



# Article An Efficient Representation-Based Subspace Clustering Framework for Polarized Hyperspectral Images

Zhengyi Chen <sup>1,2,3</sup>, Chunmin Zhang <sup>1,2,3,\*</sup>, Tingkui Mu <sup>1,2,3</sup>, Tingyu Yan <sup>1,2,3</sup>, Zeyu Chen <sup>1,2,3</sup> and Yanqiang Wang <sup>1,2,3</sup>

- <sup>1</sup> Institute of Space Optics, Xi'an Jiaotong University, Xi'an 710049, China
- <sup>2</sup> School of Science, Xi'an Jiaotong University, Xi'an 710049, China
- <sup>3</sup> Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter (Xi'an Jiaotong University), Ministry of Education, Xi'an 710049, China
- \* Correspondence: zcm@xjtu.edu.cn; Tel.: +86-029-8266-8016

Received: 7 May 2019; Accepted: 24 June 2019; Published: 26 June 2019



Abstract: Recently, representation-based subspace clustering algorithms for hyperspectral images (HSIs) have been developed with the assumption that pixels belonging to the same land-cover class lie in the same subspace. Polarization is regarded to be a complement to spectral information, but related research only focus on the clustering for HSIs without considering polarization, and cannot effectively process large-scale hyperspectral datasets. In this paper, we propose an efficient representation-based subspace clustering framework for polarized hyperspectral images (PHSIs). Combining with spectral information and polarized information, this framework is extensible for most existing representation-based subspace clustering algorithms. In addition, with a sampling-clustering-classification strategy which firstly clusters selected in-sample data into several classes and then matches the out-of-sample data into these classes by collaborative representation-based classification, the proposed framework significantly reduces the computational complexity of clustering algorithms for PHSIs. Some experiments were carried out to demonstrate the accuracy, efficiency and potential capabilities of the algorithms under the proposed framework.

Keywords: hyperspectral images; polarization; subspace clustering; sparse representation

# 1. Introduction

Polarization, which can describe surface roughness and edge properties of objects and increase the contrast between objects and background, has been demonstrated over recent decades to provide useful information for atmosphere monitoring, land surface characterizing and material classification [1–4]. In addition, hyperspectral images (HSIs) consist of high-resolution spectral correlation and rich spatial information that support land-cover classification and clustering [5–8]. As a combination of HSIs and polarization, polarized hyperspectral images (PHSIs) that provide the multidimensional information of polarization, spectral, spatial and radiant features are expected to possess great potential in object detection and clustering tasks.

There have been some studies on the clustering of Polarimetric Synthetic Aperture Radar (PolSAR) images which is the product of active detection by radars and receivers [9]. However, in the field of passive detection that relies on the natural light reflected by objects, existing studies have only focused on HSIs' clustering so far [10,11]. With the development of spectropolarimeter, for example, the channeled interference imaging spectropolarimeter (CIISP) [12], PHSIs become very convenient to be obtained. This helps adding polarization into HSIs' clustering to be possible. However, there are only few related studies on polarized hyperspectral clustering. Therefore, proposing a clustering

method that considers both of polarization and spectral information is a necessary complement to hyperspectral clustering and an extension of polarization applications as well.

Recently, a robust technique known as the representation-based clustering model has achieved remarkable success in the face recognition and motion segmentation fields [13–15]. Based on the assumption that pixels with different spectra from one kind of target lies in the same subspace, some clustering algorithms for HSIs were therefore proposed to make full use of the spectral-spatial information [16,17]. While due to the high computational complexity of representation-based clustering model, those algorithms usually require a long computation time and large computational memory. Especially for HSIs and PHSIs, which are usually accompanied by both high dimensionality and large scale, this shortcoming limits the application of those algorithms.

In this paper, a clustering framework for a representation-based clustering model combining the polarization and spectral properties of targets is presented. This clustering framework whose architecture is as shown in Figure 1 will be scalable for most of existing representation-based clustering algorithms. In order to reduce the computational complexity, a new sampling-clustering-classification strategy is adopted in the proposed framework. This sampling-clustering-classification strategy firstly samples the points of PHSIs by applying superpixel segmentation twice and then clusters the selected in-sample data with proposed polarized hyperspectral clustering method. Finally, the out-of-sample data will be matched into the clusters of in-sample data. The experimental results demonstrate that the proposed clustering framework significantly reduced the running time while ensuring overall accuracy.



Figure 1. Architecture of proposed efficient representation-based subspace clustering framework for polarized hyperspectral images.

The remainder of this paper is structured as follows. Section 2 reviews the general form of representation-based clustering algorithms and briefly introduces the representation-based clustering for hyperspectral images. In Section 3, we propose the efficient representation-based clustering framework for polarized hyperspectral images. The experimental results are given in Section 4. Finally, Section 5 concludes this paper.

#### 2. Representation-Based Subspace Clustering for Hyperspectral Images

#### 2.1. Representation-Based Subspace Clustering

Subspace clustering usually contains two tasks: (1) projecting the data into low-dimensional subspaces and (2) calculating the cluster membership of the dataset using statistical methods or spectral clustering. Additionally, the core of spectral clustering is the construction of a similarity graph of which

each vertex denotes a data point and the edge weights represent the similarities between connected points. The pairwise distance (PD) which computes the similarity based on the distance (e.g., the Euclidean distance) between two data points, and the reconstruction coefficients (RC) which denotes each data point as a linear representation of the other points, are two widely used approaches to build a similarity graph. Many of the recent studies, for example, the sparse subspace clustering (SSC) [13], low-rank representation subspace segmentation (LRR) [14] and least squares regression subspace segmentation (LSR) [15] have shown that the RC has a superior performance. These methods collectively called representation-based subspace clustering have a similar form. They first compute a self-representation matrix *C* by solving

$$\min_{C} \lambda \|Y - YC\|_{x} + \Omega(Y, C), \tag{1}$$

where  $Y \in \mathbb{R}^{d \times n}$  is the data matrix with each column being a sample vector,  $\|\bullet\|_x$  denotes a proper norm,  $\Omega(Y, C)$  is the regular term or penalty term, and  $\lambda > 0$  is a trade-off parameter.

After solving the optimization problem (1) by Alternating Direction Method of Multipliers (ADMM) [18,19], the affinity matrix is defined as:

$$W = (|C^*| + |C^*|^T)/2.$$
<sup>(2)</sup>

Finally, spectral clustering is applied to obtain the clustering result [20,21].

For different representation-based subspace clustering algorithms,  $\|\bullet\|_x$  and  $\Omega(Y, C)$  have different forms. Table 1 summarizes the choices of several existing representation-based models, where  $\|\bullet\|_1$  is the  $l_1$ -norm,  $\|\bullet\|_2$  is the  $l_2$ -norm,  $\|\bullet\|_*$  is the nuclear norm,  $\|\bullet\|_{2,1}$  is the  $l_{2,1}$ -norm and  $\|\bullet\|_F$  is the Frobenius norm.

**Table 1.** The choices of  $\|\bullet\|_x$  and  $\Omega(Y,C)$  of several existing representation-based algorithms.

Algorithms	●   <sub>x</sub>	$\Omega(Y, C)$	Subject to
SSC [13]	$  Y - YC  _F^2$	$  C  _2$	$diag(C) = 0, C^T 1 = 1$
LRR [14]	$  Y - YC  _1$ , or $  Y - YC  _{2,1}$ or $  Y - YC  _F$	$\ C\ _{*}$	$C^T 1 = 1$
LSR [15]	$  Y - YC  _2^2$	$  C  _2$	$diag(C) = 0, C^T 1 = 1$

#### 2.2. Representation-Based Subspace Clustering for HSIs

In the representation-based clustering, HSIs with a size of  $m \times n \times D$  ( $m \times n$  is the size of images and D is the number of spectral bands) can be transformed to  $2D mn \times D$  dataset where mn is the total number of pixels. Then all the pixels can be considered as a dataset  $Y = [y_1, y_2, ..., y_{mn}]$  in  $\mathbb{R}^D$  in which every single data point can be regard as selected from a union of k affine subspaces  $V_1 \cup V_2 \cup \cdots \cup V_k$ of dimensions  $\{d_i\}_{i=1}^k$  with  $\sum d_i = D$ .

Compared with other high-dimensional datasets, HSIs have some special features especially spatial features and spectral features. Directly applying the representation-based clustering, SSC algorithm for example, to HSIs is usually not able to exploit the spectral and spatial information. It is easy to find that similar data points have a high probability of sharing the same subspace. Based on this fact, some constraints from the spatial and spectral perspective are added [16,17]. Thus, the representation coefficient matrix *C* for HSIs can be achieved by solving the following optimization problem:

$$\min_{C} \|W \bullet C\|_{1} + \|Y - YC\|_{F}^{2}$$
s.t. diag(C) = 0, C<sup>T</sup>1 = 1 
(3)

where *W* is an added weighted matrix and • means the Hadamard product of two matrices.

Considering that the function of the weighted matrix is to choose the more highly correlated hyperspectral pixels to represent the target signals, the more highly correlated pixels should be given larger weight. For example, in [16] the weighted matrix *W* is defined in the following way:

$$W_{i,j} = \frac{1}{\|y_i - y_j\|_2^2 + \Delta}, \, i, j = 1, 2, \cdots mn, \, and \, i \neq j, \tag{4}$$

where and  $y_j$  are the spectral vectors corresponding to the ith pixel and the jth pixel of a given HSIs, and  $\Delta$  is a very small constant to avoid the extreme case of the weight of two very similar neighboring pixels.

# 3. Proposed Efficient Representation-Based Subspace Clustering Framework for Polarized Hyperspectral Images

#### 3.1. Representation-Based Clustering Framework for PHSIs

Although existing representation-based subspace clustering algorithms can reach a good performance in HSIs' clustering, they are not able to deal with two input datasets, which are the spectral information dataset and polarized information dataset, at the same time. In order to achieve the purpose of adding polarization information into HSI clustering, it is necessary to modify the subspace clustering for the characteristics of polarized spectral information. The most important part of subspace clustering algorithms is the composition of self-representation matrix *C*. If *C* can reflect both kinds of information, the clustering result will be a combination of two datasets. Considering that if two pixels are in the same cluster, they should be more likely to be assigned to the same subspace whether from the perspective of spectra or polarization, the optimization problem (1) becomes:

$$\min_{C} \Omega(Y^{(s)}, C) + \alpha \Omega(Y^{(p)}, C) + \lambda_{1} ||Y^{(s)} - Y^{(s)}C||_{x} + \lambda_{2} ||Y^{(p)} - Y^{(p)}C||_{x}$$

$$s.t. \, diag(C) = 0, \, C^{T}1 = 1$$
(5)

where  $\alpha$ ,  $\lambda_1$  and  $\lambda_2$  are trade-off parameters,  $Y^{(s)}$  and  $Y^{(p)}$  are spectral dataset and polarized dataset, respectively.

Furthermore, we change the optimization problem to the following form:

$$\min_{C^{(s)}, C^{(p)}} \Omega(Y^{(s)}, C^{(s)}) + \alpha \Omega(Y^{(p)}, C^{(p)}) + \lambda_1 \|Y^{(s)} - Y^{(s)}C^{(s)}\|_x + \lambda_2 \|Y^{(p)} - Y^{(p)}C^{(p)}\|_x$$

$$s.t. C^{(s)} = C^{(p)}, diag(C^{(s)}) = 0, diag(C^{(p)}) = 0, C^{(s)T}1 = 1, C^{(p)T}1 = 1$$

$$(6)$$

Optimization problem (6) can be solved using ADMM [18,19]. Then, we can get two coefficient matrices  $C^{(s)}$  and  $C^{(p)}$ . In addition, we use the following ideas to get the final representation matrix C:

(1) if two pixels are similar in  $Y^{(p)}$ , their corresponding  $Y^{(s)}$  should be more likely to be in the same subspace;

(2) also, if two pixels are similar in  $Y^{(s)}$ , their corresponding  $Y^{(p)}$  should be more likely to be in the same subspace.

So, we can get the final coefficient matrix *C* by:

$$C = (W_p \bullet C^{(s)} + \alpha W_s \bullet C^{(p)}) / (1 + \alpha).$$
(7)

Based on Euclidean distance,  $W_p$  is defined as:

$$W_p(i,j) = \exp[\sigma_p(\frac{1}{\|Y_i^{(p)} - Y_j^{(p)}\|_2^2} - 1)],$$
(8)

where  $\Delta$  is a very small constant that prevents the denominator from being zero,  $\sigma_p$  is a parameter which makes  $W_p$  adjustable on different datasets.

Additionally, in order to better represent the similarity of the spectral shape,  $W_s$  based on the spectral angle is proposed as:

$$W_{s}(i,j) = exp[\sigma_{s}(\frac{\sum_{m} Y^{(s)}(i,m) \cdot Y^{(s)}(j,m)}{\sqrt{\sum_{m} Y^{(s)}(i,m)^{2}} \cdot \sqrt{\sum_{m} Y^{(s)}(j,m)^{2}}})].$$
(9)

In the same way as in Section 2, *C* is utilized to build the similarity graph. The clustering result can then be achieved by applying the spectral clustering algorithm to the similarity graph [20,21]. The modified versions of algorithms listed in Table 1 are called Polarized Spectral Sparse Subspace Clustering (PS-SSC), Polarized Spectral Low-Rank Representation clustering (PS-LRR), and Polarized Spectral Least Squares Regression clustering (PS-LSR), respectively.

## 3.2. The Sampling-Clustering-Classification Strategy

Due to the large size of polarized hyperspectral images, representation-based subspace clustering programs usually require a large computer memory and long computing time, which is not in line with actual needs. For example, if the size of image is 512 \* 512, the coefficient matrix *C* will be a 262,144 \* 262,144 matrix which makes clustering a hard task. In some related works, superpixel theory is used to reduce the time complexity and space complexity of algorithms [22,23]. However, even if we reduce the size of the input data from 512 \* 512 pixels to thousands of superpixels, it will still take a lot of time to solve the optimization problem. For this reason, taking the SSC model as an example, a new sampling-clustering-classifying strategy is proposed and the PS-SSC based on this strategy is called Fast Polarized Spectral Sparse Subspace Clustering (FPS-SSC).

The first step of FPS-SSC algorithm is sampling. The original polarized hyperspectral images are firstly segmented by simple linear iterative clustering (SLIC) algorithm into  $N_1$  superpixels  $S = [S_1, S_2, ..., S_{N_1}]$  [24]. Then, we select a small number of data points as in-sample data and apply the PS-SSC algorithm over them. It is a challenging problem to select some key points as in-sample data. Although random sampling performs well in some cases, there is still some instability [25]. Due to the spatial characteristics of HSIs, here we propose a superpixel-based method to select in-sample data and its architecture is as shown in Figure 2.



Figure 2. Architecture of proposed superpixel-based selection method of in-sample data.

In addition to the superpixel segmentation mentioned above, the SLIC is applied a second time to segment the original images into another superpixel set  $S'=[S'_1, S'_2, \cdots, S'_{N_2}]$  with a number of  $N_2$  which is much fewer than  $N_1$ . We can easily get the geometric center coordinates  $GC' = [(x'_1, y'_1), (x'_2, y'_2), \cdots, (x'_{N_2}, y'_{N_2})]$  of each superpixel of S'. Then, the geometric centers GC' are projected to the superpixel set S, and the superpixels in which the projected points located are the selected in-sample superpixels.

By applying superpixel segmentation twice, the in-sample data can be selected stably. Then we can get the cluster assignment of in-sample data by applying the PS-SSC over them, which is the second step of FPS-SSC.

Naturally, the last part of FPS-SSC is to obtain the cluster membership of out-of-sample data by performing classification over it. Inspired by [25], we apply collaborative representation-based classification to measure the adjacency relationship among data points. In addition, the collaborative representation-based classification algorithm is modified to accommodate polarized hyperspectral images.

For each out-of-sample data we can get the representation matrix  $c_i$  regarding all in-sample data as a dictionary:

$$\min_{c_i} \|c_i\|_2^2 + \beta_1 \|O_i^{(s)} - I^{(s)}c_i\|_2^2 + \beta_2 \|O_i^{(p)} - I^{(p)}c_i\|_2^2,$$
(10)

where  $O_i^{(s)}$  and  $O_i^{(p)}$  are the spectral data and the polarization data of each out-of-sample data  $O_i$ ,  $I^{(s)}$  and  $I^{(p)}$  are the spectral data and the polarization data of all in-sample data, respectively.

The solution of optimization problem (10) can be written as:

$$c_i^* = \left[\beta_1(I^{(s)})^T I^{(s)} + \beta_2(I^{(p)})^T I^{(p)} + I\right]^{-1} \left[\beta_1(I^{(s)})^T O_i^{(s)} + \beta_2(I^{(p)})^T O_i^{(p)}\right].$$
(11)

Once the optimal  $c_i^*$  is achieved, for each cluster  $R_k$ , the residual between  $O_i$  and the in-sample points in  $R_k$  can be obtained as:

$$r_k(O_i) = \frac{\beta_1 ||O_i^{(s)} - I^{(s)}D_k(i,j)c_i^*||_2 + \beta_2 ||O_i^{(p)} - I^{(p)}D_k(i,j)c_i^*||_2}{||D_k(i,j)c_i^*||_2},$$
(12)

where  $D_k$  is a diagonal matrix based on the in-sample clustering result:

$$D_k(i, j) = \begin{cases} 1, & i = j, \text{ and } j \text{th in} - \text{sample data belongs to } k \text{th cluster} \\ 0, & \text{otherwise} \end{cases}$$
(13)

 $O_i$  is then matched into the cluster that has the minimum residual:

$$identity(O_i) = \underset{k}{\operatorname{argmin}}[r_k(O_i)], \tag{14}$$

where *identity*( $O_i$ ) denotes the assignment of the out-of-sample data  $O_i$ , which means  $O_i$  belongs to *k*th cluster if  $r_k(O_i)$  is the minimum value of  $\{r_k(O_i)\}_{k=1}^K$ .

The proposed efficient clustering framework for polarized hyperspectral images is also scalable to LRR, LSR and other representation-based subspace clustering models. In addition, algorithms applying this framework can be summarized as Algorithm 1.

#### Algorithm 1 FPS-SSC, FPS-LRR and FPS-LSR

#### Input:

Spectral dataset  $Y^{(s)}$  and polarized dataset  $Y^{(p)}$ ; the desired number of clusters k,  $\lambda_1$ ,  $\lambda_2$ ,  $\sigma_s$ ,  $\sigma_p$ ,  $\beta_1$  and  $\beta_2$ . **Main algorithm:** 

- (1) Sampling: apply the superpixel-based method to segment images into  $N_1$  superpixels and to select  $N_2$  in-sample data.
- (2) Clustering: perform polarized hyperspectral clustering algorithm (PS-SSC or PS-LRR or PS-LSR) over in-sample data.
- (3) Classification: match each out-of-sample data into the clusters of in-sample data with (10)–(14).

#### **Output:**

A 2-D matrix which records the labels of the clustering result of the polarized hyperspectral images.

#### 4. Experimental Results and Discussion

In this section, some experiments are carried out to demonstrate the accuracy, efficiency and potential capabilities of the proposed polarized hyperspectral clustering framework. As described above, this method can be applied to most representation-based clustering algorithms. Here we mainly use the polarized spectral version and fast polarized spectral version of SSC as examples, and take LRR and LSR as supplementary analysis. We also compare the clustering performance of these methods with a benchmark method k-means and a novel non-subspace based clustering algorithm rank-two nonnegative matrix factorization (R2NMF) [26].

This section consists of four parts. Section 1 introduces the datasets we used for experiments, and Section 2 shows the clustering results of different algorithms. The analyses of sensitivity of parameters and selection of in-sample data and the number of superpixels are in Sections 3 and 4, respectively.

#### 4.1. Instrument and Data

In order to evaluate the performance of the proposed method, here we obtained two polarized hyperspectral datasets. Dataset A was measured with a channeled interference imaging spectropolarimeter (CIISP) proposed by Zhang et al. The CIISP can simultaneously obtain four stokes vectors ( $S_0$ ,  $S_1$ ,  $S_2$  and  $S_3$ ) in 128 wave bands from 480 to 960 nm of a two-dimensional space target through one-dimensional push sweep. The experimental data was collected by the authors of this paper at a distance of 2 km, and the size of images is 512 × 512. Since the CIISP instrument is designed for spaceborne, collected images at this distance are somewhat blurred. This point posed a challenge to the clustering accuracy of the algorithms.

Dataset B that includes four polarization angles (0°, 45°, 90°, and 135°) in 33 bands from 400 to 720 nm is from the Northwestern Polytechnical University, and can be downloaded from the website (http://jszy.nwpu.edu.cn/zhaoyq.html).

After getting four stokes vectors, since  $S_3$  is usually negligible under natural conditions, the degree of linear polarization (DOLP) can be obtained by:

$$DOLP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0}.$$
 (15)

Then we can regard  $S_0$  as ordinary hyperspectral information and the DOLP as polarization information. Figure 3a–d shows the  $S_0$  and DOLP images of dataset A at 690 and 778 nm, and Figure 3e–h shows the  $S_0$  and DOLP images of dataset B at 480 and 690 nm.



**Figure 3.** Images of dataset A and dataset B: (**a**)  $S_0$  image of dataset A at 690 nm; (**b**) degree of linear polarization (DOLP) image of dataset A at 690 nm; (**c**)  $S_0$  image of dataset A at 778 nm; (**d**) DOLP image of dataset A at 778 nm; (**e**)  $S_0$  image of dataset B at 480 nm; (**f**) DOLP image of dataset B at 690 nm; (**h**) DOLP image of dataset B at 690 nm; (**i**) false color image of dataset A; (**j**) 3249 superpixels of dataset A; (**k**) false color image of dataset B; (**l**) 3249 superpixels of dataset B.

Based on the SLIC algorithm, three bands (RGB) of  $S_0$  were selected to construct a false color image, and then the false color image was segmented into thousands of superpixels. Note that for dataset A, since the spectral range (480–960 nm) of the CIISP does not include 435.8 nm which means red, the false color image is just an approximation. In the case where the number of superpixels is too small, there will be a sacrifice of image spatial resolution, which leads to a reduction in clustering accuracy. While if the number is too large, the computational time required for clustering will increase dramatically. In this paper, the number was placed at about 3000. Figure 3j,l shows the superpixel segmentation results.

#### 4.2. Clustering Results and Discussion

Dataset A includes five different kinds of targets: blue-gray painted wall; windows without curtains; windows with curtains; white painted wall and window edges; and leaves. Although the window edges are composed of metal, the surface is a layer of white paint, so it belongs to the same cluster as the white painted wall. Additonally, if there are curtains in the window, a large part of the light passing through the glass window will be reflected by the curtains, otherwise the light will enter the room and only a small part will be reflected back. Therefore, the existence of curtains will cause the difference in  $S_0$  and DOLP intensity. In addition, dataset B includes seven kinds of targets: white wall, blue wall, white smooth ground, black smooth ground, rough ground, vegetation 1 and vegetation 2.

The cluster maps of the various clustering methods of dataset A and dataset B are shown in Figures 4 and 5, respectively, and the corresponding quantitative evaluation of the clustering results is provided in Tables 2 and 3. Considering the limitations of the spatial resolution of the instruments themselves, it is difficult to evaluate the clustering performance at the boundary of the two targets and some vague areas, so these pixels which are unlabeled in Figures 4j and 5j were ignored when calculating the accuracy (AC).



**Figure 4.** Cluster maps of the different methods of dataset A: (**a**) k-means with  $S_0$  dataset only; (**b**) rank-two nonnegative matrix factorization (R2NMF) with  $S_0$  dataset only; (**c**) sparse subspace clustering (SSC) with  $S_0$  dataset only; (**d**) SSC with DOLP dataset only; (**e**) Polarized Spectral -SSC; (**f**) PS-low-rank representation subspace segmentation (LRR); (**g**) PS-least squares regression subspace segmentation (LSR); (**h**) in-sample clustering result of Fast Polarized Spectral (FPS)-SSC; (**i**) FPS-SSC; (**j**) FPS-LRR; (**k**) FPS-LSR; (**l**) ground truth.



**Figure 5.** Cluster maps of the different methods of dataset B: (**a**) k-means with S<sub>0</sub> dataset only; (**b**) R2NMF with S<sub>0</sub> dataset only; (**c**) SSC with S<sub>0</sub> dataset only; (**d**) SSC with DOLP dataset only; (**e**) PS-SSC; (**f**) PS-LRR; (**g**) PS-LSR; (**h**) in-sample clustering result of FPS-SSC; (**i**) FPS-SSC; (**j**) FPS-LRR; (**k**) FPS-LSR; (**l**) ground truth.

Method	Classes	k-Means	R2NMF	SSC-S <sub>0</sub>	SSC-DOLP	PS-SSC	PS-LRR	PS-LSR	FPS-SSC	FPS-LRR	FPS-LSR
	Blue-gray painted wall	69.93	81.67	73.24	10.05	63.18	62.94	57.14	80.70	84.31	79.63
	Windows	50.94	76.90	79.74	64.75	72.99	81.28	84.29	70.18	59.60	64.40
AC (%)	Windows with curtains	30.72	0.07	1.78	89.08	69.40	88.99	73.37	83.59	81.65	81.86
	White wall and window edges	45.94	52.18	65.71	13.50	64.60	66.12	76.52	71.93	75.85	76.77
	Leaves	73.60	77.07	98.58	97.46	97.69	96.12	95.44	96.34	94.87	96.04
OA (%)		57.49	63.25	66.02	35.84	69.07	72.46	71.02	79.44	80.40	79.36
Running time (s)		31.32	18.03	230.2	223.2	378.6	371.3	112.9	58.33	59.89	55.00

**Table 2.** Quantitative evaluation of the different clustering algorithms of dataset A.

**Table 3.** Quantitative evaluation of the different clustering algorithms of dataset B.

Method	Classes	k-means	R2NMF	SSC-S <sub>0</sub>	SSC-DOLP	PS-SSC	PS-LRR	PS-LSR	FPS-SSC	FPS-LRR	FPS-LSR
AC(%)	White wall	95.95	71.75	80.88	88.77	89.45	58.04	58.04	86.16	90.90	64.37
	Blue wall	0	74.02	71.69	0.00	0.00	81.89	75.09	63.59	69.19	74.87
	Black smooth ground	14.57	0	93.29	91.26	87.92	91.43	92.46	91.30	91.39	93.26
	White smooth ground	90.56	84.58	90.89	87.01	89.49	91.73	91.15	89.20	91.96	91.80
	Rough ground	2.54	81.62	81.57	95.82	93.87	81.52	62.99	93.71	86.20	76.55
	Vegetation 1	18.42	46.10	62.09	42.03	71.70	75.53	88.21	80.82	82.79	90.56
	Vegetation 2	0.51	10.90	89.57	51.05	89.48	85.61	72.42	82.22	79.12	65.31
OA (%)		50.11	64.61	75.70	64.04	80.68	80.20	81.46	85.10	85.28	84.58
Running time (s)		11.05	8.35	235.2	219.2	369.6	373.2	115.6	55.45	59.81	44.03

It can be obviously observed from Tables 2 and 3 that adding polarization into HSIs' clustering improved the overall accuracy (OA) significantly. For dataset A, directly applying the SSC algorithm over the  $S_0$  dataset can distinguish between blue-gray painted wall, windows without curtains, white painted wall and window edges, leaves, but the accuracy of windows with curtains is only 1.78%. While directly applying the SSC algorithm over the DOLP dataset can achieve a high accuracy of 89.08% on windows with curtains with low accuracy on blue-gray painted wall, windows without curtains, white painted wall and window edges. Both of two methods perform well on leaves, because leaves have distinct features on  $S_0$  and DOLP. As a combination of spectral information and polarized information, the PS-SSC algorithm has the advantages of SSC ( $S_0$ ) and advantages of SSC (DOLP) simultaneously. The accuracy of windows with curtains is 86.15% and the accuracy of other classes remains at a good level, which leads to a higher OA. It is similar for PS-LRR and PS-LSR that reach 72.46% and 71.02% of OA, respectively.

The dataset B does not have the window with curtains which is totally different from other targets in DOLP as in dataset A. Except for the edge contour, the DOLP intensity of different targets in dataset B is not significantly different. In this case, as can be seen from Table 3, adding polarization still improves the accuracy of clustering by about 5%.

However, on the other hand, due to the complexity of the optimization process, the time cost of these three methods is too high. The PS-SSC and the PS-LRR take more than 350 s of running time, and even though there is no iteration in solving the optimization problem in the PS-LSR algorithm, it still takes about 2 min. The k-means takes much less time than these subspace-based methods, but the OAs in both two datasets are much lower. Although the R2NMF can reach a higher OA than k-means with less running time, its OA is still not as good as subspace-based methods.

When the sampling-clustering-classification strategy is adopted, the running time is greatly reduced to less than 1 min. The FPS-SSC, FPS-LRR and FPS-LSR all achieved OAs of about 80% on dataset A and 85% on dataset B, which are even much higher than the PS-SSC, PS-LRR and PS-LSR. The PS version algorithms are equivalent to taking all 3249 superpixels as in-sample data in FPS version algorithms. Usually, for sampling strategies, the more sampling points there are, the higher the accuracy should be. However, the experimental results show a different trend. This is because the representation-based models themselves can achieve high accuracies on small-scale datasets, but on large-scale datasets, they may not perform as well as on small-scale ones. In addition to reducing the computational complexity, this is another reason why we adopt a sampling strategy to avoid clustering directly on polarized hyperspectral datasets. The relationship between OA and the number of selected in-sample data is discussed in Section 4.

### 4.3. Sensitivity of Parameters

Obviously, the various parameters required in the algorithms have an important impact on the final clustering results. In different practical application scenarios, the parameters need to be reset. It is desired for parameters to achieve the best performance by taking the same values on different datasets, so that users can easily set the parameters when applying these algorithms on various datasets collected by different instruments and different measurement environments.

In addition to the number of in-sample data, there are five trade-off parameters ( $\alpha$ ,  $\lambda_1$ ,  $\lambda_2$ ,  $\beta_1$  and  $\beta_2$ ) and two weight parameters ( $\sigma_s$  and  $\sigma_p$ ) in the proposed framework. For subspace clustering algorithms, regularization parameter  $\lambda$  is decided by the following formulation [13]:

$$\lambda = \frac{\lambda'}{\mu},\tag{16.a}$$

$$\mu = \min_{i} \max_{j \neq i} |y_i^T y_j|, \tag{16.b}$$

where  $\lambda'$  is the adjustment coefficient and  $\mu$  is fixed for a certain dataset. The five trade-off parameters, extending (16.a) and (16.b) to them, can be replaced with parameters not related to the dataset by:

$$\lambda_1 = \frac{\lambda'_1}{\min_{i} \max_{j \neq i} \left| Y_i^{(s)T} Y_j^{(s)} \right|},\tag{17.a}$$

$$\lambda_2 = \frac{\lambda'_2}{\min_{i} \max_{j \neq i} \left| Y_i^{(p)} Y_j^{(p)} \right|'}$$
(17.b)

$$\beta_1 = \frac{\beta'_1}{\min_{n} \max_{m \neq n} \left| O_{i_m}^{(s)T} I_n^{(s)} \right|'}$$
(17.c)

$$\beta_{2} = \frac{\beta'_{2}}{\min_{n} \max_{m \neq n} \left| O_{i_{m}}^{(p)} T I_{n}^{(p)} \right|},$$
(17.d)

$$\alpha = \alpha' \frac{\min_{i} \max_{j \neq i} \left| (W_p C^{(s)})_i^T (W_p C^{(s)})_j \right|}{\min_{i} \max_{j \neq i} \left| (W_s C^{(p)})_i^T (W_s C^{(p)})_j \right|}.$$
(17.e)

By changing one of those parameters while other parameters are fixed, the sensitivity of this parameter is determined. A change curve of the overall accuracy and in-sample accuracy with various values of different parameters is drawn in Figure 6.

Taking the OA of 79% or more of dataset A and the OA of 83% or more of dataset B as acceptable ranges, we can get the following phenomena from Figure 6:

(1)  $\lambda_1'$  performs well on dataset A and dataset B in the range of eight orders of magnitude (10–10<sup>8</sup>) with the peaks at 10<sup>3</sup>, and  $\lambda_2'$  can keep a high level of OA for both datasets in the range of 50–200;

(2)  $\beta_1'$  can reach the acceptable OA taking value from  $10^2$  to  $10^6$  for dataset A and from  $10^4$  to  $10^8$  for dataset B, and  $\beta_2'$  meets this requirement in eight orders of magnitude (10–10<sup>8</sup>) with the peaks at  $10^3$  for both datasets;

(3)  $\alpha'$  can achieve the highest OA for the two datasets in the range of 10–20, and the peaks appear when  $\alpha'$  is 15;

(4) For  $\sigma_s$  and  $\sigma_p$ , it is difficult to find fixed values for different datasets like the trade-off parameters, but they have a wide range of acceptable values on each dataset.

Summarizing the above points, we can reach the conclusion that the five trade-off parameters  $(\lambda_1', \lambda_2', \beta_1', \beta_2' \text{ and } \alpha')$  and the two weight parameters  $(\sigma_s \text{ and } \sigma_p)$  all have low parameter sensitivity. In addition, especially for the trade-off parameters, the best values on different datasets are very close.

#### 4.4. Selection of In-Sample Data and the Number of Superpixels

In the proposed framework, images are firstly segmented into lots of superpixels by the SLIC algorithm. Intuitively, the final OA as well as the running time is closely related to the number of superpixels. Figure 7a,b respectively indicates the OA and running time of FPS-SSC, FPS-LRR, FPS-LSR in the cases where the numbers of superpixels are 1024, 3249, 5329, 7225 and 10,404 when the number of in-sample points is 429. As the number of superpixels increases, the OA does not keep increasing but remains at a stable level when the number is greater than 3000. While, the running time grows up monotonically with the number of superpixels. Therefore, as a trade-off choice, the suggested number of superpixels for  $512 \times 512$  images is 2000–5000.



**Figure 6.** Analysis of parameters. Change in the overall accuracy and in-sample accuracy with various values of: (a)  $\lambda_1'$  of dataset A and dataset B; (b)  $\lambda_2'$  of dataset A and dataset B; (c)  $\beta_1'$  of dataset A and dataset B; (d)  $\beta_2'$  of dataset A and dataset B; (e)  $\alpha'$  of dataset A and dataset B; (f)  $\sigma_s$  of dataset A; (g)  $\sigma_s$  of dataset B; (h)  $\sigma_p$  of dataset A; (i)  $\sigma_p$  of dataset B.

The OA also depends on the selection method of the in-sample data. Figure 7c shows the comparison between random sampling and the superpixel-based method presented in Section 3, and the red dotted line indicates the mean and standard deviation of the results of applying FPS-SSC algorithm with random sampling for 10 times. Since the random sampling method is not in a position to ensure if the selection of in-sample data is suitable, it is accompanied by large standard deviation and great uncertainty. While applying the superpixel-based method, although its OA may be lower than the best results of random sampling, it achieves more than 75% of OA under different numbers of sampling data, and the results are much higher than the mean of the results of random sampling. Another benefit of the superpixel-based method is the stability, which means we only need one experiment to get a reliable clustering result.



**Figure 7.** Analysis of the number of superpixels and the selection of in-sample data: (**a**) change in overall accuracy (OA) with various numbers of superpixels; (**b**) change in running time with various numbers of superpixels; (**c**) comparison between random sampling and proposed method; (**d**) change in OA with various numbers of in-sample data; (**e**) change in In-sample Accuracy (IA) with various numbers of in-sample data; (**f**) change in running time with various numbers of in-sample data.

In addition, the clustering result relies on the number of in-sample data. The number of in-sample data affects the OA in two ways: (1) as the number increases, the OA generally shows an upward trend; (2) the number also affects the clustering accuracy of in-sample data, which is another important factor in OA. Figure 7d indicates the OA of FPS-SSC, FPS-LRR, FPS-LSR in the cases where the numbers of in-sample data are 214, 309, 429, 509, 622 and 764. Additionally, Figure 7e,f shows the corresponding In-sample Accuracy (IA) and running time, respectively. From the trend point of view, when the value of A is larger, the larger the OA. However, when the number takes values 509 and 622, the OA drops. This happens because when the number equals to 509 or 622, the accuracy of in-sample data is significantly reduced.

Additionally, as expected, the running time increases monotonically with the number of in-sample data. Following consideration of the cost of time, the suggested number of in-sample data is about one tenth of the number of superpixels. Specifically, for images with a size of  $512 \times 512$ , a high OA as well as relatively short running time can be achieved when the number of in-sample data is 400–500.

#### 5. Conclusions

In this paper, we added polarization into the clustering of hyperspectral images and proposed a representation-based subspace clustering framework for polarized hyperspectral images by treating each kind of land-cover class as a subspace. Furthermore, faced with the shortcoming that representation-based subspace clustering requires too much running time, an improved efficient version adopting the sampling-clustering-classification strategy was presented. Extensive experimental results clearly show that algorithms using proposed framework (FPS-SSC, FPS-LRR, FPR-LSR) can achieve a superior clustering performance with short running time and low sensitivity of parameters.

The proposed method can be further improved by the adaptive determination of the weight parameters and automatic determination of the cluster number by subspace detection, which will be in our future work.

**Author Contributions:** All the authors made significant contributions to this work. Z.C. devised the approach and analyzed the data; C.Z. and T.Y. designed the experiments and provided the data; T.M., Z.C. and Y.W. provided advice for the preparation and revision of the work.

**Funding:** This research was funded by the Major Program of National Natural Science Foundation of China (Grant No. 41530422), the 863 Program (Grant No. 2012AA121101), the National Natural Science Foundation of China (Grant No. 61775176, 61405153, 61501361), the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 32-Y30B08-9001-13/15), and the Fundamental Research Funds for the Central Universities of China (Grant No. xjj2017105).

Acknowledgments: The authors are grateful to the editors and the reviewers for their constructive comments that have helped improve this work significantly.

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

# References

- 1. Curran, P.J. A photographic method for the recording of polarized visible light for soil surface moisture indications. *Remote Sens. Environ.* **1978**, *7*, 305–322. [CrossRef]
- 2. Deschamps, P.-Y. The POLDER mission: Instrument characteristics and scientific objectives. *IEEE Trans. Geosci. Remote Sens.* **1994**, 32, 598–615. [CrossRef]
- 3. Wolff, L.B. Polarization-based material classification from specular reflection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 1059–1071. [CrossRef]
- 4. Wolff, L.B.; Boult, T.E. Constraining object features using a polarization reflectance model. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 635–657.
- 5. Palubinskas, G.; Datcu, M.; Pac, R. Clustering algorithms for large sets of heterogeneous remote sensing data. *Proc. IGARSS* **1999**, *3*, 1591–1593.
- 6. Barni, M.; Garzelli, A.; Mecocci, A.; Sabatini, L. A robust fuzzy clustering algorithm for the classification of remote sensing images. *Proc. IGARSS* **2000**, *5*, 2143–2145.
- 7. Paoli, A.; Melgani, F.; Pasolli, E. Clustering of Hyperspectral Images Based on Multiobjective Particle Swarm Optimization. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 4175–4188. [CrossRef]
- 8. Li, S.; Zhang, B.; Li, A.; Jia, X. Hyperspectral imagery clustering with neighborhood constraints. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 588–592. [CrossRef]
- 9. Bo, R.; Biao, H. Sparse Subspace Clustering-Based Feature Extraction for PolSAR Imagery Classification. *Remote Sens.* **2018**, *10*, 391.
- 10. Murphy, J.M.; Maggioni, M. Unsupervised Clustering and Active Learning of Hyperspectral Images with Nonlinear Diffusion. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1829–1845. [CrossRef]
- 11. Saranathan, A.M.; Parente, M. On Clustering and Embedding Mixture Manifolds using a Low Rank Neighborhood Approach. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3890–3903. [CrossRef]
- 12. Chunmin, Z.; Qiwei, L. High throughput static channeled interference imaging spectropolarimeter based on a Savart polariscope. *Opt. Express* **2016**, *24*, 23314–23332.
- 13. Elhamifar, E.; Vidar, R. Sparse subspace clustering: Algorithm, theory, and application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2765–2781. [CrossRef] [PubMed]
- 14. Liu, G.C.; Lin, Z.C.; Yu, Y. Robust subspace segmentation by low-rank representation. In Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel, 21–24 June 2010; pp. 663–670.
- 15. Canyi, L.; Hai, M. Robust and efficient subspace segmentation via least squares regression. In Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 347–360.
- 16. Zhang, H.; Zhai, H.; Zhang, L.; Li, P. Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3672–3684. [CrossRef]
- 17. Zhai, H.; Zhang, H.; Zhang, L.; Li, P.; Plaza, A. A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 43–47. [CrossRef]
- Afonso, M.V.; Bioucas-Dias, J.M.; Figueiredo, M.A.T. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.* 2011, 20, 681–695. [CrossRef] [PubMed]
- 19. Mota, J.F.; Xavier, J.W.F.; Aguiar, P.M.Q.; Puschel, M. D-ADMM: A communication-efficient distributed algorithm for separable optimization. *IEEE Trans. Signal Process.* **2013**, *61*, 2718–2723. [CrossRef]

- 20. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2000, 22, 888–905.
- 21. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On spectral clustering: Analysis and an algorithm. *Proc. Adv. Neural Inf. Process. Syst.* **2002**, *2*, 849–856.
- 22. Zhang, G.; Jia, X.; Kwok, N.M. Super pixel based remote sensing image classification with histogram descriptors on spectral and spatial data. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 4335–4338.
- 23. Zhang, G.; Jia, X.; Hu, J. Superpixel-based graphical model for remote sensing image mapping. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5861–5871. [CrossRef]
- 24. Radhakrishna, A. SLIC Superpixels Compared to State-of-the-art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282.
- 25. Peng, X.; Zhang, L.; Yi, Z. Scalable sparse subspace clustering. In Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 430–437.
- 26. Gillis, N.; Kuang, D.; Park, H. Hierarchical Clustering of Hyperspectral Images Using Rank-Two Nonnegative Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2066–2078. [CrossRef]



© 2019 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 (http://creativecommons.org/licenses/by/4.0/).