



Unsupervised Hyperspectral Band Selection Based on Maximum Information Entropy and Determinantal Point Process

Zhijing Yang¹(✉), Weizhao Chen¹, Yijun Yan², Faxian Cao¹,
and Nian Cai¹

¹ School of Information Engineering, Guangdong University of Technology,
Guangzhou 510006, China

yzhj@gdut.edu.cn

² Department of Electronic and Electrical Engineering,
University of Strathclyde, Glasgow G1 1XW, UK

Abstract. Band selection is of great important for hyperspectral image processing, which can effectively reduce the data redundancy and computation time. In the case of unknown class labels, it is very difficult to select an effective band subset. In this paper, an unsupervised band selection algorithm is proposed which can preserve the original information of the hyperspectral image and select a low-redundancy band subset. First, a search criterion is designed to effectively search the best band subset with maximum information entropy. It is challenging to select a low-redundancy spectral band subset with maximizing the search criteria since it is a NP-hard problem. To overcome this problem, a double-graph model is proposed to capture the correlations between spectral bands with full use of the spatial information. Then, an improved Determinantal Point Process algorithm is presented as the search method to find the low-redundancy band subset from the double-graph model. Experimental results verify that our algorithm achieves better performance than other state-of-the-art methods.

Keywords: Unsupervised band selection · Maximum information
Graph model · Determinantal Point Process (DPP)

1 Introduction

Hyperspectral data is a three-dimension image with the spatial information and abundant spectral information which makes great development of hyperspectral images in the field of remote sensing [1, 2]. Multidimensional spectral bands can improve the representation of the ground objects. However, not all the bands play an important role in hyperspectral image processing [3]. That is because high dimensional hyperspectral data may cause a lot of problems, such as redundancy of information, noise bands and Hughes phenomenon. Therefore, an effective dimension reduction method is necessary for the hyperspectral data. Traditional dimensionality reduction methods mainly include two types: band selection and feature extraction [5–7]. Compared with feature extraction, band selection can effectively protect the original information as much as possible.

The research of band selection can be mainly divided into two directions: supervised band selection [8, 9] and unsupervised band selection [10–12]. Supervised band

selection method can select a spectral band subset that have a strong correlation with class labels. However, it's time-consuming to label the hyperspectral images. Therefore, the unsupervised band selection method is more applicable. In this paper, we propose an unsupervised band selection method to select a low-redundancy band subset with maximum amount of information.

The unsupervised band selection process mainly contains two steps: the selection criteria and search strategy. At present, the unsupervised selection methods can be divided into the ranking-based methods [13], clustering-based methods [17], and greedy-algorithms [15]. However, most of the above-mentioned band selection methods only consider the property of the band combination ignoring the latent structure between spectral bands in the high dimensional space [3]. This defect will cause the lack of global considerations. In addition, most band selection methods do not consider the correlation between pairwise bands from the spatial information. Spatial information helps to make the measurement of spectral redundancy more comprehensive. In [3], Yuan et al. proposed a multiple graph to measure similarity between spectral bands by spectral clustering method. However, the spectral clustering has very high computation complexity and high memory requirements. In view of this, this paper proposes a double-graph structure to describe the complex correlation between spectral bands with both pixels information and neighborhood information of pixels.

In addition, selecting a low-redundancy band subset from the graph model with traditional search methods is a challenging work. DPP [16] is a probabilistic model that can select a diverse subset of features. The k-DPP [17], as an improved algorithm of the DPP, can select a low-redundancy band subset with dimension k . The original k-DPP can only be applied to select from single graph model without full use of the spatial information of hyperspectral images. Furthermore, it does not consider the physical properties of each band in the selecting process.

To solve the above problems, a double-graph model DPP is proposed with full use of pixel information and its domain information. To attain a high-performance band subset, a spectral band selection criterion: maximum band subset information [4] is used which means that the selected spectral band subset has a rich amount of information.

2 The Proposed Framework

2.1 Hyperspectral Data Representation

A hyperspectral dataset is represented by $B = \{b_1, b_2, b_3, \dots, b_l\} \in R^{n \times l}$, where n represents the number of pixels in each spectral band and l is the total number of all spectral bands. $b_i (1 \leq i \leq l)$ indicates the i -th spectral band. Given a pixel point p_x^j , we use \hat{p}_x^j to represent the mean of the neighbourhood of pixel p_x^j . Then, the neighbour information can be rewritten as a new dataset $\hat{B} = \{\hat{b}_1, \hat{b}_2, \hat{b}_3, \dots, \hat{b}_l\} \in R^{s \times l}$, where s is the total number of $\hat{b}_i (1 \leq i \leq l)$.

2.2 Double Graph Model

In [3], a multi-graph structure with spectral clustering method is proposed that can effectively represent the relationship between bands. In this section, we construct a

double graph model to capture the complex relationships between pairwise spectral bands with statistical property of space. In graph model, each vertex represents a band b_i and each edge corresponds to the correlation between pairwise bands. The correlation of vertices can be described as an adjacency matrix which can be defined by a Gram matrix. The adjacency matrix of bands is defined as follows:

$$L_B = B^T B \tag{1}$$

And the adjacency matrix of neighbour information data can be expressed as follows:

$$L_{\hat{B}} = \hat{B}^T \hat{B} \tag{2}$$

L_B and $L_{\hat{B}}$ are both $l \times l$ positive semidefinite similarity matrices. According to the reference [18], it is clear that

$$\det(L_Y) = \text{Vol}^2(\{b_i\}_{i \in Y}) \tag{3}$$

where $\det(L_Y)$ is equal to the squared k -dimensional volume of the parallelepiped spanned by the $\{b_i\}_{i \in Y}$ of B corresponding to bands in subset Y , where k is the dimension of subset Y .

2.3 Search Criterion

To select a band subset with rich information, the evaluation criteria of maximum information entropy [4] is used here. The information entropy of the spectral band subset is defined as:

$$\max H_k = -\frac{1}{k} \sum_{i=1}^k \sum_{x=1}^n P(p_x^i) \log P(p_x^i) \tag{4}$$

where p_x^i represents the pixel value of the i -th spectral band and k is the number of band subset. Equation (4) can be used as the evaluation criteria for the performance of each band subset, which is called **MI** for short.

2.4 MI-DPP Band Selection Method

The k-DPP [14] can be expressed as follows:

$$\sum_{|Y'|=k} \det(L_{Y'}) = \det(L + I) \sum_{|Y'|=k} P_L(Y') \tag{5}$$

where $P_L(Y')$ represents the probability of any band subset Y' with dimension k . According to Eq. (5), k-DPP can be expressed as follows:

$$P_L^k(Y) = \frac{\det(L_Y)}{\sum_{|Y'|=k} \det(L_{Y'})} \tag{6}$$

According to reference [13], we can get

$$P_L^k = \frac{1}{e_k^N} \det(L + I) P_L = \frac{1}{e_k^N} \sum_{|Y|=k} P^{V_Y} (Y') \prod_{n \in Y} \lambda_n \quad (7)$$

Equation (7) is the mathematical expression of k-DPP which shows that P_L^k is a mixture of the elementary DPPs P^{V_Y} .

With full use of the spatial information, our algorithm can be written as follows:

$$P_{L_B, L_{\hat{B}}}^k(Y) = \frac{1}{e_{B,k}^N} \sum_{|Y|=k} u P^{V_Y^B} (Y') \prod_{n \in Y} \lambda_n^B + \frac{1}{e_{\hat{B},k}^N} \sum_{|Y|=k} \hat{u} P^{V_Y^{\hat{B}}} (Y') \prod_{n \in Y} \lambda_n^{\hat{B}} \quad (8)$$

where u and \hat{u} are scalars that balance the weight of two k-DPPs. Equation (8) shows that our algorithm is a mixture of two elementary DPPs $P^{V_Y^B}$ and $P^{V_Y^{\hat{B}}}$, where λ_n^B and $\lambda_n^{\hat{B}}$ are the eigenvalues of the adjacent matrices L_B and $L_{\hat{B}}$ respectively. $e_{B,k}^N = \sum_{|Y|=k} \prod_{n \in Y} \lambda_n^B$ and $e_{\hat{B},k}^N = \sum_{|Y|=k} \prod_{n \in Y} \lambda_n^{\hat{B}}$ are the eigenvalue polynomials of the adjacent matrices L_B and $L_{\hat{B}}$, respectively. $P_{L_B, L_{\hat{B}}}^k(Y)$ is the probability of sampling a band subset Y . Combined with the search criterion MI mentioned in Sect. 2.3, we call the proposed algorithm as MI-DPP. Details of the sampling process of MI-DPP are summarized in Algorithm 1.

Algorithm 1 Sampling Process of MI-DPP

Input:

- B : a hyperspectral image
- k : number of selected bands

Output:

Y : index of feature vectors

- 1) Computing similarity matrix: $L_B = B^T B$ $L_{\hat{B}} = \hat{B}^T \hat{B}$
- 2) Characteristic decomposition: $\{\lambda_n^B, \mathbf{v}_n^B\}_{n=1}^l \leftarrow L_B$ $\{\lambda_n^{\hat{B}}, \mathbf{v}_n^{\hat{B}}\}_{n=1}^l \leftarrow L_{\hat{B}}$
- 3) for $\mathbf{n} = 1, \dots, 2, \mathbf{1}$ do
- 4) select eigenvectors sets V_B and $V_{\hat{B}}$ with cardinality \mathbf{k} with probability $\lambda_n^B \frac{e_{B,k-1}^{n-1}}{e_{B,k}^n}$ and $\lambda_n^{\hat{B}} \frac{e_{\hat{B},k-1}^{n-1}}{e_{\hat{B},k}^n}$
- 5) end for
- 6) while $|\mathbf{V}_B| > \mathbf{0}$ and $|\mathbf{V}_{\hat{B}}| > \mathbf{0}$ do
- 7) select band \mathbf{i} with probability:

$$P_r(\mathbf{i}) = u \frac{1}{|\mathbf{V}_B|} \sum_{\mathbf{v}_n^B \in V} (\mathbf{v}_n^B{}^T \mathbf{e}_i)^2 + \hat{u} \frac{1}{|\mathbf{V}_{\hat{B}}|} \sum_{\mathbf{v}_n^{\hat{B}} \in V_{\hat{B}}} (\mathbf{v}_n^{\hat{B}}{}^T \mathbf{e}_i)^2$$
- 8) If $\mathbf{u} \sim U[\mathbf{0}, \mathbf{1}] < P_r(\mathbf{i})$
 $\mathbf{Y} \leftarrow \mathbf{Y} \cup \{\mathbf{i}\}$
- 9) end if
- 10) $\mathbf{V}_B \leftarrow \mathbf{V}_{B\perp}, \mathbf{V}_{\hat{B}} \leftarrow \mathbf{V}_{\hat{B}\perp}$ an orthonormal basis for the subspace of $\mathbf{V}_B, \mathbf{V}_{\hat{B}}$ orthogonal to \mathbf{e}_i
- 11) end while

Output: Y

We repeat Algorithm 1 for α times to obtain α spectral band subsets Y with the dimension k . Then we select the band subset that meets the maximum search criteria H_k from the α spectral band subsets. With the proposed algorithm MI-DPP, the selected band subset will have the advantages of low redundancy and high information.

3 Experimental Result

In this section, we test the performances of the proposed algorithm and compare with several well-known unsupervised selection algorithms, including MIC [18] and Lscore [19].

3.1 Hyperspectral Datasets

AVRIS sensor: Indian Pines dataset

The India Pines dataset was obtained by the Airborne Visible/Infrared Imaging Spectrometer sensor (AVIRIS) in 1992 [20]. The image covers Indian Pines test site in North-western Indiana which has 220 spectral bands covering the spectrum range of 0.2–2.4 μm . And 20 water absorption spectral bands [104–108], [150–163] are removed and remaining 200 spectral bands form the dataset. It has 145×145 spatial pixels, including 16 classes ground truth.

Table 1. The numbers of training and test samples for each category of the Indian Pines image

Class	Training	Test	Samples
1. Alfalfa	4	50	54
2. Corn - notill	142	1292	1434
3. Corn - mintill	83	751	834
4. Corn	23	211	234
5. Grass - pasture	49	448	497
6. Grass - trees	73	674	747
7. Grass - pas - turemowed	3	23	26
8. Hay - windrowed	48	431	489
9. Oats	2	18	20
10. Soybean - notill	97	871	968
11. Soybean - mintill	245	2223	2468
12. Soybean - clean	59	555	614
13. Wheat	20	192	212
14. Woods	127	1167	1294
15. Building - grass-trees	39	341	380
16. Stone - steel-towers	9	86	95
Total	1023	9343	10366

3.2 Experimental Parameter Setup

In this section, we will introduce the parameter settings. The weight scalar u and \hat{u} are set to be $u = \hat{u} = 0.5$. The number of iterations $\alpha = 5$. The number of selected bands increased by 10 and up to 100 bands. We use the SVM algorithm [21] to test the performance of the selected bands. The parameters C and γ of RBF kernel of SVM are in the range of $10^{(1:4)}$ and $2^{(-4:4)}$, and are determined by five-fold cross validation. In all experiments, 10% of the samples were randomly selected for training, and the rest of the samples were used for testing. In Table 1, we show the numbers of training samples and test samples for Indian Pines dataset. The results of the experiment are evaluated through three widely used indexes: overall accuracy (OA), average accuracy (AA) and kappa coefficient [22].

3.3 Runtime Analysis

As showed in [18], k-DPP has a time complexity of $\mathcal{O}(Nk^3)$. It spends $\mathcal{O}(N^3)$ time to decompose s adjacent kernel matrices of bands. The MI-DPP algorithm takes total time $\mathcal{O}(nN^3 + nNk^3)$ to sample k bands, where n is the number of the similar kernel matrices. The computation time of the algorithm will greatly increase when N and k are large. However, the numbers of N and k are normally small in the selecting progress. We pre-calculate the information entropy for each band and establish lookup tables for Indian Pines dataset. Establishing lookup tables for Indian Pines dataset takes only 0.05 s. All the experiments are carried out on Matlab with Intel CPU E5-1620@3.5 GHz with 32 GB RAM.

In Fig. 1, the computing time of the three algorithms is shown as the number of selected spectral bands increases from 10 to 100. It's obvious that the calculation time of the proposed MI-DPP is the least comparing with the other algorithms when the dimension of band subset is low. As the number of selected spectral bands increases,

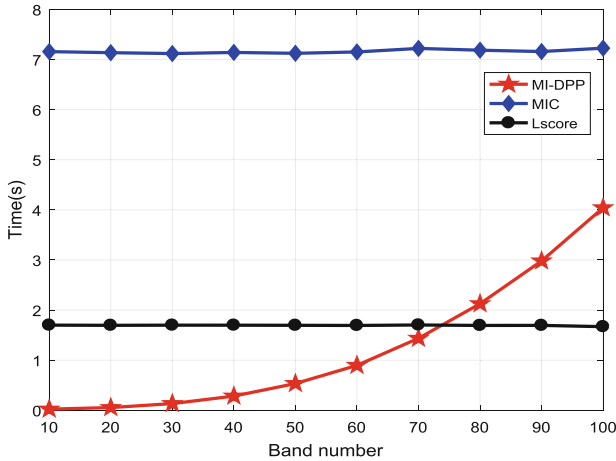


Fig. 1. Runtime of MI-DPP, MIC, Lscore on the Indian Pines

the calculation time of the MIC and Lscore does not change significantly. The calculation time of the algorithm MI-DPP will increase as the dimension of selected band subset increases. In general, the proposed algorithm MI-DPP takes less time than the other band selection algorithms.

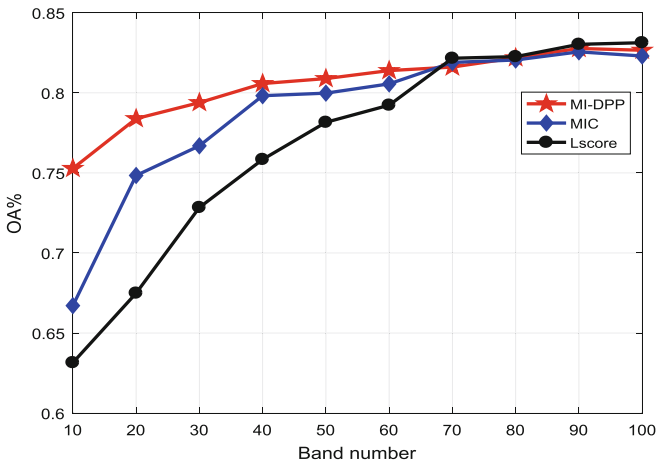


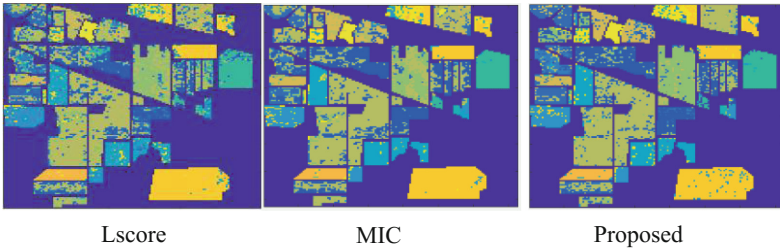
Fig. 2. Classification accuracy of the three band selection algorithms on Indian Pines

3.4 Classification Results

In this subsection, we test the performance of the selected subsets of spectral bands on Indian Pines image. In Fig. 2, we compared the classification accuracy of the band subset selected by different algorithms. When the dimension of band subsets is in the range of 10–70, our proposed band selection method MI-DPP is significantly better than other algorithms. The algorithm Lscore performs the worst. When the dimension of spectral band subsets is in the range of 70–100, the performances of the three algorithms are similar. This is because as the number of bands is increasing, the amount of information in the spectral band increases. In the following, we will analyse the performance of the proposed MI-DPP in the low dimension in detail. We set the dimension of selected band subset to be 10 for each dataset and perform the classification tests. The results are shown in Table 2. In Fig. 3, we show the classification maps of three band selection algorithms on Indian Pines. It can be seen that among these three algorithms, Lscore obtains the worse result. Compared with the other two algorithms, the classification accuracy of the optical band selected by the proposed MI-DPP in all categories has been significantly improved. This is because the spectral band subset selected by the algorithm MI-DPP are more diverse and low-redundant which can provide abundant information.

Table 2. Classification accuracy of the band set on the Indian pines image selected by Lscore, MIC and the proposed MIMN-DPP

Class	Lscore	MIC	Proposed
1. Alfalfa	22.0 ± 19.7	32.6 ± 25.1	44.6 ± 16.6
2. Corn - notill	46.7 ± 3.5	55.3 ± 3.8	60.2 ± 4.8
3. Corn - mintill	34.2 ± 3.4	32.6 ± 5.8	55.6 ± 2.5
4. Corn	10.9 ± 7.5	12.4 ± 6.3	53.7 ± 7.9
5. Grass - pasture	78.1 ± 3.9	83.2 ± 4.4	88.5 ± 2.7
6. Grass-trees	88.4 ± 3.1	90.8 ± 1.9	91.8 ± 2.1
7. Grass - pas - turemowed	26.9 ± 78.9	63.4 ± 21.7	82.2 ± 10.1
8. Hay - windrowed	82.1 ± 6.1	95.6 ± 1.9	95.7 ± 2.5
9. Oats	11.1 ± 11.7	16.1 ± 13.7	57.2 ± 27.3
10. Soybean - notill	55.6 ± 3.6	44.3 ± 6.8	65.1 ± 6.1
11. Soybean - mintill	71.5 ± 2.3	75.1 ± 2.7	79.6 ± 3.1
12. Soybean - clean	36.4 ± 3.4	52.7 ± 4.7	66.1 ± 3.1
13. Wheat	91.3 ± 4.7	90.1 ± 4.9	94.4 ± 3.1
14. Woods	92.6 ± 1.3	93.6 ± 0.6	94.5 ± 1.3
15. Building - grass-trees	24.0 ± 5.4	37.1 ± 3.5	47.2 ± 6.5
16. Stone - steel-towers	81.4 ± 8.7	77.0 ± 16.6	87.7 ± 8.6
OA (%)	63.1 ± 0.1	66.7 ± 0.9	75.2 ± 0.7
AA (%)	53.3 ± 2.6	59.5 ± 3.2	72.8 ± 2.4
Kappa (%)	57.4 ± 0.9	61.5 ± 1.5	71.5 ± 0.8

**Fig. 3.** Classification maps of three band selection algorithms on Indian Pines

4 Conclusion

In this paper, an unsupervised band selection algorithm called MI-DPP is proposed for hyperspectral images. We use the evaluation criterion to evaluate the spectral band performance, aiming to maintain the maximum band information. A double graph model is used to capture the complex relationship between bands. Furthermore, an improved k-DPP is proposed to sample a diversity and low-redundancy band subset from the two-graph structure. The experimental results on Indian Pines image dataset show that the proposed algorithm has a good expression in the task of band selection.

Acknowledgements. This work is supported by the National Nature Science Foundation of China (nos. 61471132 and 61372173), and the Training program for outstanding young teachers in higher education institutions of Guangdong Province (no. YQ2015057).

References

1. Qiao, T., Yang, Z., Ren, J., et al.: Joint bilateral filtering and spectral similarity-based sparse representation: a generic framework for effective feature extraction and data classification in hyperspectral imaging. *Pattern Recogn.* **77**, 316–328 (2017)
2. Qiao, T., Ren, J., et al.: Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis. *IEEE Trans. Geosci. Remote Sens.* **55**(1), 119–133 (2017)
3. Yuan, Y., Zheng, X., Lu, X.: Discovering diverse subset for unsupervised hyperspectral band selection. *IEEE Trans. Image Process.* **26**(1), 51–64 (2017)
4. Feng, J., Jiao, L., Liu, F., Sun, T., Zhang, X.: Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images. *Pattern Recogn.* **51**, 295–309 (2016)
5. Zabalza, J., Ren, J., Yang, M., et al.: Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing. *ISPRS J. Photogram. Remote Sens.* **93**, 112–122 (2014)
6. Zabalza, J., et al.: Structured covariance principal component analysis for real-time onsite feature extraction and dimensionality reduction in hyperspectral imaging. *Appl. Opt.* **53**(20), 4440–4449 (2014)
7. Ren, J., Zabalza, J., Marshall, S., Zheng, J.: Effective feature extraction and data reduction in remote sensing using hyperspectral imaging. *IEEE Sig. Process. Mag.* **31**(4), 149–154 (2014)
8. Sotoca, J.M., Pla, F.: Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recogn.* **43**(6), 2068–2081 (2010)
9. Yang, H., Du, Q., Su, H., Sheng, Y.: An efficient method for supervised hyperspectral band selection. *IEEE Geosci. Remote Sens. Lett.* **8**(1), 138–142 (2011)
10. Sui, C., Tian, Y., Xu, Y., Xie, Y.: Unsupervised band selection by integrating the overall accuracy and redundancy. *IEEE Geosci. Remote Sens. Lett.* **2**(1), 185–189 (2015)
11. Wang, C., Gong, M., Zhang, M., Chan, Y.: Unsupervised hyperspectral image band selection via column subset selection. *IEEE Geosci. Remote Sens. Lett.* **12**(7), 1411–1415 (2015)
12. Zhang, M., Ma, J., Gong, M.: Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization. *IEEE Geosci. Remote Sens. Lett.* **4**(5), 773–777 (2017)
13. Chang, C., Wang, S.: Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **44**(6), 1575–1585 (2006)
14. Jia, S., Ji, Z., Qian, Y., Shen, L.: Unsupervised band selection for hyperspectral imagery classification without manual band removal. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **5**(2), 531–543 (2012)
15. Geng, X., Sun, K., Ji, L., Zhao, Y.: A fast volume-gradient-based band selection method for hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **52**(11), 7111–7119 (2014)
16. Kulesza, A., Taskar, B.: Determinantal point processes for machine learning. *Found. Trends Mach. Learn.* **5**(2–3), 123–286 (2012)
17. Kulesza, A., Taskar, B.: k-DPPs: fixed-size determinantal point processes. In: *Proceedings of International Conference on Machine Learning*, pp. 1193–1200 (2016)

18. Mitra, P., Murthy, C.A., Pal, S.K.: Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(3), 301–312 (2002)
19. He, X., Cai, D., Niyogi, P.: Laplacian score for feature selection. In: *NIPS*, pp. 507–514 (2006)
20. Green, R.O., et al.: Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.* **65**(3), 227–248 (1998)
21. Bazi, Y., Melgani, F.: Toward an optimal SVM classification system for hyper-spectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **44**(11), 3374–3385 (2006)
22. Foody, G.M.: Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **80**, 185–201 (2002)