

Article

Discriminative Feature Metric Learning in the Affinity Propagation Model for Band Selection in Hyperspectral Images

Chen Yang ^{1,2,*} , Yulei Tan ¹, Lorenzo Bruzzone ³, Laijun Lu ¹ and Renchu Guan ^{4,*}

¹ College of Earth Sciences, Jilin University, Changchun 130061, China; tanyl0511@163.com (Y.T.); lulj1956@163.com (L.L.)

² Lab of Moon and DeepSpace Exploration, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

³ Department of Information Engineering and Computer Science, University of Trento, 38050 Trento, Italy; lorenzo.bruzzone@ing.unitn.it

⁴ College of Computer Science and Technology, Jilin University, Changchun 130012, China

* Correspondence: yangc616@jlu.edu.cn (C.Y.); guanrenchu@jlu.edu.cn (R.G.); Tel.: +86-431-8580-3938 (C.Y.)

Received: 26 May 2017; Accepted: 27 July 2017; Published: 30 July 2017

Abstract: Traditional supervised band selection (BS) methods mainly consider reducing the spectral redundancy to improve hyperspectral imagery (HSI) classification with class labels and pairwise constraints. A key observation is that pixels spatially close to each other in HSI have probably the same signature, while pixels further away from each other in the space have a high probability of belonging to different classes. In this paper, we propose a novel discriminative feature metric-based affinity propagation (DFM-AP) technique where the spectral and the spatial relationships among pixels are constructed by a new type of discriminative constraint. This discriminative constraint involves chunklet and discriminative information, which are introduced into the BS process. The chunklet information allows for grouping of spectrally-close and spatially-close pixels together without requiring explicit knowledge of their class labels, while discriminative information provides important separability information. A discriminative feature metric (DFM) is proposed with the discriminative constraints modeled in terms of an optimal criterion for identifying an efficient distance metric learning method, which involves discriminative component analysis (DCA). Following this, the representative subset of bands can be identified by means of an exemplar-based clustering algorithm, which is also known as the process of affinity propagation. Experimental results show that the proposed approach yields a better performance in comparison with several representative class label and pairwise constraint-based BS algorithms. The proposed DFM-AP improves the classification performance with discriminative constraints by selecting highly discriminative bands with low redundancy.

Keywords: hyperspectral imagery; chunklet and discriminative information; discriminative feature metric; affinity propagation; band selection

1. Introduction

Hyperspectral imagery (HSI) can acquire images in hundreds of narrow and adjacent spectral channels [1], which provide rich information to distinguish different ground objects and have a high potential for detailed land cover classification. However, this also involves several significant challenges for HSI classification due to the following critical factors: (1) the Hughes phenomenon [2,3] (also known as the curse of dimensionality), involving the classification accuracy possibly decreasing with an increase in the number of spectral bands when a limited number of labeled samples is available;

(2) high correlation between spectral bands bringing a large amount of redundant information, which has a negative effect on classification performance; and (3) high spatial resolution increasing the intraclass variability and decreasing the interclass variability, which thus reduces the statistical separation between different land cover classes in the spectral domain [4]. Therefore, it is necessary to reduce the dimensionality and remove redundant information contained in HSI, while improving the discriminability of pixels belonging to different classes.

Dimensionality reduction (DR) is a necessary preprocessing step that decreases the number of hyperspectral channels for improving the classification performance of HSI. Generally, DR techniques can be classified into two major types: band (feature) selection (BS) or feature extraction (FE) [5,6]. The FE [7] methods use all of the features to find a transformation that maps the original data into a low-dimensional subspace. The BS [8] aims to identify a suitable subset of the original feature set, which allows optimization of the classification results. Compared to FE, the BS techniques can inherit the original spectral meanings of HSI data [9]. In this paper, we focus on BS techniques.

According to the selection strategy used, the BS techniques can be divided into ranking-based and clustering-based methods. In ranking-based BS, each band is sorted by a band prioritization criterion (e.g., the maximum variance-based criterion [10] and the non-Gaussianity (NG) [6] criterion used in information divergence (ID) [11]), before top-ranked bands are selected. However, this type of BS method does not consider spectral correlation and cannot remove the redundant bands. The clustering-based methods perform clustering on each band to group them into different clusters according to their similarity (or correlation). Accordingly, redundancy reduction can be obtained by selecting the representative bands. However, the fact that the classical clustering-based methods are sensitive to initialization implies that the selected subset of bands may be unstable and may change depending on the particular initialization. Recently, an exemplar-based clustering algorithm, namely affinity propagation (AP) [12], was applied to BS in HSI [13,14]. AP identifies clusters based on similarity measures, which is characterized by a relatively small error rate and a fast execution speed (especially for large datasets). The above AP-based methods can ensure that the selected bands have a relatively low redundancy and are stable. However, they do not consider the discriminative capability of single bands as they give the bands the same prior suitability (preference). This can significantly affect the final selection results of bands. A desirable BS process in HSI for classification tasks is to select highly discriminant bands and reduce redundant information. For achieving this goal, it is necessary to define a feature metric (or similarity measure of bands) in the AP so that both can assess the class separability of each band and the spectral correlation between two bands.

Based on the availability of prior information, BS algorithms can be further classified into supervised and unsupervised. Supervised methods [15,16] select a discriminative band subset by measuring the separation among classes with prior knowledge. In unsupervised methods [17,18], the informative bands can be evaluated by statistical measures and are selected by feature clustering. Recently, semi-supervised methods that exploit a small amount of prior information have been applied to reduce the HSI dimensions. Two main types of prior information can be used in HSI analysis: land cover class labels and pairwise constraints (i.e., positive constraints and negative constraints on pairs of pixels, which indicate only if a pair of samples belongs to the same class or not without knowing the class label). Earlier studies proposed that the metric (or similarity) should be learned in a supervised way through class labels. However, often, land cover class labels are difficult to obtain, and thus, it is difficult to define a reliable and complete training set for exploiting supervised methods. In contrast, pairwise constraints are relatively easy to achieve. Class labels and pairwise constraints are generally defined at the pixel level, with each single pixel considered as an independent information entity.

HSI represents the real land surfaces that typically extend for a few pixels in the image. Thus, nearby pixels in HSI have a high probability of belonging to the same class, indicating the existence of high spatial correlation [19,20]. Recently, a new learning paradigm, namely adjustment learning (AL), has been studied for image retrieval [21]. In the AL scheme, data points can be identified as small group sets, which are known to originate from the same class (but the label is unknown). These small group

sets are termed as “chunklets”, which can be constructed by positive constraints. Following this, a simple and efficient metric learning method called relevant component analysis (RCA) was developed, which was based on the AL [22,23]. The RCA uses chunklets to find a transformation matrix, which can improve data representation. In previous work, we integrated the RCA and the AP into a joint framework for HSI band selection [24,25]. However, only considering chunklet information or positive constraints is not enough. Negative constraints play a fundamental role in defining discriminant criteria. Two extensions of the RCA have been developed, which are the discriminative component analysis (DCA) [26] and the extended relevant component analysis (ERCA) [27]. Compared with the RCA, a discriminative set can be formed in the DCA by negative constraints for describing the relationship of chunklets, which delivers a type of discriminative information. In comparison, the ERCA simply focuses on positive and negative constraints.

A desirable BS process in HSI for classification tasks is to select a subset of bands with low redundancy and a strong capability to distinguish different classes by using easily-acquired prior information. For achieving this goal, we introduce a new type of discriminative constraint, namely chunklet and discriminative information, into the BS process by modeling the spectral and spatial relationship between pixels in this paper. The chunklet method groups spectrally-close and spatially-close pixels together without requiring explicit knowledge of their class labels, while discriminative information describes the relationship between any two chunklets providing valuable information for distinguishing classes. In greater detail, a discriminative feature metric (DFM) is modeled with the discriminative constraints in terms of an optimal criterion based on the DCA. The learned DFM can effectively assess both class separability and spectral correlation associated with HSI channels. Following this, the AP is used to search the subset of high-discriminative and low-redundancy bands. The proposed BS technique is referred to as discriminative feature metric-based affinity propagation (DFM-AP). The effectiveness of the proposed BS method is analyzed on two different hyperspectral datasets in terms of accuracy and robustness with the constructed discriminative constraints.

2. Methods

In this section, the discriminative feature metric-based affinity propagation (DFM-AP) is proposed. The full implementation process of the DFM-AP consists of the following three parts: (1) construction of discriminative constraints; (2) discriminative feature metric (DFM) learning; and (3) selection of representative band subsets based on AP clustering with the DFM. A detailed description of these three parts is given in the following subsections.

2.1. Construction of Discriminative Constraints

Let $X = \{x_1, x_2, \dots, x_N\} \subset R^{B \times N}$ be an HSI dataset, where x_i represents a vector of spectral responses, B is the number of bands and each band contains N pixels. Two key observations on the spectral and spatial relationship information are as follows:

Spectral similarity: HSI pixels that are similar in the feature space have a high probability of belonging to the same class (and vice versa).

Spatial correlation: HSI pixels that are spatially near each other have a high probability of belonging to the same class, while pixels that are far away from each other in the spatial domain may belong to different classes.

Following the observations described above, two types of prior information can be derived, which naturally integrates the spectral similarity and spatial correlation information. One is expressed in terms of pairwise constraints, while the other is expressed in terms of discriminative constraints. The pairwise constraints can always be used to construct discriminative constraints without any explicit knowledge of the pixel labels. We introduce the construction process of discriminative constraints, as shown in Figure 1. In this figure, blue, orange, red and green points denote four sets of

positive constraints, while orange and green, as well as red and green points are examples of negative constraints. Assuming that K chunklets are generated according to the given positive constraints, the k -th chunklet is termed as $C_k = \{x_{k1}, x_{k2}, \dots, x_{kn_k}\}$ ($k = 1, 2, \dots, K$), where n_k is the number of pixels in the k -th chunklet. For each chunklet, a discriminative set D is formed by the negative constraints in order to represent discriminative information between any two sets of chunklets. For the k -th chunklet, each element in the discriminative set D_k indicates one of the K chunklets that can be discriminated from the k -th chunklet, which is also known as the discriminative chunklet. Two chunklets can be discriminated from each other if there is at least one negative constraint between them. As shown in Figure 1, the positive constraints can be grouped into four chunklets (i.e., C_1, C_2, C_3 and C_4), and any two chunklets with one negative constraint are discriminative chunklets (i.e., C_2 and C_4 , as well as C_3 and C_4). The chunklets and the discriminative information are grouped together and termed as discriminative constraints.

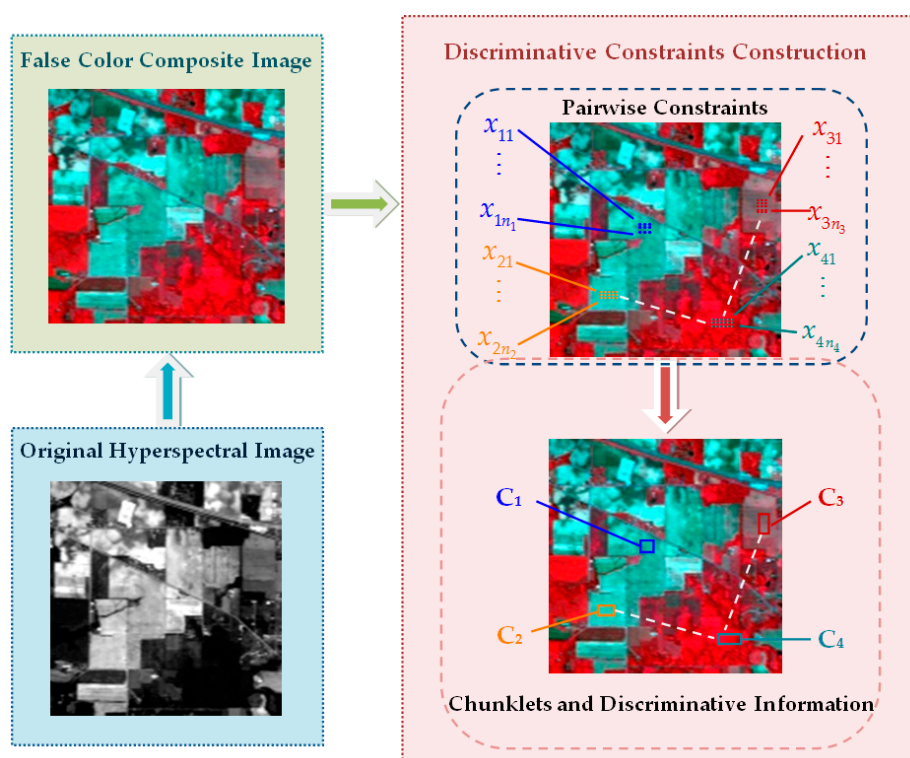


Figure 1. Example of the process of the construction of the discriminative constraints.

Indeed, the two types of information are not equivalent since discriminative constraints should be constructed with spatially-near pairwise constraints, while pairwise constraints do not need to hold both spectral and spatial relationships. It is important to note that the discriminative constraints can reflect the spectral and spatial relationship information between pixels. However, these are not limited to a pair of pixels and can sometimes be obtained directly by photo-interpretation or by automatic image analysis techniques without knowing the specific pairwise constraints.

2.2. Learning Discriminative Feature Metric

The aim of BS is to find a subset of representative bands $Y = \{y_1, y_2, \dots, y_b\}$ ($b \ll B$), which can better reflect discriminative chunklets, where y_1, y_2, \dots, y_b are the selected representative bands from X .

The DFM learned here is based on the optimal criteria of the DCA and learns a distance metric by a data transformation for both minimizing the total variance of data points within the same chunklets

and maximizing the total variance between the chunklets with discriminative information. To achieve this goal, two covariance matrices are introduced, which are the between-chunklets covariance matrix \hat{C}_b and within-chunklets covariance matrix \hat{C}_w . They can be written as follows:

$$\hat{C}_b = \frac{1}{K_d} \sum_{k=1}^K \sum_{t \in D_k} (m_k - m_t)(m_k - m_t)^T \quad (1)$$

$$\hat{C}_w = \frac{1}{K} \sum_{k=1}^K \frac{1}{n_k} \sum_{l=1}^{n_k} (x_{kl} - m_k)(x_{kl} - m_k)^T \quad (2)$$

where $m_k = \frac{1}{n_k} \sum_{l=1}^{n_k} x_{kl}$ is the mean of the k -th chunklet; m_t is the mean of the t -th chunklet; x_{kl} is the l -th pixel of the k -th chunklet; n_k is the total number of pixels in the k -th chunklet; $K_d = \sum_{k=1}^K |D_k|$ denotes the cardinality of a set; and D_k indicates one of the K chunklets, which can be discriminated from the k -th chunklet.

The Fisher theory or criterion [28,29] and its derived algorithms such as Fisher's linear discriminant analysis (FLDA) and some modified methods [30–32], have been widely accepted as effective techniques for HSI classification and analysis. According to the Fisher theory, the solution of DCA corresponds to a learned transformation matrix, which maximizes the ratio of between-chunklets and within-chunklets covariance in a matrix. Accordingly, the transformation matrix W can be defined as:

$$W = \hat{C}_b \cdot \hat{C}_w^{-1} \quad (3)$$

The DFM between different bands x_i and x_j can be computed as follows:

$$\text{DFM}(x_i, x_j) = W(x_i, x_j) \quad (4)$$

$i = 1, 2, \dots, B; \quad j = 1, 2, \dots, B; \quad i \neq j$

The DFM of a single band x_i can be expressed as:

$$\text{DFM}(x_i, x_i) = -DFTS \cdot \left| \frac{1}{W(x_i, x_i)} \cdot \frac{Max}{Min} \right| \quad (5)$$

$i = 1, 2, \dots, B$

where *Max* and *Min* are the maximum and minimum values of $1/W(x_i, x_i)$. By setting an appropriate threshold, which is also known as the discriminative feature threshold scalar (DFTS), one can select the spectral band subset Y in which the classes can be well separated with discriminative constraints.

2.3. Discriminative Feature Metric-Based Affinity Propagation

AP [12] takes the similarity values between pairs of data points as input measures and aims to find exemplars to maximize the sum of similarities between data points and corresponding exemplars. A common choice for similarity is the negative Euclidean distance. It does not require that the number of clusters is prespecified, as it is controlled by setting the parameter "preference", namely the self-similarity. The "preference" represents the prior suitability of a data point to be the exemplar, which can be set to a global (shared) value or customized for specific data points. In this present study, the learned DFM is adopted to measure the similarity of bands and is used in the AP algorithm as an input measure. The DFM between different bands denotes the band correlation, while the DFM of a single band denotes its preference (band priority), which is its prior suitability to serve as a representative band. The proposed approach is referred to as the discriminative feature metric-based affinity propagation (DFM-AP). The details and procedure associated with DFM-AP are as follows.

Proposed Discriminative Feature Metric-Based Affinity Propagation

Input:

Hyperspectral dataset X , Discriminative Constraints C and D .

Output:

Representative bands set Y

Procedure:

1. Initialize values of exemplars and parameters.
 - Let $Y = X$: consider all bands to be initial clustering exemplars, which are namely the representative bands;
2. Compute the between-discriminative chunklets matrix \hat{C}_b and within-chunklets scatter matrix \hat{C}_w using Equations (1) and (2).
3. Calculate the DFM for all spectral bands;
 - Set DFTS to get the expected number of bands;
 - Calculate the DFM between different bands using Equation (4);
 - Compute the DFM of single band according to Equation (5).
4. Update availability and responsibility.

Two types of messages, responsibility r and availability a , communicate between bands, with each taking a different kind of competition into account. Let $a(x_i, x_j)$ (availability) denote the degree of band x_i as the cluster center for band x_j . To begin with, let $a(x_i, x_j) = 0$; $r(x_i, x_j)$ (responsibility) is the degree of the band x_j , which is suitable for being the cluster center for band x_i . The cluster centers of all of the bands, $a(x_i, x_j)$ and $r(x_i, x_j)$, are computed as follows:

$$a(x_i, x_j) = \begin{cases} \sum_{p \neq j} \max\{0, r(x_p, x_j)\} & i = j \\ \min\left\{0, r(x_j, x_j) + \sum_{p \neq j, i} \max\{0, r(x_p, x_j)\}\right\} & i \neq j \end{cases} \quad (6)$$

$$r(x_i, x_j) = DFM(x_i, x_j) - \max_{p \neq j} \{DFM(x_i, x_p) + a(x_i, x_p)\} \quad (7)$$

The damping factor is used in over-relaxation methods to avoid numerical oscillations when computing responsibilities and availabilities with simple updating rules. The two types of messages could be damped according to the following equations:

$$\begin{aligned} R^{t+1} &= \alpha R^{t-1} + (1 - \alpha) R^t \\ A^{t+1} &= \alpha A^{t-1} + (1 - \alpha) A^t \end{aligned} \quad (8)$$

where R and A represent responsibility and availability vectors, respectively; α is the factor of damping; and t is the number of iterations.

5. Obtain the cluster centers set C of all of the spectral bands.

For any band x_i , a larger sum of $a(x_i, x_j)$ and $r(x_i, x_j)$ means a greater possibility of band x_j to be the final cluster center of band x_i . The band x_j determines its cluster center according to the following equation:

$$\max_{x_j \in C} \{a(x_i, x_j) + r(x_i, x_j)\} \quad (9)$$

6. Identify the representative bands Y and their number b .

Repeat Steps 4 and 5, until cluster boundaries are unchanged for a given number of iterations. At convergence, we obtain the final subset of representative bands and the number of bands to be used for HSI classification.

3. Experimental Results

3.1. Dataset Description

Two different types of hyperspectral datasets are used in the experiments.

The first dataset Indian Pines, Indian Pines 92AV3C [33], was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over northwest Indiana on 12 June 1992. It has 145×145 pixels and 220 spectral bands ranging from 400–2500 nm with a spatial resolution of 20 m. After discarding the lower signal-to-noise (SNR) bands (104–108, 150–163 and 220), 200 bands were considered. The dataset includes 16 land cover classes, which represent different crop types, vegetation and man-made structures with 10,366 true ground labeled pixels. Figure 2 shows the false color composite image of Bands 57, 27 and 17 and the available true map of the Indian Pines dataset.

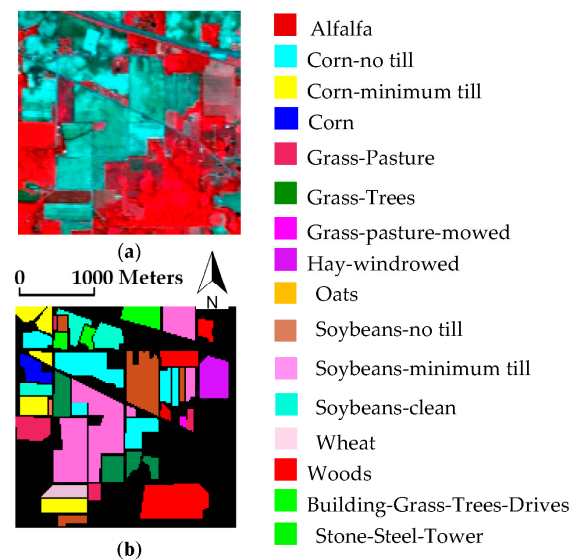


Figure 2. (a) False color composite of image of Bands 57 (927.61 nm), 27 (656.67 nm) and 17 (557.49 nm); and (b) available ground truth map (Indian Pines dataset).

The second dataset, Xuzhou, was acquired by Hyperion over the city of Xuzhou, Jiangsu Province, China on 7 February 2004 (Entity ID: EO1H1210362004038110PZ_AGS_01). It has 250×566 pixels and shows an urban area. The original image contains 242 spectral channels with a wavelength range from 356–2577 nm with only 198 bands being calibrated. After removing uncalibrated and noisy bands (1–10, 58–81, 120–130, 165–182, 185–187, 200, 201 and 221–242), 152 channels were selected to use as input parameters in the process of band selection. The image contains 6233 ground truth pixels classified into five land cover classes, which are water, vegetation, woodland, built-up and bare land. Figure 3 shows a false color composition and the available true ground map of the Xuzhou dataset.

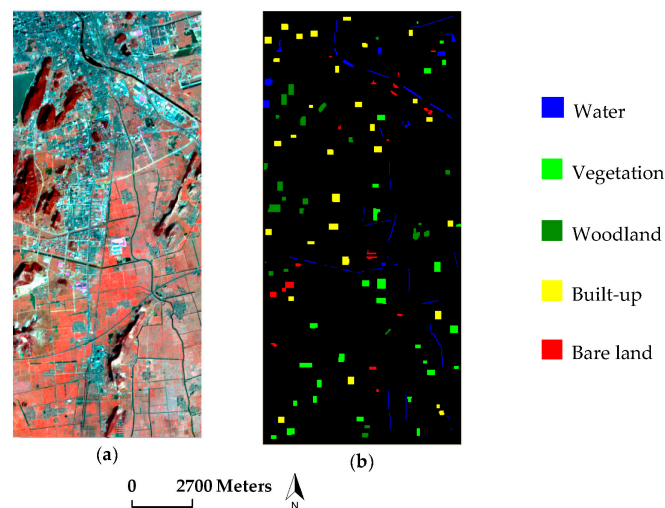


Figure 3. (a) False color composite of Bands 84 (983.08 nm), 33 (681.20 nm) and 26 (609.97 nm); and (b) true ground map (Xuzhou dataset).

3.2. Experimental Design

For evaluating the performance of the proposed DFM-AP, we compared it with the seven unsupervised, semi-supervised and supervised BS methods in our experiments as described below. Table 1 describes the acronyms of compared methods and evaluation criteria used in experiments.

Unsupervised BS:

- ID [6,11]: all bands are analyzed by a band prioritization criterion, known as the NG;
- Maximum variance-based principal component analysis (MVPCA) [10]: MV criteria used to prioritize bands;
- AP [12]: standard AP with the negative Euclidean distance;
- Adaptive AP (AAP) [14]: AP with the negative spectral angle mapper (SAM) and an exemplar number determination procedure for getting fixed selected band numbers;

Semi-supervised BS:

- RCA-based AP (FM-AP) [24,25]: AP using a feature metric (FM) based on the criterion of the RCA with chunklets;
- ERCA-based AP (pairwise feature metric (PFM)-AP): the similarity of bands used in AP is based on the optimization criterion of ERCA with pairwise constraints.

Supervised BS:

- FLDA-based AP (LFM-AP): the similarity in AP is based on FLDA with class labels.

We also compared the results obtained with those achieved by using all of the original bands (baseline).

For Indian Pines and Xuzhou datasets, we randomly selected 630/90 and 1980/13 pixels as positive/negative constraints from the available true ground pixels. A total of 63 and 11 chunklets (each chunklet contains 10 and 180 pixels for different land cover classes, which are unknown in advance) are formed by the given positive constraints. For each dataset, a discriminative set is formed by negative constraints, in which a chunklet is discriminated from another if there is at least one negative constraint between them. In practice, we just need to consider certain pixels in any two chunklets that have a negative relationship. The randomly-selected pairwise constraints in the experiments are used for the PFM-AP learning, while the transformed chunklets and discriminative

constraints are used for the FM-AP and DFM-AP learning. In order to obtain statistically-significant results, a widely-used supervised classifier, namely the support vector machine (SVM), is employed to evaluate the performance of BS methods. The randomly-selected constraints are transformed into class labels with available ground-truth data for FLDA learning and SVM training, while the remaining pixels are used for estimating the performance of BS methods. To assess whether results were statistically significant, we repeated this process for five trials/times and used the average overall accuracy (AOA) and the related standard deviation (SD) as evaluation criteria. It is important to note that no labeled samples were used in our BS process.

Table 1. Description of the acronyms of compared methods and evaluation criteria used in experiments.

Acronyms	Description	Acronyms	Description
ID	Information divergence	LFM-AP	Label feature metric-based affinity propagation
NG	Non-Gaussianity	PFM-AP	Pairwise feature metric-based affinity propagation
MVPCA	Maximum variance-based principal component analysis	DFM-AP	Discriminative feature metric-based affinity propagation
AP	Affinity propagation	SVM	Support vector machine
AAP	Adaptive affinity propagation	AOA	Average overall accuracy
FM-AP	Feature metric-based affinity propagation	SD	Standard deviation

3.3. Results

In the experimental analysis, we first quantified the performance of the proposed DFM-AP with respect to the other seven reference BS methods and to the baseline. Secondly, we analyzed the behavior of the AOA provided by the DFM-AP compared to the number of chunklets. Thirdly, we assessed the classification accuracy of the DFM-AP with different percentages of discriminative information. Finally, a parameter sensitivity analysis of DFM-AP is given.

3.3.1. Accuracy Compared to the Number of Selected Bands

The AOA and SD obtained by using all of the spectral channels and the selected bands subsets were compared (see Figure 4 and Table 2).

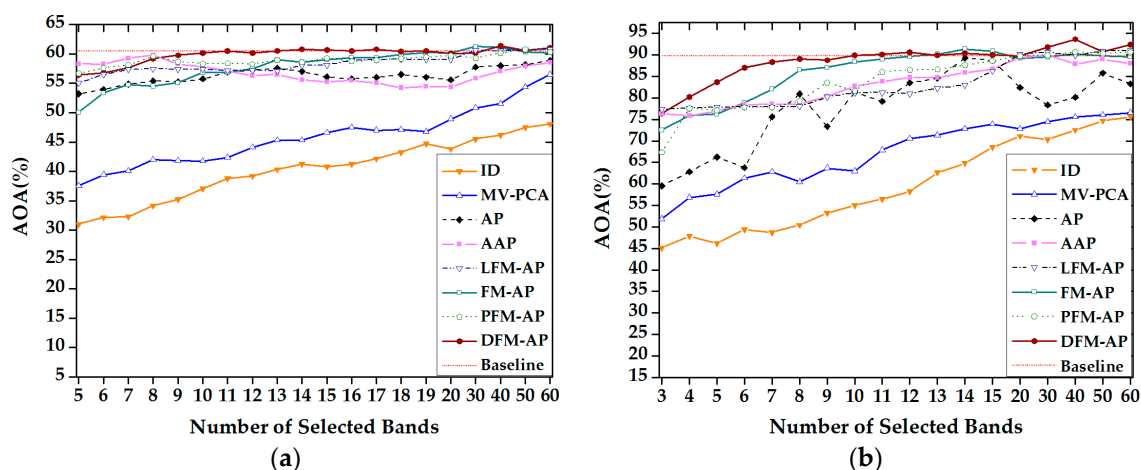


Figure 4. AOA on five trials provided by the SVM compared to the number of selected bands obtained by the ID, MV-PCA, AP, AAP, LFM-AP, FM-AP, PFM-AP and proposed DFM-AP methods with the (a) Indian Pines and (b) Xuzhou datasets. The results achieved by using all of the spectral channels are also reported (baseline).

Table 2. SD of the seven compared methods and proposed DFM-AP compared to the number of selected bands (NB) in the range of 10–60 for the Indian Pines and Xuzhou datasets.

Datasets	SD		MVPCA	ID	AP	AAP	LFM-AP	FM-AP	PAM-AP	DFM-AP
	NB									
Indian Pines	10		1.642	1.426	1.382	2.368	2.667	2.005	1.061	1.045
	20		1.746	1.526	1.458	0.767	2.985	1.106	1.304	2.006
	40		1.996	1.129	2.126	1.831	2.480	0.910	1.398	2.205
	60		1.582	1.468	2.308	2.116	1.427	0.689	1.399	1.878
Xuzhou	10		1.860	1.757	2.363	2.972	2.437	1.661	6.476	1.812
	20		1.969	1.812	2.007	2.045	2.954	1.874	3.624	1.972
	40		1.783	1.826	1.964	2.510	2.972	1.985	3.085	1.596
	60		1.813	1.719	2.174	2.747	2.700	1.956	3.135	1.989

For the Indian Pines and Xuzhou datasets, we can observe that the AP-based methods yielded a higher AOA than the other two BS methods (i.e., the MVPCA and ID). In greater detail, compared with the AP, AAP and PFM-AP, the proposed DFM-AP, LFM-AP and FM-AP exhibited more stable performances with a lower SD. It is important to note that there were a few times when the PFM-AP and AAP achieved a higher AOA (when 5–8 bands were selected for the Indian Pines dataset). This is probably due to the fact that spectral values of different land cover classes (e.g., corn and soybean-types) are very close in this dataset, which thus requires more bands to discriminate them. Nevertheless, the AOA obtained by AAP shows a sharp decline and is lower than the baseline when the number of selected bands increases. Furthermore, the proposed DFM-AP approach always provides the best AOA and a lower SD compared to the FM-AP and LFM-AP with an increase in the number of selected bands. For Indian Pines and Xuzhou datasets, the DFM-AP obtained only 14 and 10 bands with AOA values (60.80% and 89.87%, respectively) that were higher than those yielded when using all 200 and 152 bands (60.55% and 89.76%, respectively). In comparison, more bands (30 and 13) have to be selected to achieve a higher accuracy for FM-AP than the baseline in the two datasets.

Figure 5 shows the wavelength of 14 selected bands compared to the radiance value of land cover classes at the sensor for the proposed DFM-AP and FM-AP in the Indian Pines dataset. It was seen the FM-AP selected many adjacent bands in each trial as it uses only chunklets, which are affected by the imbalance and small sample size of the dataset. On the contrary, the bands selected by the proposed DFM-AP have a more dispersed distribution in the five trials. Moreover, they cover regions of the spectrum with large intervals, which indicates their highly discriminative capabilities for different categories.

3.3.2. Comparison of Accuracy Compared to the Number of Chunklets

This experiment explores the effect of the number of chunklets on the AOA of DFM-AP in the two considered datasets (see Figure 6). For the Indian Pines and Xuzhou datasets, the originally generated pairwise constraints were divided into 126/189 and 55/110 chunklets (each contains 5/3 and 36/18 pixels), respectively.

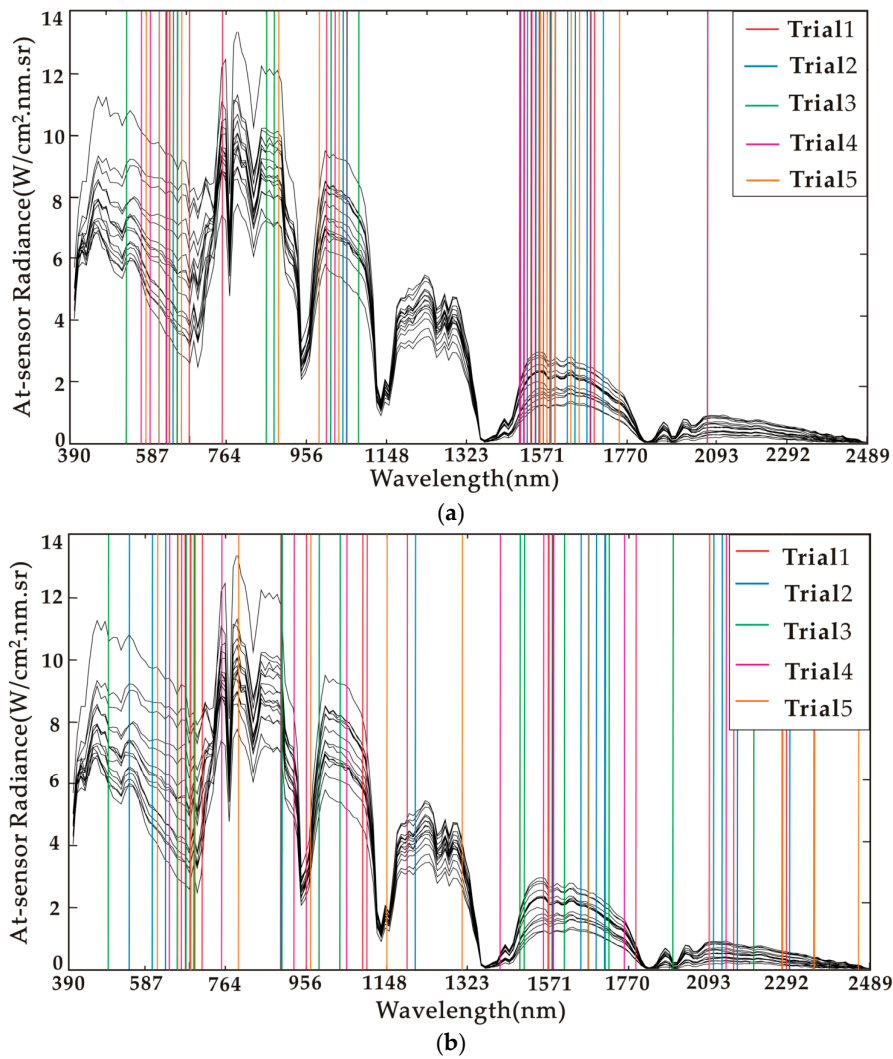


Figure 5. Wavelength of selected bands (straight lines) compared to the radiance value at the sensor for each land cover class (16 curves) for the: (a) FM-AP; and (b) proposed DFM-AP (Indian Pines dataset).

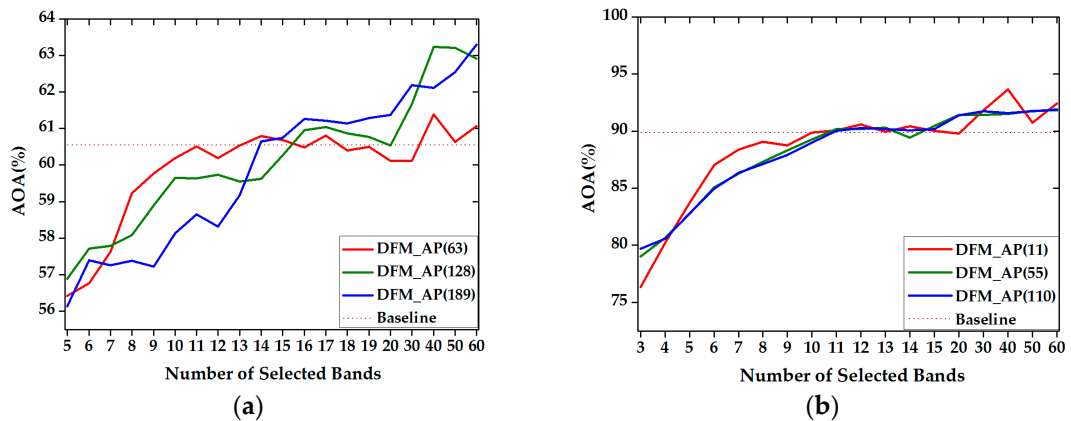


Figure 6. AOA compared to the number of selected bands obtained by the proposed DFM-AP for different numbers of chunklets in the: (a) Indian Pines; and (b) Xuzhou datasets.

3.3.3. Comparison of Accuracy with Discriminative Information

This experiment assesses the performance of the proposed DFM-AP when considering a different amount of discriminative information. We randomly generated 0.5-, 1-, 2-, 3- and 4-times the number of the original negative constraints as a discriminative set, while maintaining the number of positive constraints/chunklets attached to the Indian Pines and Xuzhou datasets (see Table 3).

Table 3. AOA of the proposed DFM-AP compared to the number of negative constraints with different numbers of selected bands (NB) for the Indian Pines and Xuzhou datasets.

Dataset	AOA NB	Discriminative Information (# Times the Number of Original Negative Constraints)				
		0.5	1	2	3	4
Indian Pines	12	57.39	60.18	59.70	60.69	57.87
	20	59.60	60.11	60.21	62.23	60.90
	30	60.09	60.12	62.27	61.97	60.46
	40	61.06	61.39	62.04	62.13	60.47
	50	61.52	60.63	62.23	63.04	60.98
	60	62.18	61.06	62.11	63.45	60.14
Xuzhou	8	58.46	89.08	89.27	89.58	90.15
	20	88.02	89.79	90.41	90.93	91.07
	30	90.18	91.82	92.67	91.02	91.93
	40	91.60	93.65	93.52	93.95	93.82
	50	90.15	90.72	92.59	92.23	92.73
	60	92.07	92.43	91.86	92.52	93.37

3.3.4. Sensitivity Analysis

There are two user-defined parameters in the proposed DFM-AP:

- α , which affects the convergence speed; and
- DFTS, which determines the number of selected bands (i.e., representative bands).

The parameter α should be at least 0.5 and less than one. In general, it can be set between 0.8 and 0.9. If the algorithm does not converge, the value can be increased, but the execution time increases as well. In the experiments, the α values of AP, FM-AP and DFM-AP were set to 0.85 and 0.9, while the average execution times were 6.8, 5.2 and 5.9 seconds with 14 selected bands in the Indian Pines dataset (PC workstation (Intel(R) Core(TM) i7-3720QM CPU @ 2.60 GHz, 2.60 GHz with 16.0 GB of RAM)), respectively. All software is implemented in Microsoft visual C++.Net. Generally, the α can be set to 0.9. A larger α means a better convergence of AP. If the algorithm does not converge, the value α can be increased, but numerical precision issues can arise if it goes beyond 0.99 [34]. Readers can refer to a previous study [25] for more details on the selection of the parameter α value. For the second parameter, the number of representative bands is close to being monotonically related to the DFTS. Lower values of DFTS resulted in the selection of many bands, while high values led to a small number of bands in all sampling conditions of the two datasets, although these had a different range of values (see Figure 7). As done in other BS techniques, one can run DFM-AP several times with different DFTS values searching for the desired number of bands. The DFTS ranges and values can also be obtained automatically [35].

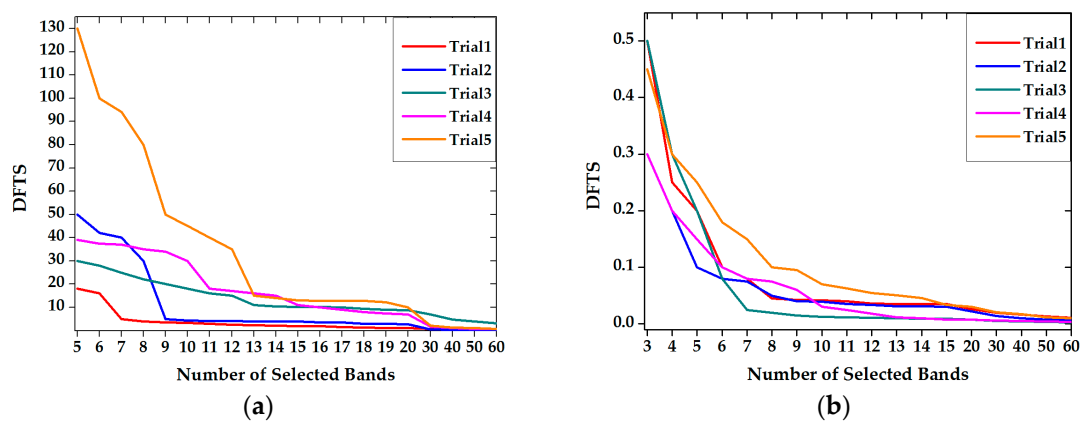


Figure 7. Number of selected bands compared to the value of parameter discriminative feature threshold scalar (DFTS) for the proposed DFM-AP in the: (a) Indian Pines; and (b) Xuzhou datasets.

4. Discussion

Based on the experimental results presented in the previous section, we can observe that the proposed DFM-AP is almost always superior to the other BS methods in terms of AOA and SD in the considered datasets. In this section, we focus our attention on a discussion of the effectiveness of BS algorithms with respect to the selection strategy and the prior information.

4.1. Clustering-Based Methods versus Ranking-Based Methods

According to the experimental results presented in the previous section, we can conclude that the clustering-based BS methods (i.e., the proposed DFM-AP, FM-AP, PFM-AP, LFM-AP, AAP and AP) produced higher classification accuracies than the ranking-based methods (i.e., ID and MVPCA). Among these BS algorithms, ranking-based methods (i.e., ID and MVPCA) obtained the lowest AOA because they neglect the redundancy of selected bands. On the contrary, redundant bands can be removed by selecting the representative bands in the clustering process. At the same time, the AP-based methods consider all data points to be equally suitable as exemplars or can use constraints to guide the clustering process for finding effective representative bands.

4.2. BS Performance Compared to Different Prior Information

From the experimental results, several conclusions can be drawn. First, we can observe from Figure 4 that, as expected, the prior information used in similarity metrics can generally improve the accuracy. The AOA of the proposed DFM-AP, FM-AP, PFM-AP and LFM-AP are higher than those of the unsupervised methods (i.e., the AAP, AP and the ID, MVPCA) with an increase in the number of selected bands.

With regards to the proposed DFM-AP, FM-AP, PFM-AP and LFM-AP methods, four different types of prior information, namely the discriminative constraints (chunklets and discriminative information), chunklets, pairwise constraints and class labels, have been used for finding subsets of bands with a relatively high capability for class discrimination. In the proposed DFM-AP, as well as LFM-AP and PFM-AP, this is done by learning the optimal data transformation matrices via both maximizing the total variance between discriminative chunklets, different classes and negative pixels in addition to minimizing the total variance of pixels in the same chunklets, the same classes and positive pixels. This leads to a final feature metric. The FM-AP only focuses on minimizing the total variance of pixels in the same chunklet. From this point of view, the organization of the modeling of the prior information in pixels plays an important role in the BS process. Discriminative constraints combine pixels into small group sets with the spectral and spatial characteristics of HSI and build the connections between pixel groups. Therefore, a higher intraclass variability and lower interclass

variability between two land cover classes can be more suitable to a certain degree. However, the pairwise constraints and class labels are based on pairwise pixel connections, which cannot guarantee the spatial correction among pixels, especially when the prior information is dispersed. By analyzing Table 2 in greater detail, it was seen that the proposed DFM-AP approach always provided a higher AOA and lower SD than the PFM-AP and LFM-AP with an increase in the number of selected bands in the Xuzhou dataset. Thus, the proposed DFM-AP achieves a relatively higher stability than the other three clustering-based methods. At the same time, it improves the classification performance by selecting highly discriminative and low redundancy bands (see Figure 5).

4.3. BS Performance Compared to the Amount of Prior Information

This sub-section discusses the performances of the proposed DFM-AP with different numbers of chunklets and different discriminative information (see Figure 6 and Table 3).

From Figure 6, it was observed that the AOA of the Indian Pines dataset increases with an increase in the number of chunklets. This is due to the fact that the Indian Pines dataset shows high intraclass variability, and the diversification of chunklets models this variability. In comparison, the DFM-AP is not very sensitive to the number of chunklets for the Xuzhou dataset. This may be due to the fact that spectral values of the same land cover classes in the Xuzhou dataset are very close, as they show low intraclass variability. It is important to note that the accuracy decreases in both datasets with an increase in the number of chunklets over a given value. This is due to the fact that each chunklet contains only two pixels at this value, as seen in the case of PFM-AP.

By analyzing Table 3, it was seen that the AOA values obtained by the proposed DFM-AP on the Indian Pines and Xuzhou datasets were higher than the baseline and had a fewer number of bands (12 and eight bands, respectively). These values were also higher than those obtained with the FM-AP (+3.24% and +3.67%) when three- or four-times the number of original discriminative bands were considered on the two datasets. This proves that the performances of the proposed DFM-AP in identifying highly discriminative subsets of bands are significantly better than those obtained by the previous FM-AP when introducing a given amount of discriminative information. When using a limited (0.5-times) amount of discriminative information, more bands (about 30 bands) should be selected to obtain a higher accuracy than the baseline. Nevertheless, the proposed DFM-AP exhibited a more stable behavior with respect to the FM-AP when the number of selected bands was further increased.

According to the above-mentioned discussion, the performance of the DFM-AP technique depends on the adopted DFM and on the number of discriminative constraints. However, the desired number of discriminative constraints to be selected is not known a priori, as it varies with the HSI data considered. Determination of an optimal number of discriminative constraints will be investigated in the future. On the other hand, the proposed DFM-AP based on the solution of DCA corresponds to a learned transformation matrix needed for learning the DFM. The eigenanalysis has been used in the learning process, but the relationship between eigenvalues and eigenvectors was not fully exploited. Therefore, we will explore other FM criteria via eigen (spectral) decomposition to further improve the classification performance in future studies.

5. Conclusions

In this study, we have presented a novel DFM-AP approach to identify a high-discriminative and low-correlated subset of bands for HSI classification. Considering the difficulty in the collection of labeled pixels, new types of discriminative constraints (i.e., chunklets and discriminative information) are constructed with spectral and spatial relationship information without requiring any land cover class label. The DFM-AP learns a DFM as a similarity metric, which is based on the optimal criterion of the DCA with constructed chunklets and discriminative information that aims to measure both the correlations and the discriminative capability of bands. Following this, the subset of bands is selected by using the AP clustering. The performances of the proposed DFM-AP have been compared

against seven BS methods with different types of prior information. The results show that the proposed DFM-AP achieves the highest AOA with a low SD. Furthermore, the experiments on the influence of both the number of chunklets and the discriminative information on the AOA show that the chunklet groups can improve BS performance effectively with the introduction of a small amount of discriminative information, because of the joint exploitation of pixel aggregation and separability.

Acknowledgments: This work was funded by the National Natural Science Foundation of China (61572228 and 11403049), the National Key Basic Research Program of China (2015CB453000) and the Science-Technology Development Plan Project of Jilin Province of China (20160101247J and 20140520070JH). The authors would like to thank Landgrebe from Purdue University for providing AVIRIS images and ground truth map, P. J. Du from Nanjing University and S. C. Liu from Tongji University for providing Hyperion images and the ground truth map used in this paper. Hyperion data are available from the U.S. Geological Survey. The authors would like to thank the Editor and referees for their suggestions that improved the manuscript.

Author Contributions: All coauthors made significant contributions to the manuscript. Chen Yang presented the key idea of the DFM-AP method and designed the contrast experiments. Yulei Tan performed the comparison experiments between the proposed DFM-AP and the seven other band selection methods. Lorenzo Bruzzone provided the background knowledge of hyperspectral processing and helped to revise the manuscript. Renchu Guan helped to design the procedures of experiments in the manuscript.

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

References

1. Smith, R.B. *Introduction to Hyperspectral Imaging*; TNTmips, MicroImages: Lincoln, NE, USA, 2012; pp. 1–24.
2. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; John Wiley & Sons Press: New York, NY, USA, 2001.
3. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [[CrossRef](#)]
4. Bruzzone, L.; Demir, B. A review of modern approaches to classification of remote sensing data. In *Land Use and Land Cover Mapping in Europe*; Springer: Dordrecht, The Netherlands, 2014; Volume 18, pp. 127–143.
5. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, S110–S122. [[CrossRef](#)]
6. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2006.
7. Dopido, I.; Villa, A.; Plaza, A.; Gamba, P. A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 421–435. [[CrossRef](#)]
8. Serpico, S.B.; Bruzzone, L. A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **1994**, *39*, 1360–1367. [[CrossRef](#)]
9. Sun, W.; Jiang, M.; Li, W.; Liu, Y. A symmetric sparse representation based band selection method for hyperspectral imagery classification. *Remote Sens.* **2016**, *8*, 238. [[CrossRef](#)]
10. Chang, C.-I.; Du, Q.; Sun, T.L.; Althouse, M.L. A joint band prioritization and band-deccorelation approach to band selection for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2631–2641. [[CrossRef](#)]
11. Chang, C.-I.; Wang, S. Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1575–1585. [[CrossRef](#)]
12. Frey, J.F.; Dueck, D. Clustering by passing messages between data points. *Science* **2007**, *315*, 972–976. [[CrossRef](#)] [[PubMed](#)]
13. Qian, Y.; Yao, F.; Jia, S. Band selection for hyperspectral imagery using affinity propagation. *IET Comput. Vis.* **2009**, *3*, 213–222. [[CrossRef](#)]
14. Su, H.J.; Sheng, Y.H.; Du, P.J.; Liu, K. Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection. *Appl. Opt.* **2012**, *51*, 2656–2663. [[CrossRef](#)] [[PubMed](#)]
15. Guo, B.; Gunn, S.R.; Damper, R.I.; Nelson, J.D.B. Band selection for hyperspectral image classification using mutual information. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 522–526. [[CrossRef](#)]
16. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.

17. Martínez-usó, A.; Pla, F.; Sotoca, J.M.; García-sevilla, P. Clustering based hyperspectral band selection using information measures. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 4158–4171. [[CrossRef](#)]
18. Sun, K.; Geng, X.; Ji, L.; Lu, Y. A new band selection method for hyperspectral image based on data quality. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2697–2703.
19. Ji, R.; Gao, Y.; Hong, R.; Liu, Q. Spectral-spatial constraint hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1811–1824.
20. Cheng, G.; Zhu, F.; Xiang, S.; Wang, Y. Semisupervised hyperspectral image classification via discriminant analysis and robust regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 595–608. [[CrossRef](#)]
21. Shental, N.; Hertz, T.; Weinshall, D.; Pavel, M. Adjustment learning and relevant component analysis. In Proceedings of the European Conference on Computer Vision, Copenhagen, Denmark, 28–31 May 2002; pp. 776–790.
22. Bar-Hillel, A.; Hertz, T.; Shental, N.; Weinshall, D. Learning distance functions using equivalence relations. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 11–18.
23. Bar-Hillel, A.; Hertz, T.; Shental, N.; Weinshall, D. Learning a Mahalanobis metric from equivalence constraints. *J. Mach. Learn. Res.* **2005**, *6*, 937–965.
24. Yang, C.; Liu, S.; Bruzzone, L.; Guan, R.C. A semisupervised feature metric based band selection method for hyperspectral image classification. In Proceedings of the 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Shanghai, China, 4–7 July 2012; pp. 1–4.
25. Yang, C.; Liu, S.C.; Bruzzone, L.; Guan, R.C.; Du, P.J. A feature-metric-based affinity propagation technique to feature selection for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1152–1156. [[CrossRef](#)]
26. Hoi, S.C.H.; Liu, W.; Lyu, M.R.; Ma, W.Y. Learning distance metrics with contextual constraints for image retrieval. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 17–22 June 2006; pp. 2072–2078.
27. Yeung, D.Y.; Chang, H. Extending the relevant component analysis algorithm for metric learning using both positive and negative equivalence constraints. *Pattern Recognit.* **2006**, *39*, 1007–1010. [[CrossRef](#)]
28. McLachlan, G.J. *Discriminant Analysis and Statistical Pattern Recognition*; John Wiley: New York, NY, USA, 1992.
29. Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B.; Muller, K. Fisher discriminant analysis with kernels. In Proceedings of the IEEE Signal Processing Society Workshop, Madison, WI, USA, 25 August 1999; pp. 41–48.
30. Chang, C.-I.; Ji, B.-H. Fisher’s linear spectral mixture analysis. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2292–2304. [[CrossRef](#)]
31. Du, Q. Modified fisher’s linear discriminant analysis for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 503–507. [[CrossRef](#)]
32. 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.* **2012**, *50*, 1185–1198. [[CrossRef](#)]
33. Baumgardner, M.F.; Biehl, L.L.; Landgrebe, D.A. *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*; Purdue University Research Repository: West Lafayette, Indiana, 2015. [[CrossRef](#)]
34. Frey, J.F.; Dueck, D. Affinity Propagation FAQ. Available online: <http://www.psi.toronto.edu/affinitypropagation/faq.html> (accessed on 26 May 2017).
35. Frey, J.F.; Dueck, D. Affinity Propagation. Available online: <http://www.psi.toronto.edu/index.php?q=affinity%20propagation> (accessed on 26 May 2017).

