sets confirm that our proposed algorithms outperform the other state-of-the-art algorithms.

Keywords: hyperspectral image (HSI) classification; sparse representation; collaborative representation; pairwise elastic net; neighbor information

1. Introduction

A hyperspectral image is a 3D remote sensing image containing hundreds of bands, from visible to infrared spectra. Due to their abundant spectral information, HSIs have become an actual application in the field of remote sensing, such as skin imaging [1], ground elements identifying [2] and mineral exploration [3]. To date, many classification algorithms for hyperspectral datasets have been proposed. Among the techniques, the support vector machine (SVM) [4], Gaussian mixture-model (GMM) [5] and the Gaussian maximum-likelihood classifier (MLC) [6] are all proved to be effective for solving HSI classification problem. The most concerning research methods in recent years can be roughly divided into two categories: representation-based algorithms and deep learningbased algorithms. On the one hand, in order to make full use of the spectral and spatial information of HSIs, some effective spectral-spatial feature extraction methods have been combined with sparse models to improve the characterization capability of models, such as [7–10]. On the other hand, since the deep convolutional neural network (CNN) with deep architecture has been proven to be very effective in using image features, this type of method using deep CNN for hyperspectral feature extraction has stimulated various studies [11–14].

This paper is mainly focused on the HSI classification algorithm based on representation learning. The classification principle of the method is to assume that each testing pixel can be reconstructed with labeled training pixels. Then, the abundance coefficients of the testing pixel can be obtained with the penalty of l_1 -norm or l_2 -norm, which is named



Citation: Li, H.; Zhang, Y.; Ma, Y; Mei, X; Zeng, S.; Li, Y. Pairwise Elastic Net Representation Based Classification for Hyperspectral Image Classification. *Entropy* **2021**, *23*, 956. https://doi.org/10.3390/ e23080956

Academic Editor: Amelia Carolina Sparavigna

Received: 16 June 2021 Accepted: 15 July 2021 Published: 26 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). sparse representation classification (SRC) [15] and collaborative representation classification (CRC) [16]. In [17], Chen et al. first introduced the sparsity model into hyperspectral classification and proposed the joint sparse representation classification (JSRC) method by incorporating the contextual information. In [18], considering that different atoms have different importance for the reconstruction process, Li et al. proposed the nearest regularized subspace (NRS) classifier with Tikhonov regularization. By wisely combing SRC and KNN, in Ref. [19] a class-dependent sparse representation classifier (cdSRC) was proposed. However, some research [18,20] shows that the collaboration of approximation enhances classification results rather than competition. Therefore, in Ref. [21], a joint within-class CRC was provided to solve the HSI classification tasks. In [22], the kernel version of CRC was further considered and the Kernel-based CRC (KCRC) was proposed. There are also some investigations dedicated to improving classification effectiveness. On the one hand, some focus on the more simple and robust dictionary to reduce computation costs. On the other hand, some take the neighborhood spatial information as an important factor in improving classification accuracy. In Ref. [23], the nonlocal joint collaborative representation (NJCRC) algorithm was proposed by utilizing a subdictionary whose atoms are obtained by the k-nearest neighbor (K-NN) with testing samples rather than the whole dictionary atoms. In [24], Fang et al. introduced the shape irregular neighbor region into the joint SRC model and proposed the shape adaptive joint sparse representation (SAJSRC).

It is worth noting that both the SRC-based algorithms and CRC-based algorithms have their limitations. In these representation-based classification models, the obtained abundance coefficients reflect the importance of each training sample for reconstruction. Accordingly, the primary concern of this type of method is the solution of the abundance coefficient. Ideally, the test pixels should be linearly represented by atoms from the same category. The nonzero terms of sparse coefficients should be located at the position of the corresponding class. For SRC, it tends to select as few atoms as possible. The too sparse property will lead to the deviation of the absolute reconstruction error, and the sparsity will be weakened when the number of training atoms sets is small. For CRC, it tends to select all the atoms for reconstruction, and the class discrimination will be weak when including mixed-class information. Intuitively, SRC and CRC should be balanced to achieve better classification performance is necessary.

To solve the above problem, in Ref. [25], the elastic net representation-based classification (ENRC) method was proposed. The elastic net originally raised in [26] encourages both sparsity and grouping by forming a convex combination of the CRC and SRC governed by a selectable parameter. Furthermore, the elastic net can yield a sparse estimate with more than n nonzero weights. Based on these advantages, the ENRC improves of HSI classification performance. However, the optimal balance factors are all obtained by traversing the manufactured parameter space. This makes the algorithm time-consuming and complex. Additionally, the pixelwise fusion algorithm cannot make full use of the spatial information of the HSI.

Fortunately, the recent literature [27] has pointed out that the pairwise elastic net (PEN) model using similarity measures between regressors can establish a local balance between SRC and CRC. It can achieve more flexible grouping than ENRC. Moreover, PEN allows the customization of the sparsity relationship between any two features. Hence, in this work, we propose the pairwise elastic net representation-based classification (PENRC) method to overcome the indigenous disadvantages of ENRC, SRC and CRC. It can automatically achieve the balance between l_1 -norm and l_2 -norm so that more robust weight coefficients can be estimated, and further realizing better between-class sparse and intraclass collaborative classification performance.

Specifically, the main contribution of the proposed PENRC can be briefly summarized as follows. First, considering the computation cost when using all the dictionaries, we adopt the KNN to select the labeled atoms, which are more similar to the testing pixel as an optimal sub-dictionary. Then, unlike the ENRC, which assigns only a single global tradeoff between sparsity and collaboration, we introduce a similar matrix about sub-dictionary atoms in penalties, resulting in the local sparsity and collaboration tradeoff and be more flexible than ENRC. After obtaining the abundance coefficients, we use the principle of minimum reconstruction error to decide the final label. We also provide a further extension of our algorithm by incorporating the neighbor information of each pixel.

In summary, it is expected that the abundance coefficients from PENRC reveal a more powerful discriminant ability, thereby outperforming the original SRC, CRC and ENRC.

The remaining parts of the paper are organized as follows: Section 2 briefly introduces the two classical SRC and CRC classifiers. Section 3 details the proposed PENRC mechanism. Section 4 gives the experimental results on chosen two datasets. Finally, Section 5 concludes this paper.

2. Related Works

Denoting a testing pixel as $y = [y_1, ..., y_B] \in R^{B \times 1}$ and the dictionary composed of training atoms with class order as $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_C] \in R^{B \times N}$, where *B* is the number of spectral bands, $N = \sum_{c=1}^{C} N_c$ is the training atoms number and *C* is the total number of categories. The sub-dictionary $\mathbf{X}_c \in R^{B \times N_c}$ is the set of training atoms in *c*-th class.

2.1. Sparse Representation for HSI Classification

The sparse model assumes that a testing pixel can be linearly approximated with few dictionary atoms suitably [15]. Then, for a testing pixel y, the purpose of SRC model is to obtain the corresponding abundance coefficients by minimizing the reconstruction error $||y - X\alpha^{SRC}||_2^2$ with the sparse constraint term $||\alpha^{SRC}||_1$. Mathematically, the object function can be represent as follows:

$$\hat{\boldsymbol{\alpha}}^{SRC} = \arg\min \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{SRC}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}^{SRC}\|_{1'}^{\prime}$$
(1)

where λ_1 is the balancing parameter. The weight vector $\boldsymbol{\alpha}^{SRC} \in \mathbb{R}^{N \times 1}$ is sparse and only have few nonzero terms. It can be obtained by solving Equation (1) with basis pursuit (BP) or basis pursuit denoising (BPDN) algorithms [28,29]. When l_2 -norm is directly used, Equation (1) can be solved by subspace pursuit (SP) and orthogonal matching pursuit (OMP) algorithms [30].

After obtaining the weight vector α^{SRC} , we can assign the final class label which corresponding the mimimum reconstruction error to the testing pixel:

$$class(\boldsymbol{y}) = \arg\min_{c=1,\dots,C} \|\boldsymbol{y} - \hat{\boldsymbol{y}}_{c}\|_{2}^{2}$$
$$= \arg\min_{c=1,\dots,C} \|\boldsymbol{y} - \boldsymbol{X}_{c} \hat{\boldsymbol{\alpha}}_{c}^{SRC} \|_{2}^{2},$$
(2)

where α_c^{SRC} is the subset of sparse vector α^{SRC} which belongs to *c*-th class.

2.2. Collaborative Representation for HSI Classification

1

Unlike SRC model, the CRC assumes that a testing pixel can be linearly combined with all the training set [21]. The CRC attempts to obtain abundance coefficients by minimizing the reconstruction error $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{CRC}\|_2^2$ with the term $\|\boldsymbol{\alpha}^{CRC}\|_2$. Thus, the CRC can be expressed as:

$$\hat{\boldsymbol{\alpha}}^{CRC} = \arg\min \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{CRC}\|_2^2 + \lambda_2 \|\boldsymbol{\alpha}^{CRC}\|_{2'}$$
(3)

where λ_2 balances the influence of the reconstruction error and constraint term. Equation (3) can be simply solved with a closed form. Assuming that the derivative of the above cost function and is zero, we can obtain the optimal value of α^{CRC} :

$$\mathbf{x}^{CRC} = \left(\mathbf{X}^T \mathbf{X} + \lambda_2 \mathbf{I}\right)^{-1} \mathbf{X}^T \boldsymbol{y},\tag{4}$$

where **I** is an identity matrix with the size of $N \times N$. After obtaining α^{CRC} , the final class label *c* of testing pixel can be determined with the minimm residual rule as introduced in last section.

For the above representation-based classification methods, training atoms tend to be "competitive" in SRC due to the sparse constraints. With l_2 -norm, all atoms participate in the representation process equally. Thus, CRC tends to be "cooperative". Researchers compared the performance of SRC with CRC in literature [21,22]. Moreover, the experiments showed that in some cases, SRC performances better than CRC while CRC performance was better in other cases. For example, in remote sensing images, the SRC algorithm gave rise to a more remarkable improvement with some mixed pixels [31]. Thus, it is an effective way to combine SRC and CRC appropriately. In fact, in Ref. [25], FRC and ENRC algorithms to combine SRC with CRC were proposed. However, the dictionary chosen in [25] consists of all the training samples and brings a large computational burden. In addition, the algorithms in [25] only set a global trade-off between SRC and CRC, leading to the inflexible balance of different classes.

3. Proposed PENRC

The framework of our proposed PENRC algorithm is shown in Algorithm 1. First, we built a local adaptive dictionary to reduce the amount of calculation. Given a test pixel, we used the KNN algorithm to select the *K* pixels that are most similar to the local adaptive dictionary set. Second, we constructed the PENRC model of the hyperspectral image. We used the local adaptive dictionary to construct the PEN model and obtain the abundance coefficients corresponding to the testing pixel. Then, we calculated the reconstruction error of each class according to the abundance coefficients and used the minimum reconstruction error to classify the testing pixels. In addition, in order to further improve the classification performance, we also integrated the spatial information of the pixel neighborhood into the model, named joint pairwise elastic net representation-based classification (J-PENRC).

Algorith	m 1 the Proposed PEN Algorithm
Input:	(1) $\mathbf{X} \in \mathbb{R}^{B imes N}$, the training set.
	(2) K , λ .
Procedu	re:
Step 1:	Obtain adaptive dictionary D by applying KNN.
Step 2:	Obtain weight vector $\hat{\alpha}$ according to Equation (8):
	for $i = 1$: N
	update $\hat{\alpha}_i$ by Equations (19) and (20).
Step 3:	Decide the final label $class(y)$ by the minimum reconstruction error principle by
Equati	on (14).
Outp	at:
c	lass(y).

3.1. Local Adaptive Dictionary

In representation-based methods, dictionaries are usually composed of all labeled training pixels [32,33]. In order to have a robust representation, it is necessary to ensure that the dictionary is complete (that is, enough training samples are needed). However, training samples are usually limited in practice. In addition, using all training pixels directly will lead to a large amount of computation. Therefore, to solve the above problems, we utilize the local adaptive dictionary to obtain a more robust representation.

For a testing pixel y, we utilize the KNN to construct a similar signal set **D** as the adaptive dictionary. However, due to the high dimension of the hyperspectral image, it is unreasonable to directly use Euclidean distance to measure the similarity of the spectral vector. In order to increase the separability of data, LDA [34] algorithm is used to project HSIs into low-dimensional space, which can find an optimal projection direction to minimize the intraclass distance of samples and maximize the inter-class distance. Let

 $\Gamma \in R^{B' \times B}$ indicate the LDA mapping matrix and B' represent the reduced dimension. Then, the similarity measure between the testing atom y and arbitrary training atoms x_n can be expressed as:

$$d_n = \|\mathbf{\Gamma} \mathbf{y} - \mathbf{\Gamma} \mathbf{x}_n\|. \tag{5}$$

Then, we sorted all the distance set $[x_1, x_2, \ldots, x_N]$ in descending order and obtained the dictionary indices $\{i_c\}_{c=1,..,C}$ corresponding to the first *K* large distance values. The adaptive dictionary can be denoted as:

$$\mathbf{D} = \mathbf{X}(:, i_c), c = 1, \dots, C.$$
(6)

3.2. Pairwise Elastic Net Representation Based Classification

First, we introduce the concept of correlation matrix. Consider the following two matrices R_1 and R_2 :

$$R_1 = \begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix} \quad R_2 = \begin{pmatrix} 1.0 & 0.9 & 0.0 \\ 0.9 & 1.0 & 0.3 \\ 0.0 & 0.3 & 1.0 \end{pmatrix}.$$
 (7)

We can see that the three features in the R_1 matrix have the same similarity values. At this point, it is effective to set the global trade-off between l_1 -norm and l_2 -norm. Nevertheless, for the matrix R_2 , feature 1 is very similar to feature 2 (regarding l_2 -norm), feature 1 is independent from feature 3 (regarding l_1 -norm) and feature 2 is slightly related to feature 3 (regarding elastic net). Hence, we need a flexible trade-off scheme to match the regularization term with the data structure.

Thus, the objective function of our proposed PENRC can be denoted as:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{y} - \mathbf{D}\boldsymbol{\alpha}\|_{2}^{2} + \lambda \Big(\|\boldsymbol{\alpha}\|_{2}^{2} + \|\boldsymbol{\alpha}\|_{1}^{2} - |\boldsymbol{\alpha}|^{T}\mathbf{R}|\boldsymbol{\alpha}|\Big),$$
(8)

where **R** is the similarity matrix between atoms in the adaptive dictionary $\mathbf{D} \in \mathbb{R}^{K \times K}$. Some frequently-used similarity measures are absolute atom correlation $\mathbf{R}_{ij} = |\mathbf{D}_i^T \mathbf{D}_j|$ and Gaussian kernel $\mathbf{R}_{ij} = \exp\left(-\|\mathbf{D}_i - \mathbf{D}_j\|^2/\sigma^2\right)$ et al. Considering some basic results and notation with abundance coefficients and similarity matrix:

$$\|\boldsymbol{\alpha}\|_{2}^{2} = |\boldsymbol{\alpha}|^{T} I |\boldsymbol{\alpha}|$$
(9)

$$\|\boldsymbol{\alpha}\|_{1} = |\boldsymbol{\alpha}|^{T} \mathbf{1} = \mathbf{1}^{T} |\boldsymbol{\alpha}|$$
(10)

$$\|\boldsymbol{\alpha}\|_{1} = |\boldsymbol{\alpha}|^{T} \mathbf{1} = \mathbf{1}^{T} |\boldsymbol{\alpha}|$$
(10)
$$\|\boldsymbol{\alpha}\|_{1}^{2} = |\boldsymbol{\alpha}|^{T} \mathbf{1} \mathbf{1}^{T} |\boldsymbol{\alpha}|,$$
(11)

where I is the identity matrix and **1** is a vector of all ones. Then, the fourth term in Equation (8) representing the trade-off between l_1 -norm and l_2 -norm can be explained as follows. For the completely similar features, $\mathbf{R} = \mathbf{1}\mathbf{1}^T$. Equation (8) only left l_2 -norm, reducing the impact of the l_1 constraint. For the completely dissimilar features, $\mathbf{R} = I$, and Equation (8) reduces to SRC model with only l_1 constraint. That is to say, when the two features are similar, we take the CRC method; when the two features are dissimilar, we take the SRC method; for the remaining cases, we take the ENRC method. Thus, the flexible trade-off scheme can be realized though our proposed PENRC.

To further enhance the classification performance, we also incorporate the spatial information of HSI pixel into the PENRC model. In [24], a shape adaptive (SA) region is proposed for each pixel. In our work, we utilized the neighbor information with SA and the chosen pixel can be represented by the average of all pixels in the SA window. For an arbitrary pixel y in the HSI, the corresponding SA set matrix can be denoted as

 $\mathbf{Y}_{SA} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T]$. T is the number of chosen pixels in SA. Then, the pixel \mathbf{y} introduced into spatial information can be obtained by

$$\bar{y}_{SA} = \frac{1}{T} \sum_{t=1}^{T} y_{T'}$$
 (12)

Then, the sparse coefficients α_{SA} for \bar{y}_{SA} can be denoted as:

$$\hat{\boldsymbol{\alpha}}_{SA} = \arg\min \|\bar{\boldsymbol{y}}_{SA} - \bar{\mathbf{D}}_{SA}\boldsymbol{\alpha}_{SA}\|_{2}^{2} + \lambda \Big(\|\boldsymbol{\alpha}_{SA}\|_{2}^{2} + \|\boldsymbol{\alpha}_{SA}\|_{1}^{2} - |\boldsymbol{\alpha}_{SA}|^{T}\mathbf{R}|\boldsymbol{\alpha}_{SA}|\Big),$$
(13)

Once the sparse coefficients α_{SA} is obtained, the final label can be determined by the category minimum reconstruction error:

$$\operatorname{class}(\boldsymbol{y}) = \arg\min_{\boldsymbol{c}=1,\dots,\boldsymbol{C}} \|\bar{\boldsymbol{y}}_{SA} - \bar{\mathbf{D}}_{c_{SA}}\boldsymbol{\alpha}_{c_{SA}}\|_{2}, \tag{14}$$

where, $\bar{\mathbf{D}}_{c_{SA}}$ and $\alpha_{c_{SA}}$ represent the subset of $\bar{\mathbf{D}}_{SA}$ and α_{SA} corresponding to *c*-th class, respectively.

3.3. Coordinate Descent

To solve Equation (8), we rewrite it as following:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda |\boldsymbol{\alpha}|^T \mathbf{P} |\boldsymbol{\alpha}|, \qquad (15)$$

where $\mathbf{P} = I + \mathbf{1}\mathbf{1}^T - \mathbf{R}$. As [27] proves, only if **P** has nonnegative entries and is a positive semidefinite (PSD) matrix, the second term $|\boldsymbol{\alpha}|^T \mathbf{P} |\boldsymbol{\alpha}|$ in above model is convex. However, the matrix **P** in Equation (15) is not always a PSD matrix. We can consider the following way as proved in [27]:

$$\mathbf{P}^{S}_{\theta} = \theta I + (1 - \theta) \mathbf{P},\tag{16}$$

where $\frac{\tau}{\tau+1} \leq \theta \leq 1$ and $\tau = -\min\{0, \lambda_{min}(\mathbf{P})\}$.

Then, Equation (15) can be seen as a quadratic program (QP) problem and can be solved by the QP solver. However, the QP solver does not meet the high-dimensional data requirements. In order to obtain more exact results, we use the coordinate descent method [35] in this paper. The approach can be summarized as: given a convex function $f(\alpha)$, we calculate the derivative $\frac{\partial f}{\partial \alpha_i}$; update α_i by holding all α_j (where $j \neq i$) fixed with the equation $\frac{\partial f}{\partial \alpha_i} = 0$; cyclic each α_i iteratively until the termination condition is satisfied. In PENRC, we have

$$f(\boldsymbol{\alpha}) = \arg\min \|\boldsymbol{y} - \mathbf{D}\boldsymbol{\alpha}\|_{2}^{2} + \lambda |\boldsymbol{\alpha}|^{T} \mathbf{P} |\boldsymbol{\alpha}|$$

= $\boldsymbol{y}^{T} \boldsymbol{y} - 2\boldsymbol{q}^{T} \boldsymbol{\alpha} + \boldsymbol{\alpha}^{T} \mathbf{Q} \boldsymbol{\alpha} + \lambda \sum_{i,j} \mathbf{P}_{i,j} |\boldsymbol{\alpha}_{i} \boldsymbol{\alpha}_{j}|,$ (17)

where **P** is PSD and nonnegative, $\mathbf{Q} = \mathbf{D}^T \mathbf{D}$ and $q = \mathbf{X}^T y$. Then, the derivative $\frac{\partial f}{\partial \mathbf{x}_i}$ is

$$\frac{\partial f}{\partial \boldsymbol{\alpha}_{i}} = -2\boldsymbol{q}_{i} + 2\boldsymbol{Q}_{i}^{T}\boldsymbol{\alpha} + 2\lambda \operatorname{sgn}(\boldsymbol{\alpha}_{i}) \sum_{j=1}^{K} \mathbf{P}_{i,j} |\boldsymbol{\alpha}_{j}|.$$
(18)

If the derivative is 0, we update α_i according to $\alpha_{-i} = \alpha_{\{1:K\}\setminus i}$:

$$(\mathbf{Q}_{ii} + \mathbf{P}_{ii})\boldsymbol{\alpha}_i + \operatorname{sgn}(\boldsymbol{\alpha}_i)\lambda \sum_{j \neq i} \mathbf{P}_{ij} |\boldsymbol{\alpha}_j| = \boldsymbol{q}_i - \sum_{j \neq i} \mathbf{Q}_{ij}\boldsymbol{\alpha}_j.$$
(19)

Then, we define the scalars *a*, *b* and *c*. Let $a = \mathbf{Q}_{ii} + \mathbf{P}_{ii}$, $b = \lambda \sum_{j \neq i} \mathbf{P}_{ij} |\mathbf{\alpha}_j|$ and $c = q_i - \sum_{j \neq i} \mathbf{Q}_{ij} \mathbf{\alpha}_j$. The update equation can be denoted as:

$$\mathbf{\alpha}_{i} = \begin{cases} (c+b)/a & c < -b \\ 0 & -b \le c \le b \\ (c-b)/a & c > b. \end{cases}$$
(20)

4. Results

In this part, to validate the superiority of our proposed PENRC, we compare our proposed PENRC (pixelwise) and PENRC with both the single pixel-based and spatial information-based algorithms such as the KNN [36], SRC [15], CRC [16], fused representation-based classification (FRC) method [25], elastic net representation-based classification (ENRC) method [25], nearest regularized subspace (NRS) classifier [18], shape adaptive joint sparse representation classification (SAJSRC) [24] and weighted joint nearest neighbor and sparse representation (WINN-JSR) [37]. All the experiments are conducted using MATLAB R2014b on a 2.50 GHz PC with 8.0 GB RAM.

4.1. Data Set

In this paper, we chose the three HSI data sets for experimental evaluation.

The first testing data set is Indian Pines dataset. The scene is obtained by AVIRIS sensor over the Indian Pines test site in Northwest Indiana [38]. The size of the image is 145×145 with 224 spectral reflectance bands whose wavelength ranging from 0.4 µm to 2.5 µm. Removing the crops with less coverage, we choose 9 kinds of crops in the given ground truth which are *corn-notill, corn-mintill, grass-pasture, grass-trees, hay-windrowed, soybean-notill, soybean-notill, soybean-clean* and *woods*. Figure 1a,b illustrate the corresponding false color composition and ground truth map respectively.



Figure 1. Indian Pines dataset. (a) composite color image. (b,c) ground truth.

The second data set is the Pavia Centre data set, which is acquired by the ROSIS sensor during a flight campaign over Pavia. The geometric resolution is 1.3 m. The image size is $1096 \times 715 \times 102$. Due to the lack of information in the image, some samples do not contain any information. Therefore, it must be discarded before analysis. For Pavia Centre, we chose nine classes in the given ground truth: *water*, *trees*, *asphalt*, *self-blocking bricks*, *bitumen*, *tiles*, *shadows*, *meadow* and *bare soil*. Figure 2a,b illustrate the corresponding false color composition and ground truth map, respectively.

The third one is the Pavia University data set, also collected by the ROSIS sensor. The spatial resolution is 610×340 , and it contains 103 spectral bands. The Pavia University dataset contains nine classes with the given ground truth: *asphalt, meadows, gravel, trees, paninted mental sheets, bare soil, bitumen, self-blocking bricks* and *shadows*. Figure 3a,b illustrate the corresponding false color composition and ground truth map, respectively.



Figure 2. Pavia Center dataset. (a) Composite color image. (b,c) Ground truth.



Figure 3. Pavia University dataset. (a) Ccomposite color image. (b,c) Ground truth.

4.2. Parameter Analysis

During the experiment, we used three evaluation indicators to measure the classification performance: OA, AA and Kappa [39]. OA represents the proportion of all correctly classified atoms to the total number of testing atoms, while AA is the average value of OAs in each class. Kappa indicates the percentage of classified testing pixels corrected by the number of agreements that would be expected by chance. Detailed definitions for each indicator can be referred to [40].

There are two main parameters (the number of adaptive dictionary atoms *K* and balancing parameter λ) that have a significant impact on classification results in our proposed PENRC. In this section, we analyze the impact of the two parameters by carrying out the sweep of the chosen parameter space and find the optimal parameters according to Figure 2. For Indian Pines, we chose 10% pixels per class as training samples. For Pavia Center, we chose 100 pixels per class as training samples and the same number for the Pavia University dataset. From Figure 4, we can see that OA increases first and then decreases with the *K* value increasing. Few adaptive dictionary atoms lack enough locality information, and too many dictionary atoms may introduce redundant category information. Then, we fixed the value of *K*, and the classification can be locally maximum with the appropriate value of λ . Then, from the maximum OA shown in Figure 4, we set *K* to 20 and λ to 1×10^{-3} , 1×10^{-2} and 1×10^{-4} for Indian Pines, Pavia Center and Pavia University, respectively.





Figure 4. Effects of the number of adaptive dictionary atoms *K* and balancing parameter λ . (a) Indian Pines dataset, (b) Pavia Center dataset and (c) Pavia University dataset.

4.3. Comparisons with Other Approaches

To avoid any bias, we repeated the experiments five times and reported the average classification accuracy.

For Indian Pines, we employ 10% labeled samples per class as training set and others as testing set. The detailed partition strategy is illustrated in Table 1. From Table 2, we can see the classification performance of our proposed PENRC and J-PENRC as well as chosen compared algorithms, and the optimal results for each class are indicated in bold. For certain classes, such as grass-pasture, grass-trees, hay-windrowed and woods, the classification accuracies of our proposed PENRC and J-PENRC can be above 98%, specially for hay-windrowed, which can be up to 100%. For category soybean-clean, our algorithm improves the classification accuracy by 19.08% relative to the chosen optimal comparison algorithm ENRC. Furthermore, from Table 1, we can clearly see that our algorithms are optimal in terms of OA, AA and Kappa. In order to prove the effectiveness of our algorithm more comprehensively, we also compare the OAs, which are calculated under the different number of training samples. The classification results are shown in Figure 4. The abscissa represents the number of training samples per class, and the ordinate represents the classification accuracy. The dashed line represents OAs of the pixelwise algorithms, and the solid line represents OAs of the algorithms based on spatial information. From Figure 4, we can see that even in the case of insufficient training samples, our algorithm can achieve an ideal classification result. Furthermore, our algorithm have always been optimal compared to the same kind of contrast algorithms.

Table 1. List of the number of samples involved in training and testing for each class in Indian Pines, Pavia Center and Pavia University datasets.

	India	n Pines		Pavia C	Center		Pavia University		
No.	Name of Class	Traning	Testing	Name of Class	Traning	Testing	Name of Class	Traning	Testing
1	Corn-notill	142	1286	Water	100	2500	Asphalt	100	800
2	Corn-mintill	83	747	Trees	100	2500	Meadows	100	800
3	Grass-pasture	49	434	Meadow	100	2500	Gravel	100	800
4	Grass-trees	73	657	Self-Blocking Bricks	100	2500	Trees	100	800
5	Hay-windrowed	48	430	Bare Soil	100	2500	Painted mental sheets	100	800
6	Soybean-notill	98	874	Asphalt	100	2500	Bare Soil	100	800
7	Soybean-mintill	246	2209	Bitumen	100	2500	Bitumen	100	800
8	Soybean-clean	60	533	Tile	100	2500	Self-Blocking Bricks	100	800
9	Woods	127	1138	Shadows	100	2500	Shadows	100	800

No.	KNN	SRC	CRC	FRC	ENRC	NRS	PENRC	SA-JSR	WJNN-JSR	J-PENRC
1	58.44	59.48	66.49	64.03	57.69	88.99	78.44	91.60	94.16	96.75
2	54.69	62.28	63.39	64.73	67.38	89.60	78.12	86.75	92.50	97.10
3	95.00	90.38	96.54	97.96	93.09	63.78	98.46	94.01	99.77	100
4	96.70	98.97	96.70	93.91	99.46	89.96	98.98	100	100	100
5	100	98.44	99.22	99.22	98.18	98.62	100	100	100	100
6	62.68	65.33	38.10	52.87	70.16	70.30	75.62	93.59	94.17	97.90
7	79.20	78.21	90.42	90.65	82.11	69.57	90.79	95.25	95.97	97.44
8	51.72	51.72	41.38	52.66	59.60	58.45	78.68	91.18	92.32	95.92
9	93.41	94.00	98.83	95.46	96.60	98.05	99.56	97.89	98.33	99.82
OA (%)	75.55	76.31	78.06	80.25	79.16	86.09	87.70	94.40	96.00	98.05
AA (%)	76.90	77.65	77.65	79.80	80.62	80.93	88.57	94.47	96.36	98.33
Kappa	71.27	72.14	72.14	76.54	75.57	82.39	85.44	93.42	95.31	97.71

Table 2. Classification results of Indian Pines by pixelwise algorithms (KNN, SRC, CRC, FRC, ENRC, NRS and PENRC) and spatial-based algorithms (SA-JSR, WJNN-JSR and J-PENRC). Bold indicates the best result.

For Pavia Center, we employ 100 labeled samples per class as a training set and 2500 per class as a testing set. The detailed partition strategy is illustrated in Table 1. Table 3 illustrates the classification performance of our proposed PENRC and J-PENRC compared to other chosen algorithms, and the optimal results for each class are indicated in bold. For *meadow*, the classification accuracies of our proposed PENRC can be above 99.6%. For some classes, such as *asphalt* and *tile*, the classification accuracies of our J-PENRC can be above 99%. Especially for *water*, the classification accuracy of both PENRC and J-PENRC can be up to 100%. Furthermore, Table 3 illustrates that our proposed algorithms are optimal in terms of OA, AA and Kappa compared to other chosen algorithms. In order to further prove the effectiveness of our algorithm, we also compare the OAs of chosen algorithms under the different number of training samples. The classification results are shown in Figure 5. The number of training samples is selected from 50 samples per class to 300 samples per class. It can be seen from Figure 5 that compared with similar algorithms, our algorithm always has the best classification effect.

Table 3. Classification results of Pavia Center by pixelwise algorithms (KNN, SRC, CRC, FRC, ENRC, NRS and PENRC) and spatial-based algorithms (SA-JSR, WJNN-JSR and J-PENRC). Bold indicates the best result.

No.	KNN	SRC	CRC	FRC	ENRC	NRS	PENRC	SA-JSR	WJNN-JSR	J-PENRC
1	99.11	99.67	99.67	100	99.81	100	100	100	100	100
2	89.56	76.37	82.33	84.17	79.67	91.85	89.17	93.67	87.11	94.67
3	87.89	90.21	88.00	86.31	92.33	87.67	99.67	98.72	95.51	99.33
4	84.33	79.32	24.56	87.42	80.17	76.29	93.16	99.82	96.60	97.31
5	88.89	89.50	67.50	84.50	89.67	85.26	96.09	99.00	81.83	93.08
6	88.11	77.83	97.67	79.67	76.85	97.31	80.73	68.67	97.52	99.41
7	86.44	88.81	86.10	84.43	88.23	83.83	94.25	96.83	85.14	97.28
8	95.33	97.01	99.03	97.21	98.15	99.50	99.50	95.04	96.83	99.60
9	100	93.00	82.33	93.42	95.50	99.50	94.62	99.71	100	100
OA(%)	91.07	87.06	80.19	88.56	88.93	91.24	94.11	94.83	92.97	97.78
AA(%)	91.07	87.06	80.19	88.56	88.93	91.24	94.11	94.83	92.97	97.78
Kappa	89.96	86.46	78.15	87.13	87.54	90.51	93.38	94.19	92.14	97.50

With regard to the Pavia University dataset, we randomly selected 100 labeled samples per class as a training set and 800 per class in the rest as a testing set (such as the *shaows* class, which only contains 947 labeled samples). The detailed partition strategy is illustrated in Table 1. Table 4 presents the classification result of our proposed PENRC and J-PENRC with other comparison algorithms, and the optimal results for each class are denoted in bold.

For *bitumen*, the classification accuracy of our proposed PENRC reached 99.17%. For some classes, such as *gravel* and *bair coil*, the classification accuracy of our J-PENRC can be above 97%. Especially for *meadows*, *painted mental sheets* and *shadows*, the classification accuracy of J-PENRC can be up to 100%. In addition, in terms of OA, AA and Kappa, Table 3 also illustrates that our proposed algorithms are optimal compared to other chosen algorithms. In order to further prove the effectiveness of our algorithm, we also compared the OAs of the chosen algorithms with different numbers of training samples. The classification results are shown in Figure 5. The number of training samples is selected from 50 samples per class to 300 samples per class. It is evident that our algorithm always gains the most extraordinary performance.

Table 4. Classification results of Pavia University by pixelwise algorithms (KNN, SRC, CRC, FRC, ENRC, NRS and PENRC) and spatial-based algorithms (SA-JSR, WJNN-JSR and J-PENRC). Bold indicates the best result.

No.	KNN	SRC	CRC	FRC	ENRC	NRS	PENRC	SA-JSR	WJNN-JSR	J-PENRC
1	70.83	57.67	36.00	56.83	60.67	91.17	72.00	94.16	70.00	87.00
2	70.33	78.00	75.00	80.17	68.50	71.00	97.33	92.50	81.33	100
3	69.67	72.83	92.67	67.33	73.33	77.50	97.00	98.59	82.00	98.67
4	88.67	89.50	96.67	94.33	92.00	95.33	93.83	100	96.73	97.13
5	98.50	99.50	100	99.83	99.27	99.17	99.67	100	99.83	100
6	66.33	65.17	57.33	64.00	68.33	83.00	95.83	94.17	85.83	97.00
7	85.50	87.00	92.17	85.83	87.00	86.50	99.17	95.97	95.00	99.00
8	66.83	67.83	20.17	72.00	69.00	64.50	86.83	92.32	80.17	94.33
9	100	94.95	93.33	97.33	98.17	99.67	97.83	98.33	99.83	100
OA(%)	79.63	79.14	73.70	79.74	79.57	85.31	93.28	96.00	87.81	96.69
AA(%)	79.63	79.14	73.70	79.74	79.57	85.31	93.28	96.00	87.81	96.69
Kappa	77.08	76.31	70.42	77.21	77.02	83.48	92.44	95.31	86.29	96.17



Figure 5. Classification performance for different numbers of training samples per class. (**a**) Indian Pines dataset, (**b**) Pavia Center dataset and (**c**) Pavia University dataset.

4.4. Computational Complexity

In this section, we compare the computational complexity for each classifier with the Indian Pines, Pavia University and Pavia Centre datasets. All above experiments were executed five times to avoid any bias. Table 5 illustrates the total time of algorithm execution and verification of the three datasets. All experimental settings and the parameters were set to be the same as described above. As can be seen from Table 5, ENRC has a lower time complexity than PENRC. There are two reasons for this. First, ENRC uses artificial prior information to set a fixed weight parameter to combine l_1 -norm and l_2 -norm, while PENRC automatically learns this weight parameter through the similarity matrix. Second, the time complexity consumed by the solution approximation algorithm used by the two methods is not the same due to the difference in the math models. On the other hand,

Table 5 also lists the time complexity comparison with or without LAD. Obviously, the use of LAD substantially reduces the computational complexity of PENRC and yields a better classification performance.

Table 5. Computational complexity comparison in Indian Pines dataset.

	With/Without LAD	Running Time (s)	Overall Accuracy
ENRC	-	32.53	79.16
DENIDC	X	3472.68	85.44
I LINKC	\checkmark	72.35	87.70

5. Conclusions

In this paper, we proposed a hyperspectral image classification algorithm named PENRC. The local constrained dictionary was first constructed to reduce the computation costs. Then, by introducing a correlation matrix, the PENRC was constructed to realize the group sparsity with self-balancing between l_1 -norm and l_2 -norm. The pairwise elastic net model was proven to be capable of the grouping selection of highly correlated data via establishing local, or pairwise, tradeoffs of similarity between correlation matrices, thereby rendering more robust weight coefficients. To further improve the classification performance, we also introduced spatial information and proposed the J-PENRC model. The experimental results of real hyperspectral images verified that the proposed algorithms could outperform the existing representation-based classifiers. Compared to the existing pixelwise and spatial-based algorithm, experiments on our chosen Indian Pines verified the effectiveness of our proposed PENRC and J-PENRC in quantitative and qualitative terms.

Author Contributions: All authors have made great contributions to the work. Conceptualization,Y.Z., Y.M. and X.M.; Software, Y.Z. and X.M.; Writing—original draft, Y.M. and Y.Z.; Writing—review and editing, H.L., S.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant no. 61903279 and 61773295) and NSFC (grant no. 61906140), NSFC-CAAC (grant no. U1833119), Hubei Natural Science Foundation for Distinguished Young Scholars (2020CFA063) and National Food and Strategic Reserves Administration Foundation (grant no. LQ2018501).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HSI	Hyperspectral Image
SR	Sparse Representation
CR	Collaborative Representation
PENRC	Pairwise Elastic Bet Representation Based Classification
J-PENRC	Joint-Pairwise Elastic Bet Representation Based Classification
SVM	Support Vector Machine
GMM	Gaussian Mixture-Model
MLC	Maximum-Likelihood Classifier
SRC	Sparse Representation Classification
CRC	Collaborative Representation Classification
JSRC	Joint Sparse Representation Classification

NRS	Nearest Regularized Subspace
cdSRC	Class-dependent Sparse
KCRC	Kernel-Based CRC
NJCRC	Nonlocal Joint Collaborative Representation
K-NN	K-Nearest Neighbor
SAJSRC	Shape Adaptive Joint Sparse Representation
FRC	Fused Representation-Based Classification
ENRC	Elastic Net Representation Based Classification
PEN	Pairwise Elastic Net
SA	Shape Adaptive
QP	Quadratic Program
WINN-JSR	Joint Nearest Neighbor and Sparse Representation

References

- Bykov, A.; Zherebtsov, E.; Dremin, V.; Popov, A.; Doronin, A.; Meglinski, I. Hyperspectal Skin Imaging with Artificial Neural Networks Validated by Optical Biotissue Phantoms. In Proceedings of the Computational Optical Sensing and Imaging, Munich, Germany, 24–27 June 2019; Optical Society of America: Washington, DC, USA, 2019; p. CW1A–3.
- 2. Keshava, N. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1552–1565. [CrossRef]
- 3. Bishop, C.A.; Liu, J.G.; Mason, P.J. Hyperspectral remote sensing for mineral exploration in Pulang, Yunnan Province, China. *Int. J. Remote Sens.* **2011**, *32*, 2409–2426. [CrossRef]
- 4. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
- Li, W.; Prasad, S.; Fowler, J.E. Decision fusion in kernel-induced spaces for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 2013, 52, 3399–3411. [CrossRef]
- Li, W.; Prasad, S.; Fowler, J.E.; Bruce, L.M. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.* 2011, *50*, 1185–1198. [CrossRef]
- Zhang, Y.; Ma, Y.; Dai, X.; Li, H.; Mei, X.; Ma, J. Locality-constrained sparse representation for hyperspectral image classification. *Inf. Sci.* 2021, 546, 858–870. [CrossRef]
- 8. Dian, R.; Li, S.; Fang, L. Learning a low tensor-train rank representation for hyperspectral image super-resolution. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2672–2683. [CrossRef]
- 9. Dong, W.; Wang, H.; Wu, F.; Shi, G.; Li, X. Deep spatial–spectral representation learning for hyperspectral image denoising. *IEEE Trans. Comput. Imaging* 2019, *5*, 635–648. [CrossRef]
- Sellami, A.; Dupé, F.X.; Cagna, B.; Kadri, H.; Ayache, S.; Artières, T.; Takerkart, S. Mapping individual differences in cortical architecture using multi-view representation learning. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
- 11. Ghamisi, P.; Maggiori, E.; Li, S.; Souza, R.; Tarablaka, Y.; Moser, G.; De Giorgi, A.; Fang, L.; Chen, Y.; Chi, M.; et al. New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning. *IEEE Geosci. Remote Sens. Mag.* **2018**, *6*, 10–43. [CrossRef]
- Sellami, A.; Farah, I. Spectra-spatial graph-based deep restricted boltzmann networks for hyperspectral image classification. In Proceedings of the 2019 PhotonIcs & Electromagnetics Research Symposium-Spring (PIERS-Spring), Rome, Italy, 17–20 June 2019; pp. 1055–1062.
- 13. Mei, X.; Pan, E.; Ma, Y.; Dai, X.; Huang, J.; Fan, F.; Du, Q.; Zheng, H.; Ma, J. Spectral-spatial attention networks for hyperspectral image classification. *Remote Sens.* **2019**, *11*, 963. [CrossRef]
- 14. Lei, Z.; Zeng, Y.; Liu, P.; Su, X. Active deep learning for hyperspectral image classification with uncertainty learning. *IEEE Geosci. Remote Sens. Lett.* **2021**. [CrossRef]
- 15. Li, C.; Ma, Y.; Mei, X.; Liu, C.; Ma, J. Hyperspectral image classification with robust sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 641–645. [CrossRef]
- 16. Jia, S.; Deng, X.; Zhu, J.; Xu, M.; Zhou, J.; Jia, X. Collaborative representation-based multiscale superpixel fusion for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7770–7784. [CrossRef]
- 17. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans. Geosci. Remote Sens.* 2011, 49, 3973–3985. [CrossRef]
- 18. Li, W.; Tramel, E.W.; Prasad, S.; Fowler, J.E. Nearest regularized subspace for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, 52, 477–489. [CrossRef]
- 19. Cui, M.; Prasad, S. Class-dependent sparse representation classifier for robust hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 2014, *53*, 2683–2695. [CrossRef]
- Zhang, L.; Yang, M.; Feng, X. Sparse representation or collaborative representation: Which helps face recognition? In Proceedings
 of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 471–478.

- 21. Li, W.; Du, Q. Joint within-class collaborative representation for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2200–2208. [CrossRef]
- 22. Li, W.; Du, Q.; Xiong, M. Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 2014, 12, 48–52.
- 23. Li, J.; Zhang, H.; Huang, Y.; Zhang, L. Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary. *IEEE Trans. Geosci. Remote Sens.* 2013, *52*, 3707–3719. [CrossRef]
- 24. Fu, W.; Li, S.; Fang, L.; Kang, X.; Benediktsson, J.A. Hyperspectral image classification via shape-adaptive joint sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, *9*, 556–567. [CrossRef]
- 25. Li, W.; Du, Q.; Zhang, F.; Hu, W. Hyperspectral image classification by fusing collaborative and sparse representations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4178–4187. [CrossRef]
- 26. Hui, Z.; Hastie, T. Regularization and variable selection via the elastic net. J. R. Stat. Soc. 2005, 67, 768.
- Lorbert, A.; Eis, D.; Kostina, V.; Blei, D.; Ramadge, P. Exploiting covariate similarity in sparse regression via the pairwise elastic net. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 477–484.
- Chen, S.; Donoho, D. Basis pursuit. In Proceedings of the 1994 28th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 31 October–2 November 1994; Volume 1, pp. 41–44.
- 29. Gill, P.R.; Wang, A.; Molnar, A. The in-crowd algorithm for fast basis pursuit denoising. *IEEE Trans. Signal Process.* 2011, 59, 4595–4605. [CrossRef]
- 30. Mallat, S.G.; Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [CrossRef]
- Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2012, *5*, 354–379.
 [CrossRef]
- 32. Huang, S.; Zhang, H.; Pižurica, A. A robust sparse representation model for hyperspectral image classification. *Sensors* 2017, 17, 2087. [CrossRef] [PubMed]
- 33. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7738–7749. [CrossRef]
- 34. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 862–873. [CrossRef]
- 35. Friedman, J.; Hastie, T.; Höfling, H.; Tibshirani, R. Pathwise coordinate optimization. Ann. Appl. Stat. 2007, 1, 302–332. [CrossRef]
- 36. Ma, L.; Crawford, M.M.; Tian, J. Local manifold learning-based *k*-nearest-neighbor for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4099–4109. [CrossRef]
- 37. Tu, B.; Huang, S.; Fang, L.; Zhang, G.; Wang, J.; Zheng, B. Hyperspectral image classification via weighted joint nearest neighbor and sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4063–4075. [CrossRef]
- Gualtieri, J.A.; Cromp, R. Support vector machines for hyperspectral remote sensing classification. In Proceedings of the 27th AIPR Workshop: Advances in Computer-Assisted Recognition, Washington, DC, USA, 14–16 October 1999; International Society for Optics and Photonics: Bellingham, WA, USA, 1999; Volume 3584, pp. 221–232.
- Ma, Y.; Zhang, Y.; Mei, X.; Dai, X.; Ma, J. Multifeature-Based Discriminative Label Consistent K-SVD for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2019, 12, 4995–5008. [CrossRef]
- 40. Richards, J.A.; Richards, J. Remote Sensing Digital Image Analysis; Springer: Berlin/Heidelberg, Germany, 1999; Volume 3.