

Article

# Nonlinear Classification of Multispectral Imagery Using Representation-Based Classifiers

Yan Xu <sup>1</sup>, Qian Du <sup>1,\*</sup>, Wei Li <sup>2</sup>, Chen Chen <sup>3</sup> and Nicolas H. Younan <sup>1</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762, USA; yx131@msstate.edu (Y.X.); younan@ece.msstate.edu (N.H.Y.)

<sup>2</sup> College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China; liwei089@ieee.org

<sup>3</sup> Center for Research in Computer Vision, University of Central Florida, Orlando, FL 32816, USA; chenchen870713@gmail.com

\* Correspondence: du@ece.msstate.edu; Tel.: +1-662-325-2035

Received: 10 May 2017; Accepted: 25 June 2017; Published: 28 June 2017

**Abstract:** This paper investigates representation-based classification for multispectral imagery. Due to small spectral dimension, the performance of classification may be limited, and, in general, it is difficult to discriminate different classes with multispectral imagery. Nonlinear band generation method with explicit functions is proposed to use which can provide additional spectral information for multispectral image classification. Specifically, we propose the simple band ratio function, which can yield better performance than the nonlinear kernel method with implicit mapping function. Two representation-based classifiers—i.e., sparse representation classifier (SRC) and nearest regularized subspace (NRS) method—are evaluated on the nonlinearly generated datasets. Experimental results demonstrate that this dimensionality-expansion approach can outperform the traditional kernel method in terms of high classification accuracy and low computational cost when classifying multispectral imagery.

**Keywords:** multispectral imagery; nonlinear classification; kernel method; dimensionality expansion

## 1. Introduction

Airborne and spaceborne optical remote sensors collect useful information from the Earth's surface based on the radiance reflected by different materials. Hyperspectral sensors acquire images at contiguous spectral ranges with high spectral resolution. On the contrary, multispectral sensors acquire only several wide bands with high spatial resolution. The high spectral resolution of hyperspectral imagery provides major advantages for classification and detection. However, due to the high dimensionality, its vast data volume can cause issues in data transmission, storage, and analysis [1,2]. Although multispectral imagery has low spectral resolution and it may be difficult to distinguish materials with similar spectral signatures, its high spatial resolution and wide coverage make it still popular in practical applications.

Recently, sparse representation classifier (SRC) [3] and collaborative representation classifier (CRC) [4] have gained much attention for hyperspectral imagery classification. Different from the traditional classifiers, such as support vector machine (SVM), these representation-based classifiers do not use the training-testing fashion. Instead, in these methods, a testing pixel is classified based on representation residual using labeled samples. The nearest regularized subspace (NRS) [5] is an improved version of CRC, where samples similar to the testing pixels are allowed to have high weights in the representation. Other variants of SRC or CRC have been proposed for hyperspectral imagery. For example, in [6], a local sparse representation-based nearest neighbor is proposed to increase the performance by utilizing class-specific sparse coefficients. A weighted joint collaborative

representation based classifier is presented in [7], which adopts more appropriate weights by considering the similarity between the centered pixel and its surrounding pixels. Bian et al. proposed a multi-layer spatial-spectral representation framework for hyperspectral classification [8]. NRS is implemented as a class-specific version by using samples of each class separately in [9], and it is performed on Gabor features in [10], yielding improved classification accuracy. Representation-based approaches for hyperspectral classification and detection are summarized in [11]. However, the performance of such representation-based classifiers in multispectral image classification is limited, because the low-dimensional pixel vectors cannot offer significant discrepancy in representation residual when using training samples of different classes, producing ambiguity in label assignment.

As a classical feature expansion approach, the kernel method has been successfully applied to hyperspectral and multispectral classification. Using the kernel trick, it maps the original data to a high dimensional feature space without the need of knowing the actual mapping function. Kernel SVM (KSVM) is applied for hyperspectral image classification, which has been considered as a standard classifier [12]. Bernabe et al. employed kernel principal component analysis to extend the original principal component analysis to a nonlinear version [13]. Kernel collaborative representation with Tikhonov regularization (denoted as KNRS) is presented in [14], and Kernel sparse representation classifier (KSRC) is developed in [15]. The difficulties of the traditional kernel methods include high computational cost in the computation of Gram matrix and exhaustive searching in parameter tuning.

In this paper, we propose to use a simple strategy to generate artificial bands for multispectral imagery classification. The goal of this approach is to use explicit nonlinear functions to contrast the dissimilarity between original spectral measurements, which can provide additional spectral information for classification problems [16]. By generating new artificial bands, the spectral contrast between different classes can be increased. Our major contribution is to use the simple band ratio as the explicit nonlinear function for dimensionality expansion, which can offer better performance than the traditional kernel method in terms of high classification accuracy and low computational cost. Here, we limit the discussion in representation-based classifiers, although the discussed band expansion can be applicable to any other classifier.

The rest of this paper is organized as follows. Section 2 introduces the two representation-based classifiers, i.e., SRC and NRS. Section 3 presents the simple nonlinear band generation method. Section 4 discusses experimental result. The conclusion is drawn in Section 5.

## 2. Representation-Based Algorithms

Let the dataset with  $n$  labeled samples in  $c$  classes be  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c\} \in \mathbb{R}^{d \times n}$ , where  $d$  is the number of bands and  $\mathbf{X}_i$  includes labeled samples for the  $i$ -th class.

### 2.1. SRC

In SRC [3], a testing sample  $\mathbf{y}$  is linearly represented by all the training samples. The objective is to find a sparse weight vector  $\mathbf{a}$  that minimizes the term  $\|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2$ , i.e.,

$$\arg \min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1 \quad (1)$$

where  $\lambda$  is the regularization term. In this research, Equation (1) is solved by mexLassoWeighted.m in MATLAB [17].

After the sparse weight vector  $\mathbf{a}$  is estimated, the residual error for each class  $i$  is calculated as

$$r_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{X}_i \mathbf{a}_i\|_2^2 \quad (2)$$

where  $\mathbf{a}_i$  denotes the entries of sparse weight vector  $\mathbf{a}$  associated with the  $i$ -th class. The testing sample is assigned as

$$\text{class}(\mathbf{y}) = \arg \min_{i=1,2,\dots,c} (r_i(\mathbf{y})) \quad (3)$$

## 2.2. NRS

It has been argued that it is the collaborative representation instead of the  $l_1$  norm that actually improves the classification accuracy [4]. The NRS [5,9,10] can adaptively adjust the regularization term per sample such that only samples similar to the testing sample can actually participate in collaborative representation. Its objective function is expressed as

$$\arg \min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 + \lambda \Gamma \|\mathbf{a}\|_2^2 \quad (4)$$

where  $\Gamma$  is a diagonal matrix, which is defined as

$$\Gamma = \begin{bmatrix} \|(\mathbf{y} - \mathbf{x}^{(1)})\|_2^2 & & 0 \\ & \ddots & \\ 0 & & \|(\mathbf{y} - \mathbf{x}^{(n)})\|_2^2 \end{bmatrix} \quad (5)$$

where  $\mathbf{x}^{(i)}$  is the  $i$ -th column of the dictionary  $\mathbf{X}$ . The coefficient  $\mathbf{a}$  has a closed-form solution as

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X} + \lambda \Gamma)^{-1} \mathbf{X}^T \mathbf{y} \quad (6)$$

Similarly, the residual error is used to determine the class label as Equation (3).

In SRC and NRS, the regularization parameter  $\lambda$  needs to be tuned, which can be the optimal value for the training samples.

## 3. Nonlinear Band Generation Method

A simple way for band generation is to adopt some explicit nonlinear functions to create artificial images that serve as additional linearly independent spectral measurements [16,18]. Although any nonlinear functions may be used, in our paper, we limit our discussion on multiplication and division. Band multiplication is related to their correlation, while band ratio is often used to remove the illumination factor [19]. Three new datasets can be generated by these two methods. The first dataset uses pixel-wise multiplication, and the second dataset is generated by division, i.e., band ratio. If we combine the original dataset with the artificial bands by both division and multiplication, we have the third dataset with a total number of  $N^2$  bands.

Note that, in the traditional kernel method, the kernel trick is to avoid to explicitly identify the nonlinear functions to use. We will show that simple multiplication and division can offer better classification than the kernel trick, and band ratio (division) is the best choice for the nonlinear function while keeping the data dimensionality manageable.

### 3.1. Multiplication

Suppose two images  $\mathbf{B}_i$  and  $\mathbf{B}_j$  (pixels at the same locations) are multiplied together, then a new image  $\{\mathbf{B}_i \mathbf{B}_j\}_{i=1, j=i+1}^{N-1}$  is produced, where  $N$  is the total number of bands of the original multispectral imagery. Although multiplication can be used for a single band, we only apply multiplication to each pair of bands in order to compare with the division method yielding the same number of bands, i.e.,  $\{\mathbf{B}_i / \mathbf{B}_j\}_{i=1, j=i+1}^{N-1}$ . Combining the original multispectral dataset with the generated artificial bands with multiplication, there are a total of  $N^2/2 + N/2$  bands.

### 3.2. Division

New bands can be created as  $\{\mathbf{B}_i / \mathbf{B}_j\}_{i=1, j=i+1}^{N-1}$  by dividing the pixels at the same locations in the original bands in the multispectral dataset. If we only combine the original dataset with the bands generated by division, we get the second dataset. The total number of bands after combining the

original bands is  $N^2/2 + N/2$  in the second dataset. If we combine the original dataset with the artificial bands by both division and multiplication, we have the third dataset with a total number of  $N^2$  bands.

The proposed framework is shown in Figure 1, which includes the comparison of four cases: original bands, original bands and bands generated with multiplication (original + multiplication), original bands and bands generated with division (original + division), original bands and bands generated with both multiplication and division (original + multiplication + division).

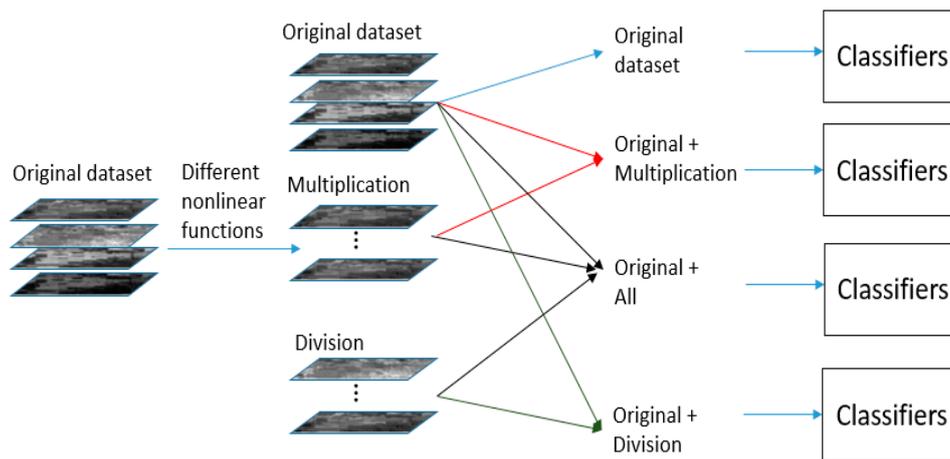


Figure 1. Framework of the band generation method.

### 3.3. Practical Consideration

In order for the generated bands to have similar dynamic ranges as the original bands, data is normalized by dividing the maximum value; in other words, after normalization, the maximum value of all the data points becomes 1. In division, the band with the larger local maximum value is chosen as the divider, or the band with non-zero minimum value is the divider.

In practice, the value of 0 often occurs at the same pixel locations, such as shadow pixels in all the bands. Then the band ratio is set to be 0. However, for pixels with very small non-zero values, such as water pixels, it may be needed to introduce a small constant in both denominator and numerator as [20]:  $\{(\mathbf{B}_i + K)/(\mathbf{B}_j + K)\}_{i=1, j=i+1}^{N-1}$ . Note that due to spectral correlation, the materials (e.g., water, shadow) consistently have low or zero reflectance values without sudden change.

## 4. Experiment Results

### 4.1. Data Description and Experimental Setup

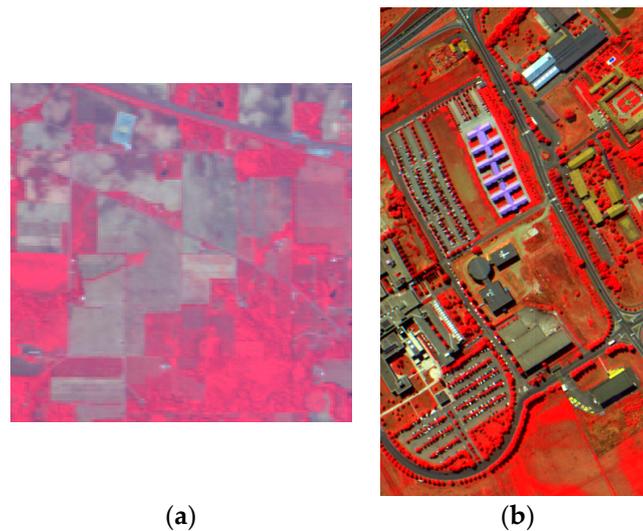
Due to lack of multispectral images with pixel level ground truth, data used in the experiments are simulated from hyperspectral images through band grouping.

The first multispectral dataset is simulated from hyperspectral Indian Pines dataset acquired by the Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) over the Indiana's Indian Pines in June 1992. The spatial size is  $145 \times 145$  with the spatial resolution 20 m/pixel, and the 220 spectral bands are from 0.4 to 2.5  $\mu\text{m}$ . We generate six bands from this dataset since it has wider spectral range. The generated six bands are to simulate blue, green, red, near infrared, short wave infrared channels by grouping band range 6~12, 13~21, 24~33, 40~54, 123~143, and 177~220 of the Indian Pines dataset [21]. Using the technique in Section 3, 15 bands are generated with multiplication, and another 15 bands are generated with division. There are, in total, 16 different classes from the original ground truth; however, we select eight classes from the original dataset from a statistic viewpoint [5]. The eight classes we used in the experiments are *Corn-no-till*, *Corn-min-till*, *Grass-pasture*, *Hay-windowed*,

*Soybean-no-till, Soybean-min-till, Soybean-clean, and woods*. The number of labeled samples are tabulated in Table 1. The false color-infrared image of this dataset is shown in Figure 2a.

**Table 1.** Number of samples per class for Indian Pines Dataset (the eight classes studied are bolded).

Class No.	Class Name	Number of Samples
C1	Alfalfa	46
<b>C2</b>	<b>Corn-no-till</b>	1460
<b>C3</b>	<b>Corn-min-till</b>	834
C4	Corn	237
<b>C5</b>	<b>Grass-pasture</b>	483
C6	Grass-trees	730
C7	Grass-pasture-mowed	28
<b>C8</b>	<b>Hay-windowed</b>	478
C9	Oats	20
<b>C10</b>	<b>Soybean-no-till</b>	972
<b>C11</b>	<b>Soybean-min-till</b>	2455
<b>C12</b>	<b>Soybean-clean</b>	593
C13	Wheat	205
<b>C14</b>	<b>Woods</b>	1265
C15	Building-grass-trees-drives	386
C16	Stone-steel-towers	93
Total		10,249



**Figure 2.** Color-infrared composites for (a) Indian Pines Dataset; (b) University of Pavia dataset.

The second multispectral dataset is generated from a hyperspectral image acquired by the reflective optics system imaging spectrometer (ROSIS) sensor. The image scene, covering the University of Pavia, has 115 spectral bands ranging from 0.43 to 0.86  $\mu\text{m}$  with the spatial size of  $610 \times 340$  pixels, and the spatial resolution is 1.3 m per pixel. This dataset consists of 102 spectral bands after removing the 12 noisy bands. We generate four bands from this dataset according to [22]. Four bands—i.e., blue, green, red, near infrared channels—are simulated by grouping band range 6~24, 25~45, 54~69, and 89~103 in the original hyperspectral dataset. Based on these four bands, six bands are generated with multiplication, and another six with division. The number of labeled samples of nine classes are shown in Table 2. The false color-infrared image of this dataset is shown in Figure 2b.

**Table 2.** Number of samples per class for University of Pavia Dataset.

Class No.	Class Name	Number of Samples
C1	Asphalt	6631
C2	Meadows	18,649
C3	Gravel	2099
C4	Trees	3064
C5	Painted metal sheets	1345
C6	Bare Soil	5029
C7	Bitumen	1330
C8	Self-Blocking Bricks	3682
C9	Shadows	947
Total		42,776

#### 4.2. Classification Results

The datasets using nonlinear band generation method are evaluated on SRC, NRS, their kernel versions with kernel trick (i.e., KSRC, KNRS), and KSVM. Each experiment is conducted 10 times to avoid any bias in sampling, and average performance of overall accuracy (OA) is reported. The number of training samples are set to 10, 30, 50, 70, 90, and 110, which are randomly selected. The regularization parameter  $\lambda$  is critical to the performance of the two classifiers, and we adopt 10-fold cross validation to choose the  $\lambda$ . Figures 3 and 4 show the thematic maps from the NRS for Indian Pines and University of Pavia datasets, respectively. Obviously, there are many misclassified pixels. However, it can be observed that the maps using the original bands only are worse than others.

Figure 5 shows the results for the datasets generated by Indian Pines. We conclude the division method provides the best performance among other band generation methods. The OA using the division method for both classifiers increases approximately 7%, compared to using the original data only. Combining multiplication and division can provide approximately the same performance as using the division only. The KSRC performs slightly better than the original SRC. When the number of training samples is large, the KNRS outperforms the original NRS. However, when the number of training samples becomes small, the KNRS may be even worse than the linear NRS. The KSVM using the original multispectral imagery is inferior to SRC or NRS on the generated bands. The advantage of using generated bands is more obvious when the number of training samples is small, which may be because the dimensionality is expanded to a reasonable level.

Figure 6 presents the SRC and NRS results for the University of Pavia dataset. For SRC, using nonlinear bands outperforms KSRC with the change of training samples. The three datasets containing nonlinearly generated bands provide comparable performance. When the number of training samples is small, the KSRC offers similar performance as the SRC on the original dataset. However, when the number of training sample increases, the KSRC provides much better accuracy than the linear SRC. For the case of NRS, with a small number of training samples, the KNRS produces approximately the same performance as its linear version. When the training size is small, using nonlinear bands can outperform the KSVM; using nonlinear bands can provide an approximately similar performance as KSVM when the training size is large.

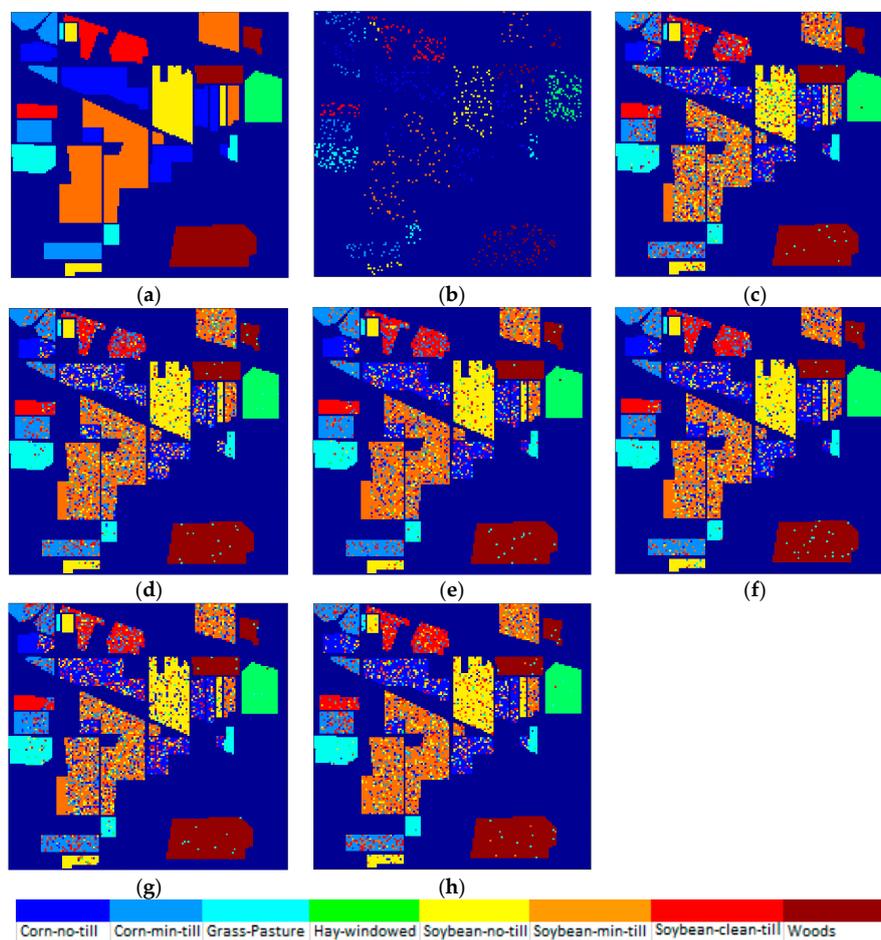
Tables 3 and 4 provide the computation cost of different algorithms in MATLAB when the training sample is 110 per class. The computer has 3.40 GHz CPU and 16.0 GB RAM. We conclude that the KSRC is computationally expensive compared to the original SRC. If bands are nonlinearly generated for the SRC, then the computational cost is only slightly higher than using the original bands. The discrepancy on computational cost between NRS and KNRS is less significant. However, KNRS costs more time than the method using NRS on the generated datasets. The KSVM is the most time consuming approach. Compared with the NRS and SRC approaches, the KSVM is more computationally expensive.

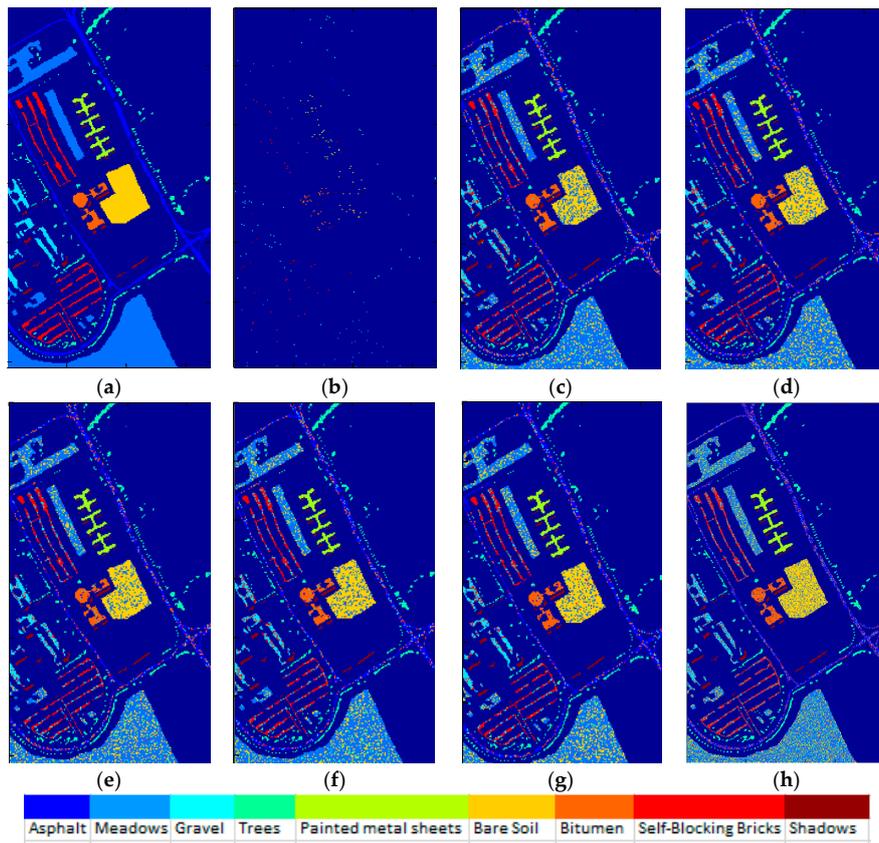
**Table 3.** Computing time (in seconds) in multispectral Indian Pines dataset using 110 samples per class.

Datasets	SRC	KSRC	NRS	KNRS	KSVM
Original	50.49	311.89	122.70	152.81	1572.29
Original + Multiplication	54.84	—	131.27	—	—
Original + Division	56.78	—	135.90	—	—
Original + Multiplication + Division	57.88	—	137.05	—	—

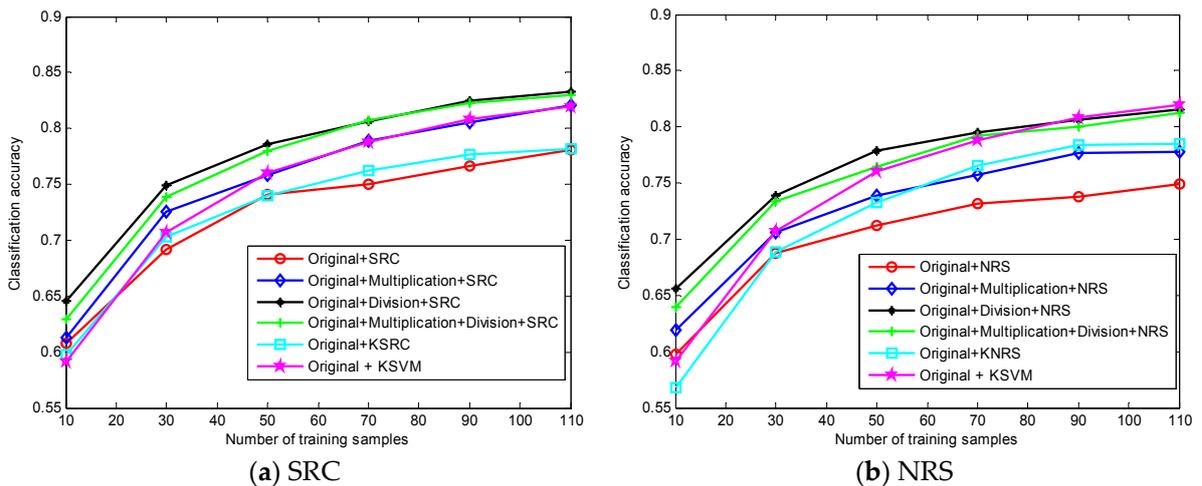
**Table 4.** Computing time (in seconds) the multispectral University of Pavia dataset using 110 samples per class.

Datasets	SRC	KSRC	NRS	KNRS	KSVM
Original	228.54	2046.9	592.97	794.09	2122.59
Original + Multiplication	240.34	—	611.35	—	—
Original + Division	245.34	—	604.75	—	—
Original + Multiplication + Division	251.48	—	620.75	—	—

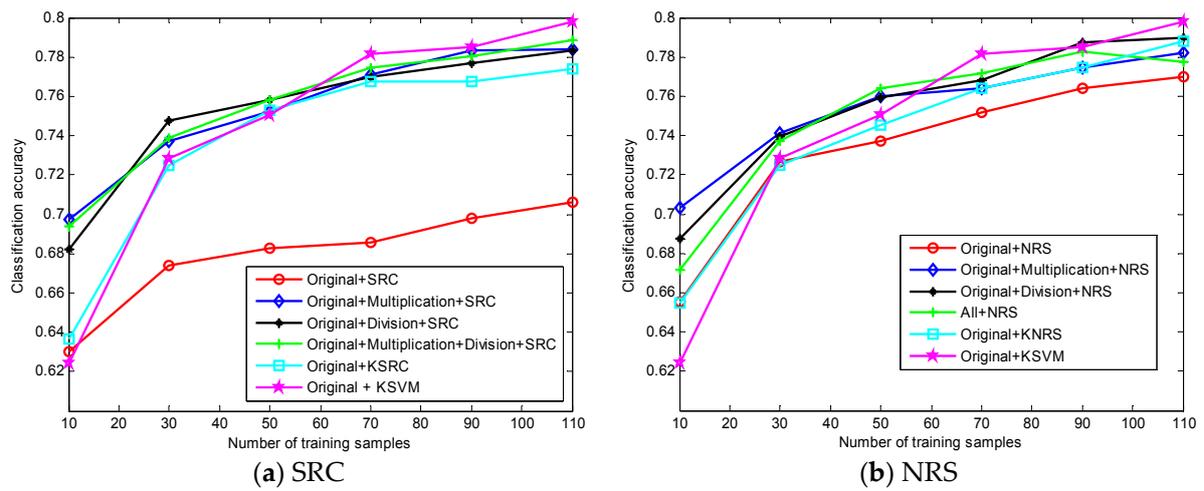
**Figure 3.** Thematic maps using 110 samples per class for the multispectral Indian Pines dataset with eight classes (and OA values). (a) Ground truth; (b) Training; (c) Original + NRS (0.7492); (d) Original + Multiplication + NRS (0.7781); (e) Original + Division + NRS (0.8159); (f) Original + Multiplication + Division + NRS (0.8124); (g) Original + KNRS (0.7852); (h) Original + KSVM (0.8193).



**Figure 4.** Thematic maps using 110 samples per class for the multispectral University of Pavia dataset with nine classes (and OA values). (a) Ground truth; (b) Training; (c) Original + NRS (0.7698); (d) Original + Multiplication + NRS (0.7820); (e) Original + Division + NRS (0.7896); (f) Original + Multiplication + Division + NRS (0.7880); (g) Original + KNRS (0.7736); (h) Original + KSVM (0.7981).



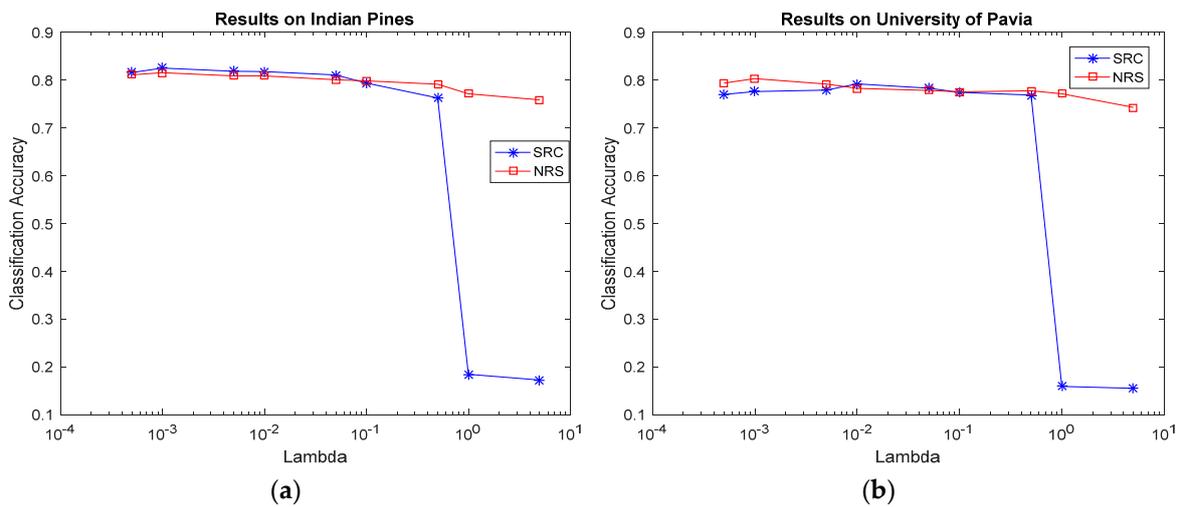
**Figure 5.** Classification on the multispectral dataset generated from the hyperspectral Indian Pines dataset.



**Figure 6.** Classification on the multispectral dataset generated from the hyperspectral University of Pavia dataset.

4.3. Parameter Tuning

The parameter  $\lambda$  is important to the representation-based classifiers. In this session, we present the effects of different  $\lambda$  on both Indian Pines and the University of Pavia datasets using NRS and SRC. Figure 7a,b show the classification accuracy changes with  $\lambda$  in Indian Pines and Pavia University datasets, respectively. The training samples are set to be 90 per class, and each experiment is conducted 10 times to estimate the average results. Since the Original + Division provides better performance with less computational cost, we test the effects of different  $\lambda$  on its generated dataset. We can conclude a relatively small  $\lambda$ , e.g.,  $10^{-2}$ , can guarantee satisfactory performance for both NRS and SRC. Obviously, NRS is less sensitive to  $\lambda$  due to the fact that the  $\Gamma$  matrix can adaptively adjust the penalty according to the similarity between the training and testing pixels.

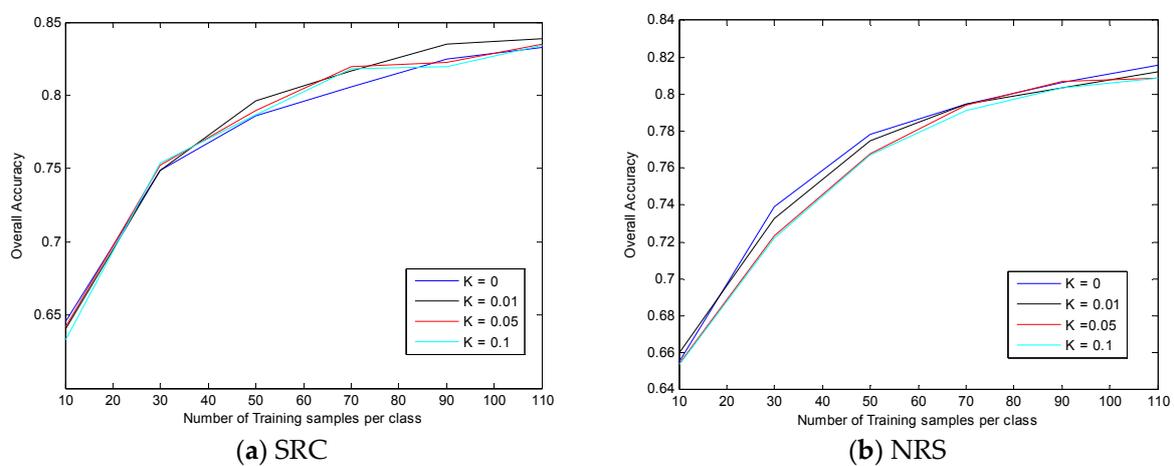


**Figure 7.** Classification Accuracy with different  $\lambda$  using NRS and SRC for: (a) multispectral Indian Pines; and (b) multispectral University of Pavia datasets.

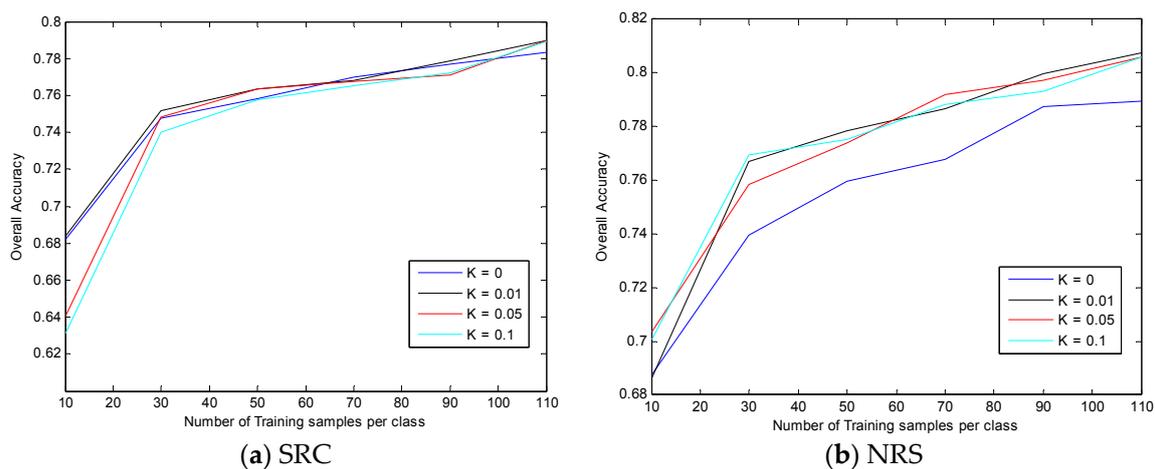
In KSRC and KCRC, the radial basis function (RBF) is chosen as the kernel function. According to [12], the parameter  $\gamma$  of the kernel function is set as the median value of  $1/(||\mathbf{x}_i - \bar{\mathbf{x}}||_2^2)$ ,  $i = 1, 2, \dots, n$ , where  $\bar{\mathbf{x}} = (1/n)\sum_{i=1}^n \mathbf{x}_i$  is the mean of all available training samples. This simple strategy offers a similar performance as using the parameter tuned by cross-validation. For the RBF kernel in the KSVM, we choose the parameter  $\gamma$  and regularization parameter  $C$  with cross-validation.

#### 4.4. Modified Band Ratio

To avoid a very small divider when calculating band ratio, a constant value of  $K$  can be added to both numerator and denominator. Figures 8 and 9 show the results for the Indian Pines and University of Pavia datasets. Since the minimum value of the Indian Pines data is about 0.12 (after normalization), the original version of band ratio with  $K = 0$  may be sufficient. In the University of Pavia dataset with many close-to-zero values, this strategy can improve the performance. Overall, a small value of  $K$ , such as  $K = 0.01$ , is an appropriate choice for both SRC and NRS.



**Figure 8.** Classification on the multispectral Indian Pines dataset using the original plus division-generated bands (original + division) with different adjustment parameter  $K$ .



**Figure 9.** Classification on the multispectral University of Pavia dataset using the original plus division-generated bands (original + division) with different adjustment parameter  $K$ .

## 5. Conclusions

This paper proposes to use nonlinear band generation method with explicit functions for multispectral classification. Two classifiers, i.e., SRC and NRS, and their kernel versions are evaluated on the new datasets. The experimental results show that this method performs better than the traditional kernel methods with higher classification accuracy and much lower computational cost. In particular, it can outperform when the number of training samples is small.

The difficulty of nonlinear band generation is choosing an appropriate nonlinear function for different datasets collected by various sensors covering all kinds of image scenes. In our experiments, it turns out that the band ratio offers the best performance. Considering its role in removing illumination factor [19], it would be a reasonable choice. Modified band ratio with a small adjustment parameter may further improve the performance when an image scene contains materials with very low reflectance.

**Author Contributions:** Yan Xu and Qian Du designed the experiments, prepared the first draft, and edited the manuscript. Wei Li, Chen Chen, and Nicolas H. Younan reviewed and edited the manuscript.

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

## References

1. Xu, Y.; Du, Q.; Younan, N. Particle swarm optimization-based band selection for hyperspectral target detection. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 554–558. [[CrossRef](#)]
2. Su, H.; Yang, H.; Du, Q.; Sheng, Y. Semi-supervised band clustering for dimensionality reduction of hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1135–1139. [[CrossRef](#)]
3. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE PAMI* **2009**, *31*, 210–227. [[CrossRef](#)] [[PubMed](#)]
4. 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.
5. Li, W.; Tramel, E.W.; Prasad, S.; Fowler, J.E. Nearest regularized subspace for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 477–489. [[CrossRef](#)]
6. Zou, J.; Li, W.; Du, Q. Sparse representation-based nearest neighbor classifiers for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2418–2422.
7. Xiong, M.; Ran, Q.; Li, W.; Zou, J.; Du, Q. Hyperspectral image classification using weighted joint collaborative representation. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1209–1213. [[CrossRef](#)]
8. Bian, X.; Chen, C.; Xu, Y.; Du, Q. Robust hyperspectral image classification by multi-layer spatial-spectral sparse representations. *Remote Sens.* **2016**, *8*, 985. [[CrossRef](#)]
9. 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](#)]
10. Li, W.; Du, Q. Gabor-filtering based nearest regularized subspace for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1012–1022. [[CrossRef](#)]
11. Li, W.; Du, Q. A survey on representation-based classification and detection in hyperspectral imagery. *Pattern Recognit. Lett.* **2016**, *83*, 115–123. [[CrossRef](#)]
12. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [[CrossRef](#)]
13. Bernabe, S.; Marpu, P.R.; Plaza, A.; Mura, M.D.; Benediktsson, J.A. Spectral-spatial classification of multispectral images using kernel feature space representation. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 288–292. [[CrossRef](#)]
14. Li, W.; Du, Q.; Xiong, M. Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 48–52.
15. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification via kernel sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 217–231. [[CrossRef](#)]

16. Ren, H.; Chang, C.-I. A generalized orthogonal subspace projection approach to unsupervised multispectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 2515–2528.
17. SPARse Modeling Software. Available online: <http://spams-devel.gforge.inria.fr/> (accessed on 15 October 2016).
18. Du, Q.; Kopriva, I.; Szu, H. Independent component analysis for classifying multispectral images with dimensionality limitation. *Int. J. Inf. Acquis.* **2004**, *1*, 201–216. [[CrossRef](#)]
19. Lillesand, T.; Kiefer, R.W. *Remote Sensing and Image Interpretation*; Wiley: Hoboken, NJ, USA, 2015.
20. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Proc.* **2004**, *13*, 600–612. [[CrossRef](#)]
21. Platt, R.V.; Goetz, A.F.H. A comparison of AVIRIS and Landsat for land use classification at the urban fringe. *Photogram. Eng. Remote Sens.* **2004**, *70*, 813–819. [[CrossRef](#)]
22. Kramer, H.J. *Observations of the Earth and Its Environment*; Springer: Berlin, Germany, 2002.



© 2017 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/>).