Pattern Recognition, Volume 102, June 2020, Article number 107213
DOI:10.1016/j.patcog.2020.107213

MIMN-DPP: Maximum-Information and Minimum-Noise Determinantal Point Processes for Unsupervised Hyperspectral Band Selection

Weizhao Chen [a], Zhijing Yang [a,*], Jinchang Ren [b,c], Jiangzhong Cao [a], Nian Cai [a], Huimin Zhao [b], Peter Yuen [d]

[a] School of Information Engineering, Guangdong University of Technology, Guangzhou, China.
[b] School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China
[c] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
[d] Electro-Optics & Remote Sensing, Centre for Electronics Warfare, Information & Cyber (CEWIC), Cranfield University, Swindon, UK

Abstract:

Band selection plays an important role in hyperspectral imaging for reducing the data and improving the efficiency of data acquisition and analysis whilst significantly lowering the cost of the imaging system. Without the category labels, it is challenging to select an effective and low-redundancy band subset. In this paper, a new unsupervised band selection algorithm is proposed based on a new band search criterion and an improved Determinantal Point Processes (DPP). First, to preserve the original information of hyperspectral image, a novel band search criterion is designed for searching the bands with high information entropy and low noise. Unfortunately, finding the optimal solution based on the search criteria to select a low-redundancy band subset is a NP-hard problem. To solve this problem, we consider the correlation of bands from both original hyperspectral image and its spatial information to construct a double-graph model to describe the relationship between spectral bands. Besides, an improved DPP algorithm is proposed for the approximate search of a low-redundancy band subset from the double-graph model. Experiment results on several well-known datasets show that the proposed optical band selection algorithm achieves better performance than many other state-of-the-art methods.

Keywords: Hyperspectral images (HSI); unsupervised band selection; maximum information and minimum noise (MIMN) criterion; Determinantal Point Processes (DPP)

*Corresponding author. Tel.: +86-20-39322438; Fax: +86-20-39322253. E-mail address: yzhj@gdut.edu.cn (Zhijing Yang)

1. Introduction

With two-dimensional information in the spatial domain and one additional dimension in the spectral domain, hyperspectral images (HSI) contains a three-dimensional structure. At a given spectral frequency or wavelength, a grey-scale image can be formed namely a band image, corresponding to the reflectance response of the target objects. The reflectivity values of different wavelengths, namely the spectral signature or spectral profile, can effectively represent the properties of the objects in terms of temperature, moisture (water contents) and change of chemical components. These advantages have enabled hyperspectral images as a unique way for effective land-mapping in the field of earth observation and remote sensing [1-2]. Therefore, it's essential to effectively utilize the spectral-spatial features for hyperspectral image analysis, especially for data classification [45-46]. Most of the recent works [47] have utilized all spectral bands for hyperspectral image classification and achieved good classification accuracy, where the redundancy of the spectral bands has not been considered.

Although increasing the number of spectral bands can enhance the representation of ground objects, actually not all the bands are useful for charactering the objects in hyperspectral image processing [3]. High dimensional data in HSI may bring a series of problems, such as data redundancy in spectral domain, noise in certain spectral bands, heavy cost of the system, and lengthy process for data acquisition, transmission and analysis. Some redundant or noise bands may cause interference to the subsequent data processing, resulting in a reduction of accuracy. In addition, high dimensional data with limited sampling pixels may lead to the curse of dimensionality, or the Hughes phenomenon [4]. Therefore, dimensionality reduction has become one crucial task in HSI for effective data acquisition and analysis.

Techniques for dimensionality reduction in HSI can be divided into two categories: (1) feature extraction [6-8]; (2) band selection [9-11]. Feature extraction [12] is the projection of the original high dimensional data onto a lower dimensional space to reveal the distinctive properties of the original data, and typical approaches include principal component analysis (PCA) and its variations [13-15]. Band selection, however, is to select a subset of bands from the original spectral band set whilst maximizing the performance of data analysis [5, 16-18] . Although the dimensionality of features can be reduced, the full

spectral data is still needed in feature extraction based methods. Band selection approaches, on the other hand, may only need a selected subset of spectral bands for the following data analysis. In comparison with feature extraction methods, band selection can better preserve the physical information and have better interpretation and expression ability of the original data [5].

In general, band selection techniques can be divided into two groups, i.e. supervised band selection [19-20] and unsupervised band selection[21-24], depending on whether the associated class labels of data are known or not. With known class labels, supervised band selection can be used to select spectral bands that have a strong correlation with the class labels. Therefore, compared with the unsupervised band selection methods, supervised band selection methods often have better performance [25]. However, the application of supervised band selection can be constrained due to lack of labelled information in practical cases. As a result, unsupervised band selection without labelling data becomes more feasible in such cases.

Conventional unsupervised band selection has two main steps, i.e. i) establishing the selection criteria for selecting the band subsets, and ii) determining the search method to select the band subset from the original dataset. Some popular selection criteria are Spectral Inter-Band Discrimination Capacity [26] and maximum information and minimum redundancy [5]. Typical search approaches for selecting band subset include the ranking-based [38], the clustering-based [39], and the greedy-search based algorithms [17]. For ranking-based methods, all spectral bands are ranked based on the given evaluation criteria and then the spectral bands with top ranks are selected for the desired number of dimensions. In [38], Chang et al proposed a ranking-based method called CEM-BCM. All spectral bands are ranked based on their correlation with the entire hyperspectral image and highly correlated bands are preferred. For clustering-based methods, spectral bands are divided into different clusters and similar bands fall into the same cluster. Then, the band subset is selected from different clusters to form a low-redundancy band subset. In [39], Sen et al proposed an Affinity Propagation-based clustering method for selecting representative spectral bands from the dataset after denoising. For greedy algorithms, the selected band subset satisfies the optimal criteria as much as possible by adding or removing spectral band. In [17], the SBS-MEV

algorithm was used to remove one band each time, which maximized the ellipsoid volume of the candidate band subset.

Most existing rank-based band selection methods only assign a single score to each band whilst ignore the latent structure between spectral bands in a high dimensional space [3]. This will result in individually selecting of each band without global considerations. In addition, many band selection methods do not consider the band redundancy from the spectral-spatial information. The spectral-spatial information can have more comprehensive measurement of redundancy than spectral information only. In [3], Yuan et al proposed a multiple graph to describe the complex structure between spectral bands based on the spectral clustering method. However, constructing a multiple graph structure by the spectral clustering needs very high computational complexity and memory requirements. This will be tackled in our proposed approach, where a double-graph structure is utilized to reduce the computational complexity whilst maintaining high performance as discussed below.

In conventional band search approaches, finding a low-redundancy band subset from a graph model is very challenging as it is a NP-hard problem [3]. As a probabilistic model, Determinantal Point Processes (DPP) [27] can be used as a search algorithm to choose the feature subset with high diversity and low redundancy. This has been widely used in tasks of image, video or texts summarization, image search and news threading [27]. As an extension to DPP, k-DPP [29] specifies the size of the selected subset to be k for efficiency. However, the original k-DPP can only be applied to a single graph model, and it only considers low-redundancy in selecting the bands rather than any other properties such as the quantity of noise in the band.

In order to tackle the aforementioned problems, a double-graph based DPP model was proposed in this paper for unsupervised band selection, where both spectral and spatial information are used to measure the redundancy between the bands. A Maximum Information entropy and Minimum Noise (MIMN) criterion is introduced for determining the best band subset by preserving a rich amount of information in the band subset whilst minimizing noise level for achieving a high accuracy of data analysis i.e. classification.

The main contributions of this paper can be summarized as follows:

1) To build a double-graph model using both the spectral and spatial information of pixels for effectively measuring the correlation between bands;

2) To propose a new maximum information and minimum noise (MIMN) search criterion for determining the optimal band subset. As maximum entropy may be attributed to a high degree of noise, minimum noise constraint can supplement for robustness;

3) By combining k-DPP with MIMN, a MIMN-DPP approach is proposed for selecting the band subset with not only low-redundancy and high diversity but also rich information and low noise. This will result in high classification accuracy and improved robustness despite of the random sampling in k-DPP.

The rest of this paper is organized as follows: Section 2 briefly introduces the relevant techniques of the proposed approach, including information entropy and DPP. In Section 3, details of the proposed MIMN-DPP algorithm for unsupervised band selection are discussed. The experiment results and discussions in comparison with several state-of-the-art approaches are given in Section 4. Finally, some concluding remarks are drawn in Section 5.

2. Related background

2.1 Information entropy

Information entropy is for quantitative measurement of the information as proposed by Shannon [28].Given a random variable X, its associated information entropy can be defined by:

$$H(X) = -\sum_{x} \mathrm{P}(x) \log \mathrm{P}(x) \tag{1}$$

where $P(x)$ is the probability density function of X. The information entropy is only related to the probability distribution of the random variable X. The more uneven the variable X is, the greater the information entropy is.

In the field of image processing, the information entropy can be used to measure the information of an image, where higher information entropy means richer information and vice versa. In this paper, X is used to denote an image, and the variable x is a pixel point of the image. The information entropy of the

image only depends on the probability distribution of the pixel values. In a hyperspectral image, each spectral band can be regarded as a two-dimensional image, where its corresponding information entropy can be determined by Eq. (1).

2.2 Determinantal Point Processes (DPP)

DPPs are elegant probabilistic models of repulsion that arise from quantum physics and random matrix theory [27], which provides an effective sampling strategy for selecting features with diversity. DPP has been successfully applied in a wide range of applications, such as text summarization, image search and news threading [27].

Given a discrete finite set $\gamma = \{b_1, b_2, \cdots, b_N\}$ with N elements, the number of total subsets of γ is 2^N. According to [27,30], DPP can be defined as follows:

$$\mathrm{P}(B \subseteq \gamma) = \det(K_B) \tag{2}$$

where B is an arbitrary subset of γ, and $P(B)$ represents the probability density function of subset B.

Let K be a $N \times N$ symmetric matrix whose elements are indexed by the corresponding elements of γ. K_B is a symmetric matrix and its entries are indexed by the corresponding elements of B; and $\det(K_B)$ is one of the principal minor of the determinant $\det(K)$.

Assume $K = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1N} \\ K_{21} & K_{22} & \cdots & K_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ K_{N1} & K_{N2} & \cdots & K_{NN} \end{bmatrix} \in R^{N \times N}$ is a marginal probability kernel, indicating the possibility of the corresponding selected elements [27]. As Eq. (2) describes the probability of the subset B, the $\det(K_B)$ must be positive, i.e. K must be a positive semidefinite matrix.

When the subset B contains only one element $B = \{b_i\}$, the probability of selecting subset B can be given by:

$$\begin{aligned} \mathrm{P}(B \subset \gamma) &= \det(K_{ii}) \\ &= \left| K_{ii} \right| \end{aligned} \tag{3}$$

where the larger value of K_{ii} implies that the corresponding element $b_i \in \gamma$ is more likely to be chosen by the DPP measure.

For the case of two-elements as $B = \{b_i, b_j\}$, we have

$$\begin{aligned} P(B \subset \gamma) &= \det(K_B) \\ &= \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix} \\ &= K_{ii} \times K_{jj} - K_{ij}^2 \end{aligned} \tag{4}$$

Obviously, the larger value of K_{ii} and K_{jj} indicates that the element b_i and b_j are more likely selected in the subset B.

The large value of K_{ij} means that b_i and b_j are unlikely to co-occur. Eq. (4) shows the most important characteristic of 'Diversity' of the DPP. K can be used as a measure of correlation between elements in the ground set γ. That is if the element b_i has a strong correlation with the element b_j, K_{ij} will have a large value. In the new subset B, taking K_{ij} as the characteristics of 'diversity', a larger value means that the two elements b_i and b_j have lower diversity.

As elements in K are within the range of 0 and 1, it is challenging to construct the marginal kernel matrix for describing real data. In order to model real data, DPP is written in the form of L-ensembles [27]. Given a symmetric positive semidefinite matrix L indexed by the elements of γ, the L-ensembles representation of DPP is given by:

$$P_L(\boldsymbol{Y} = Y) \propto \det(L_Y) \tag{5}$$

where L is a metric matrix that indicates the similarity of any two elements in the γ, Y is an arbitrary subset of γ, and $L_Y \equiv [L_{ij}]_{i,j \in Y}$ is a symmetric matrix indexed by the elements of Y. For any pair of elements $\{i, j\} \subseteq Y$, L_{ij} represents the degree of correlation between them. Eq. (5) actually indicates that the probability of any subset Y is proportional to the value of the determinant $\det(L_Y)$. We use $\sum_{Y \subseteq \gamma} \det(L_Y)$ to represent the sum of the determinants of all subsets of γ. It can be proved that [27]:

$$\sum_{Y\subseteq\gamma}\det(L_Y)=\det(L+I_N) \tag{6}$$

where I is an $N\times N$ identity matrix.

Combining Eqs. (5) and (6), the standard form of DPP can be obtained by:

$$\begin{aligned} P_L(\boldsymbol{Y}=Y) &= \frac{\det(L_Y)}{\sum_{Y\subseteq\gamma}\det(L_Y)} \\ &= \frac{\det(L_Y)}{\det(L+I_N)} \end{aligned} \tag{7}$$

The larger P_L is, the greater the probability of the subset Y being selected will be.

3. The proposed MIMN-DPP framework

In this section, we will discuss in detail the proposed MIMN-DPP method for unsupervised band selection. According to the flowchart shown in Fig.1, MIMN-DPP contains three main parts, including neighbourhood information extraction, double-graph model reconstruction and DPP-based optimal band selection. The technical details of these three parts as well as the notations of hyperspectral data are discussed as follows.

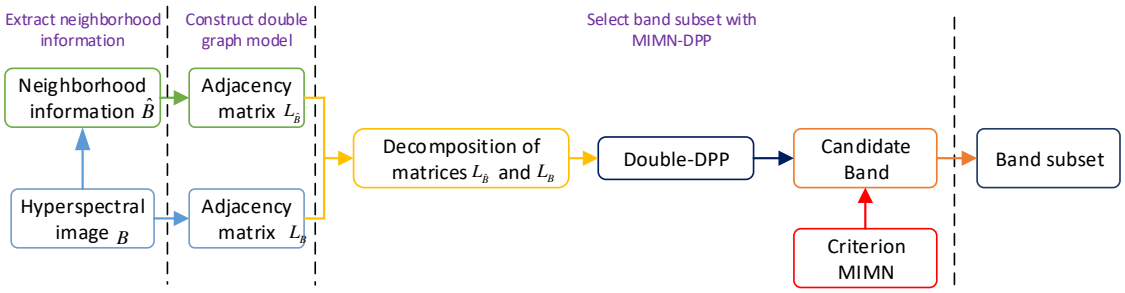


Figure 1: Flowchart of the proposed method.

3.1 Notations of Hyperspectral data

Hyperspectral data sets can be represented as $B=\{b_1,b_2,b_3,\cdots,b_l\}\in R^{n\times l}$ where n represents the number of pixels in each spectral band, l is the total number of spectral bands, and $b_m\,(1\le m\le l)$ indicates the m-th spectral band. As shown in Fig. 2, the red area represents the pixel point p_m^j, and the green area is its neighbourhood. We use $\hat{p}_m^j$ to denote the mean spectrum of the neighbourhood of p_m^j. For convenience,

we combine the means of the neighbourhood of all the pixels into a new data set and denote it by $\hat{B} = \{\hat{b}_1, \hat{b}_2, \hat{b}_3, \cdots, \hat{b}_l\} \in R^{n \times l}$.

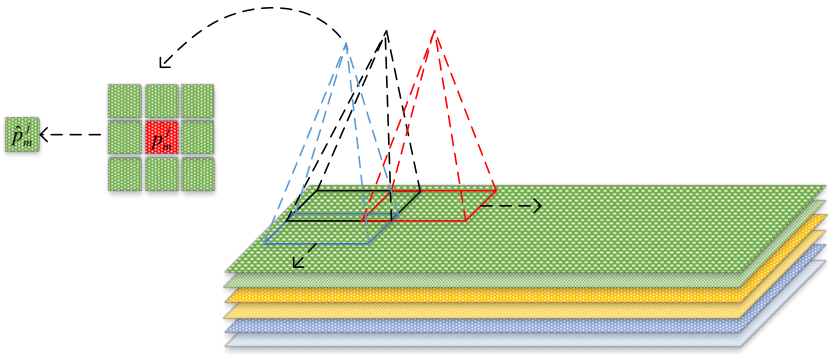


Figure 2: Neighbourhood information of p_m^j.

3.2 Double graph mode

The graph model can effectively represent the relationship between vertices. In [3], a spectral clustering based multi-graph structure is proposed to represent the correlations between bands. However, the spectral clustering methods require high computational cost. In this section, considering the pixel and its neighbourhood information, we construct a double graph model for capturing the complex relationships between the pairwise spectral bands.

An undirected graph model $G = (V, E)$ contains a set of vertices V and a set of edges E. We consider the bands of hyperspectral data as a graph model, where each vertex represents a band b_i and each edge between two vertices corresponds to the correlation between the two bands. The correlation of vertices can be represented as an adjacency matrix. We use mutual information to define the adjacency matrix, which can effectively represent the correlation between pairwise bands. The adjacency matrix of bands is defined as follows:

$$L_B = \begin{bmatrix} I(b_1,b_1) & I(b_1,b_2) & \cdots & I(b_1,b_l) \\ I(b_2,b_1) & I(b_2,b_2) & \cdots & I(b_2,b_l) \\ \vdots & \vdots & \ddots & \vdots \\ I(b_l,b_1) & I(b_l,b_2) & \cdots & I(b_l,b_l) \end{bmatrix} \in R^{l \times l} \tag{8}$$

The mutual information between two bands $b_{m1}(1 \le m1 \le l)$ and $b_{m2}(1 \le m2 \le l)$ can be estimated by a fast calculation algorithm proposed in [41] as follows:

$$I(b_{m1}, b_{m2}) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{n\sum_{j=1}^{n} e^{-\frac{1}{2h^2}((p_{m1}^i - p_{m1}^j)^2 + (p_{m2}^i - p_{m2}^j)^2)}}{\sum_{j=1}^{n} e^{-\frac{1}{2h^2}(p_{m1}^i - p_{m1}^j)^2} \sum_{j=1}^{n} e^{-\frac{1}{2h^2}(p_{m2}^i - p_{m2}^j)^2}} \tag{9}$$

where h is the kernel width which has been estimated properly in the reference [42].

Similarly, the adjacency matrix of neighbour data can also be calculated by using mutual information as:

$$L_{\hat{B}} = \begin{bmatrix} I(\hat{b}_1, \hat{b}_1) & I(\hat{b}_1, \hat{b}_2) & \cdots & I(\hat{b}_1, \hat{b}_l) \\ I(\hat{b}_2, \hat{b}_1) & I(\hat{b}_2, \hat{b}_2) & \cdots & I(\hat{b}_2, \hat{b}_l) \\ \vdots & \vdots & \ddots & \vdots \\ I(\hat{b}_l, \hat{b}_1) & I(\hat{b}_l, \hat{b}_2) & \cdots & I(\hat{b}_l, \hat{b}_l) \end{bmatrix} \in R^{l \times l} \tag{10}$$

where $\hat{b}_l$ denotes the neighbourhood of b_i.

Let Y be the selected band subset, according to Ref. [27] we have

$$\det(L_Y) = Vol^2(\{b_m\}_{m \in Y}) \tag{11}$$

where $\det(L_Y)$ is equal to the squared k-dimensional volume of the parallelepiped spanned by the $\{b_m\}_{m \in Y}$ of B corresponding bands in the subset Y, and k is the cardinal number of the subset Y.

An example is shown in Fig. 3, where Y contains two selected bands b_{m1} and b_{m2}. If the band vectors are more orthogonal, their span volumes are larger which means the band vectors are more diverse. In contrast, parallel band vectors have lowest span volumes, hence they are less likely to be selected in the same subset. This is the most important property of DPP: the more the diversity of the bands is, the greater the probability appearing in the same subset of the two bands will be. More details will be described in the following subsection.

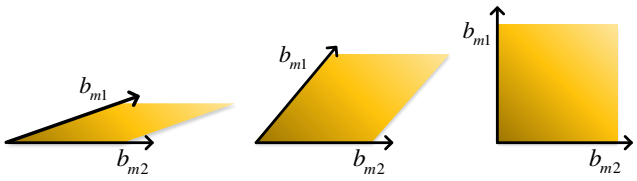


Figure 3: Geometry structure of pairwise bands.

3.3 Search criterion

In order to preserve the information of the original data, the maximization information criterion [5] is used. The information entropy of the m-th band, b_m, is defined by:

$$H(b_m) = -\sum_{j=1}^{n} \mathrm{P}(p_m^j) \log \mathrm{P}(p_m^j) \tag{12}$$

where j is the index of the pixel, n represents the total number of pixels in each spectral band and $\mathrm{P}(p_m^j)$ denotes the probability density function of p_m^j.

Let $p_{m,d}^i (1 \le d \le q)$ represent the set of the adjacent pixels of p_m^i where q denotes the size of the neighbourhood. The information entropy of each image block is used for estimating the noise level of the neighborhood. The level of noise of i-th image block in m-th band is obtained as:

$$N_i(b_m^i) = -\sum_{d=1}^{q} \mathrm{P}(p_{m,d}^i) \log \mathrm{P}(p_{m,d}^i) \tag{13}$$

If all the adjacent pixels of p_m^i have the same value, $N_i(b_m^i)$ will reach the minimal value. Conversely, if the values of all the adjacent pixels of p_m^i are randomly generated, i.e. the image block around p_m^i is noisy, $N_i(b_m^i)$ will achieve the maximum. Therefore, the smaller $N_i(b_m^i)$ is, the smoother the local block and the lower the noise level is.

As suggested in [43-44], the range of values from minimum to 1.2 multiplied by the mean of each block noise are equally divided into 150 intervals. The number of image block noise falling into each interval is counted. The noise values of image blocks in the interval, which contains most image blocks, are used to calculate their mean as the estimation noise of the whole band. The noise level of m -th band is represented by $N(b_m)$. By combining Eqs. (12) and (13) , the MIMN search criteria is derived as:

$$\max \varphi_m$$
$$\varphi_m = \frac{H(b_m)}{N(b_m)} \tag{14}$$

Equation (14) can be actually used to rank the suitability of each band. We call the above search criteria as the maximum (band) information and minimum (local) noise criterion, denoted by MIMN for short.

3.4 The proposed MIMN-DPP band selection model

In DPP, the cardinality of the band subset cannot be set in advance. To solve this problem, the *k*-DPP model proposed by Kulesza et al [29] is utilized with the cardinality set as *k*. When directly using the *k*-DPP for band selection, however, it can only select a band subset of highest diversity, rather than good performance of data analysis i.e. classification. Besides, *k*-DPP has some random operations which may lead to the selection of different band subsets and instability of following on data classification.

To solve the aforementioned two drawbacks of *k*-DPP, we combine the MIMN criterion with improved *k*-DPP to form our proposed band selection method, MIMN-DPP. The proposed method can not only select the optimal band subset of high diversity and good performance of data analysis but also mitigating the randomness of *k*-DPP for robust and stable data classification. In the following, we will detail the proposed MIMN-DPP band selection method in three parts, i.e. model construction, selection of eigenvector subset and selection of band subset , respectively.

3.4.1 Model construction of MIMN-DPP

The proposed MIMN-DPP model can be represented as:

$$\mathrm{P}^{k}_{L_B, L_{\hat{B}}}(Y) = u\frac{\det(L_B^Y)}{\sum_{|Y'|=k}\det(L_B^{Y'})} + (1-u)\frac{\det(L_{\hat{B}}^Y)}{\sum_{|Y'|=k}\det(L_{\hat{B}}^{Y'})} \tag{15}$$

where Y is the selected band subset with a cardinality of k; $|Y'|$ represents the cardinality of Y'; u is an adjusting parameter to balance the impact of the hyperspectral image information and its neighbour information. L_B and $L_{\hat{B}}$ are adjacency matrices whose entries are indexed by the corresponding bands of

hyperspectral image B and corresponding neighbourhood information $\hat{B}$, respectively. Similarly, L_B^Y and $L_{\hat{B}}^Y$ are submatrices of L_B and $L_{\hat{B}}$, where their entries are indexed by the corresponding bands of Y. In addition, $L_B^{Y'}$ and $L_{\hat{B}}^{Y'}$ are submatrices of L_B and $L_{\hat{B}}$ respectively and their entries are indexed by the corresponding bands of Y'.

As seen in Eq. (15), our model can be regarded as a mixture of two k-DPP models, one is applied to the hyperspectral imaging and the other to its neighborhood. The proposed model is to select a band subset Y with higher diversity both in the hyperspectral image and its neighbourhood information. The higher value of Eq. (15) means that bands in the selected subset Y are of high diversity and low-redundancy. In addition, in combination with the MIMN criterion, bands in the selected subset Y are of low noise in the local area for high smoothness or consistency accordingly.

According to the inference of [27], the denominator of the first term of Eq. (15) can be expressed as

$$\sum_{|Y'|=k} \det(L_B^{Y'}) = \sum_{|S|=k} \prod_{n\in S} \lambda_n^B \tag{16}$$

Similarly, the denominator of the second term of Eq. (15) can be expressed as:

$$\sum_{|Y'|=k} \det(L_{\hat{B}}^{Y'}) = \sum_{|\hat{S}|=k} \prod_{n\in \hat{S}} \lambda_n^{\hat{B}} \tag{17}$$

The eigenvalue polynomial in Eq. (16) is denoted by $e_{B,k}^N$ and $S \subseteq \{1,2,3,\cdots,l\}$ is the index of eigenvalue subset $\{\lambda_n^B\}_{n\in S}$. The cardinality of S is k. Similarly, the eigenvalue polynomial in Eq. (17) is represented as $e_{\hat{B},k}^N$. The cardinality of $\hat{S}$ is k, too.

The determinants $\det(L_B^Y)$ and $\det(L_{\hat{B}}^Y)$ can be written as:

$$\det(L_B^Y) = \det(L+I)\mathrm{P}_{L_B}^k(Y) \tag{18}$$

$$\det(L_{\hat{B}}^Y) = \det(L+I)\mathrm{P}_{L_{\hat{B}}}^k(Y) \tag{19}$$

$\mathrm{P}_{L_B}^{k}(Y)$ is a mixture of elementary DPPs [27] and it can be expressed as $\mathrm{P}_{L_B}^{k} = \frac{1}{\det(L_B + I)} \sum_{S \subseteq \{1,2,\cdots,N\}} \mathrm{P}^{V_S^B} \prod_{n \in S} \lambda_n^B$ where λ_n^B is the eigenvalue of matrix L_B. V_S^B is a set of orthonormal vectors, and $\mathrm{P}^{V_S^B}$ is called elementary DPP.

Substituting Eqs. (16)-(19) into Eq. (15), we have

$$\begin{aligned} \mathrm{P}_{L_B, L_{\hat{B}}}^{k}(Y) = & \frac{u}{e_{B,k}^{N}} \sum_{|S|=k} \mathrm{P}^{V_S^B}(Y) \prod_{n \in S} \lambda_n^B \\ & + \frac{(1-u)}{e_{B,k}^{N}} \sum_{|\hat{S}|=k} \mathrm{P}^{V_{\hat{S}}^{\hat{B}}}(Y) \prod_{n \in \hat{S}} \lambda_n^{\hat{B}} \end{aligned} . \tag{20}$$

Eq. (20) is actually one part of the final expression of our MIMN-DPP model, from which we will select a band subset Y with diversity from perspective of the image information and its neighbour information. The higher value of Eq. (20) means that the selected band subset Y is of high diversity and low-redundancy. The other part of our MIMN-DPP model is to apply the proposed MIMN criterion to the sampling process which will be introduced in Section 3.4.3.

In the following, we will introduce an improved sampling method to tackle the problem of random operations in k-DPP for improving the robustness/stability of data classification, referring to reference [29]. Similar to the k-DPP, our method also contains two main modules: the first is to select the eigenvector subset according to its eigenvalues, and the second is to sample the band indices from the eigenvector subset. However, some improved adjustments are adopted for our algorithm. Detailed descriptions of these two main modules are given subsections 3.4.2 and 3.4.3 below.

3.4.2 Selection of eigenvector subset

Unlike the original k-DPP [29], firstly, our proposed MIMN-DPP method needs to sample band subset Y from both matrices L_B and $L_{\hat{B}}$ instead of a single matrix. Therefore, as shown in Step 2 of Algorithm #1, both matrices L_B and $L_{\hat{B}}$ are decomposed to obtain the corresponding eigenvectors and eigenvalues. Two subsets of eigenvectors, S and $\hat{S}$, will be selected according to their corresponding eigenvalues

$\{\lambda_n^B\}$ and $\{\lambda_n^{\hat{B}}\}$, $n \in \{1,2,3,\cdots,l\}$. For selecting band subset, the information of two aspects S and $\hat{S}$ are combined to determine the selected bands.

Denote S as the subset of eigenvector indexes where $|S| = k$. Assume that $S' \cup n = S$, so $|S'| = k-1$. The probability of selecting the k-th eigenvector index n_B to add to S is given by [29]:

$$\mathrm{P}(n \in S) = \frac{\lambda_n^B}{e_{B,k}^n} \sum_{|S'|=k-1} \prod_{n' \in S'} \lambda_{n'}^B = \lambda_n^B \frac{e_{B,k-1}^{n-1}}{e_{B,k}^n} \tag{21}$$

Similarly, the probability of selecting the k-th eigenvector index n of the adjacent matrix $L_{\hat{B}}$ to add to $\hat{S}$ is given by:

$$\mathrm{P}(n \in \hat{S}) = \lambda_n^{\hat{B}} \frac{e_{\hat{B},k-1}^{n-1}}{e_{\hat{B},k}^n} \tag{22}$$

Details of implementation are summarized in Algorithm 1. After some iterations, we can choose two eigenvector index subsets S and $\hat{S}$ with cardinality k. The eigenvector subsets of L_B and $L_{\hat{B}}$ can be represented as $\{v_n^B\}_{n \in S}$ and $\{v_n^{\hat{B}}\}_{n \in \hat{S}}$ respectively.

3.4.3 Selection of band subset

In Algorithm #2, a band subset Y of low-redundancy and high-diversity is selected based on eigenvector subsets $\{v_n^B\}_{n \in S}$ and $\{v_n^{\hat{B}}\}_{n \in \hat{S}}$. In the meantime, to select bands with high performance in data analysis, the MIMN criterion is applied in the process as detailed below.

Let the band subset Y' with cardinality $k-1$ satisfy $Y' \cup i = Y$. The probability of selecting band i as the k-th band to add to Y can be defined as:

$$\begin{aligned} \mathrm{P}_{B,\hat{B}}(i) = u \frac{1}{\left|V_S^B\right|} \sum_{v_n^B \in V_S^B} \left(v_n^{B^T} e_i\right)^2 \\ +(1-u) \frac{1}{\left|V_S^{\hat{B}}\right|} \sum_{v_n^{\hat{B}} \in V_S^{\hat{B}}} \left(v_n^{\hat{B}^T} e_i\right)^2 \end{aligned} \tag{23}$$

where e_i is a column vector whose i-th element is one and all others are zero. The higher the probability $\mathrm{P}_{B,\hat{B}}(i)$ is, more likely the band i is selected for adding to Y.

Algorithm 1: Sampling eigenvector subset S and $\hat{S}$

Input:

B : hyperspectral image

$\hat{B}$: neighbour information

k : number of selected eigenvectors

Output:

S , $\hat{S}$: index of selected eigenvectors

Step 1: Construct adjacency matrix L_B with Eq. (8)

Construct adjacency matrix $L_{\hat{B}}$ with Eq. (10)

Step 2: Eigen decomposition:

$$L_B \to \{\lambda_n^B, v_n^B\}_{n=1}^l$$

$$L_{\hat{B}} \to \{\lambda_n^{\hat{B}}, v_n^{\hat{B}}\}_{n=1}^l$$

Step 3: $S \leftarrow \varnothing, \hat{S} \leftarrow \varnothing$, $k_B \leftarrow k$, $k_{\hat{B}} \leftarrow k$

Step 4: for $n = l, \cdots, 2, 1$ do

if $k_B = 0$ then

break

end if

if $u \sim U[0,1] < \lambda_n^B \dfrac{e_{B,k-1}^{n-1}}{e_{B,k}^n}$ then

$S \leftarrow S \cup \{n\}, k_B \leftarrow k_B - 1$

end if

end for

Step 5: for $n = l, \cdots, 2, 1$ do

if $k_{\hat{B}} = 0$ then

break

end if

if $u \sim U[0,1] < \lambda_n^{\hat{B}} \dfrac{e_{\hat{B},k-1}^{n-1}}{e_{\hat{B},k}^n}$ then

$\hat{S} \leftarrow \hat{S} \cup \{n\}, k_{\hat{B}} \leftarrow k_{\hat{B}} - 1$

end if

end for

Output: S , $\hat{S}$

To preserve the diversity of band subset, minor adjustments are adopted according to [29] rather than selecting the band directly based on Eq. (23). This is clearly defined in Step 4 of Algorithm #2.

In addition, when integrating the MIMN search criteria into the sampling process, the Step 4 of Algorithm #2 can sample the band that differs greatly from the selected band subset to maximize the diversity. We repeat this step α times to find α candidate bands to construct a temporary subset Y_{temp} . From Y_{temp} , we select the band that meets the maximum search criteria φ_m as defined in Eq. (14). This is to balance between diversity and performance of data analysis. Too large value of α will cause those bands with large φ_m to be always selected, resulting in a lack of diversity. When the value of α is too

large, it will lead to the band selection biased to those with large φ_m, resulting in poor diversity of selected bands. Meanwhile, the larger the number of iterations α is, the more bands the subset Y_{temp} will have. In each iteration, bands with a larger φ_m is more likely selected from the subset Y_{temp}. Therefore, the band to be selected will gradually focus on the one with a large φ_m. On the contrary, a too small α will result in a lack of high performance of data analysis. If the number of iterations α is too small, the number of bands in the subset Y_{temp} will be small, hence the band with a large φ_m may be excluded and the high-performance band cannot be selected. We will discuss the impact of the parameter α in the following experiment Section 4.3.

Algorithm 2: Select band subset

Input:

$S, \hat{S}$: index of selected eigenvectors

α : number of iterations

$\{v_n^B\}_{n=1}^l, \{v_n^{\hat{B}}\}_{n=1}^l$:eigenvectors of L_B and $L_{\hat{B}}$

Output:

Y : index of selected bands

Step 1: Calculate $\{\varphi_m\}_{m=1}^l$ of all bands with equation (14)

Step 2: $V_S^B \leftarrow \{v_n^B\}_{n\in S}, V_{\hat{S}}^{\hat{B}} \leftarrow \{v_n^{\hat{B}}\}_{n\in\hat{S}}$

Step 3: $Y \leftarrow \varnothing$

Step 4: While $\left|V_S^B\right| > 0, \left|V_{\hat{S}}^{\hat{B}}\right| > 0$ do

Calculate the probability $\{\mathrm{P}_{B,\hat{B}}(i)\}_i^l$ of bands with equation

if $\alpha > 0$ then

if $u \sim U[0,1] < \sum_{i=1}^{g} \mathrm{P}_{B,\hat{B}}(i)$ then

$Y_{temp} \leftarrow Y_{temp} \cup \{g\}$

end if

$\alpha \leftarrow \alpha - 1$

end if

$Y \leftarrow Y \cup \max_i(\varphi(Y_{temp}))$

$V_S^B \leftarrow V_{S\perp}^B, V_{\hat{S}}^{\hat{B}} \leftarrow V_{\hat{S}\perp}^{\hat{B}}$ orthonormal basis for subspace of $V_S^B, V_{\hat{S}}^{\hat{B}}$ orthogonal to e_i

end while

Output: Y

4. Experimental Results and Discussions

In this section, three well-known hyperspectral image datasets are used to evaluate the performance of the proposed algorithm. Eight state-of-the-art approaches including k-DPP [29], MVPCA [31], MIC [32],

Lscore [33], WaLuMI [40], MIMR-CSA [5], DSEBS [49] and SOP-SRL [48] are taken for benchmarking. We construct a 5 nearest neighbour graph for Lscore and using Cosine to calculate weights for each edge in the graph. As suggested in [5], the number of antibodies and cloned antibodies are set as 5 and 2500 respectively in MIMR-CSA. Descriptions of the datasets and experimental results are discussed as follows.

4.1. Hyperspectral Datasets

The three HSI datasets used in our experiments include Indian Pines, Pavia University dataset, and Salinas. The Indian Pines dataset was obtained by the Airborne Visible/Infrared Imaging Spectrometer sensor (AVIRIS) in 1992 [36]. The image covers Indian Pines test site in North-western Indiana. It has 220 spectral bands within the spectrum range of 0.2-2.4um. After removing 20 water absorption spectral bands [104-108] and [150-163], the remaining 200 spectral bands are used for classification. It has 145×145 spatial pixels, including 10366 pixels labelled in 16 classes as the 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	2	52	54
2. Corn-notill	71	1363	1434
3. Corn- mintill	42	792	834
4. Corn	12	222	234
5. Grass- pasture	24	473	497
6. Grass-trees	37	710	747
7. Grass-pas- turemowed	2	24	26
8. Hay- windrowed	24	465	489
9. Oats	2	18	20
10. Soybean- notill	49	919	968
11. Soybean- mintill	122	2346	2468
12. Soybean- clean	30	584	614
13. Wheat	10	202	212
14. Woods	63	1231	1294
15. Building- grass-trees	19	361	380
16. Stone- steel-towers	5	90	95
total	514	9852	10366

The Pavia University dataset was acquired using the Reflective Optics System Imaging Spectrometer optical sensor covering the Pavia University, Italy [37]. It has 115 spectral bands within the spectrum

range of 0.43-0.86 um. After removing 12 noisy spectral bands in total 103 bands are remained in the dataset. The spatial dimension is 610×340 pixels, in which 42776 pixels are labelled in 9 classes as the ground truth.

Table 2: The numbers of training and test samples for each category of the Pavia university image

Class	Training	Test	Samples
1. Asphalt	66	6565	6631
2. Meadows	186	18463	18649
3. Gravel	21	2078	2099
4. Trees	31	3033	3064
5. Painted metal sheets	13	1332	1345
6. Bare soil	50	4979	5029
7. Bitumen	13	1317	1330
8. Self-blocking bricks	37	3645	3682
9. Shadows	10	937	947
Total	427	42349	42776

The Salinas dataset was also acquired by the AVIRIS sensor over Salinas Valley, California [37]. After eliminating 20 water absorption bands, 200 spectral bands are remained in the dataset. It contains 512×214 spatial pixels, in which 54129 pixels are labelled in 16 classes as the ground truth.

Table 3: The numbers of training and test samples for each category of the Salinas image

Class	Training	Test	Samples
1.Brocoli_green_weeds_1	20	1989	2009
2.Brocoli_green_weeds_2	37	3689	3726
3.Fallow	19	1957	1976
4.Fallow_rough_plow	14	1380	1394
5.Fallow_smooth	26	2652	2678
6.Stubble	39	3920	3959
7.Celery	35	3544	3579
8.Grapes_untrained	112	11159	11271
9.Soil_vinyard_develop	62	6141	6203
10.Corn_senesced_green	32	3246	3278
11.Lettuce_romaine_4wk	10	1058	1068
12.Lettuce_romaine_5w	19	1908	1927
13.Lettuce_romaine_6wk	9	907	916
14.Lettuce_romaine_7wk	10	1060	1070
15.Vinyard_untrained	72	7196	7268
16.Vinyard_vertical	18	1789	1807
Total	534	53595	54129

4.2. Experimental parameter settings

In this section, we will introduce the parameter settings. The weight parameter u is set to be 0.5 empirically. The size of the neighbour domain, d, is set to 3 for Indian Pines dataset, and 5 for both Pavia University and Salinas datasets. The number of iterations α will be discussed in Section 4.3. The number of selected bands varies from 10 to 100 with an interval of 10. The support vector machine (SVM) classifier with the radial-basis-function (RBF) kernel [34] is used for data classification to verify the performance of the selected band subsets. For SVM, the parameter C and γ are determined by how-many-folds cross validation. In Indian Pines dataset, about 5% samples are randomly selected for training and the remaining for testing. In Pavia University and Salinas datasets, about 1% samples are randomly selected for training and the rest for validation.

The number of training samples and test samples for each labelled class of the three datasets are shown in Tables 1-3, respectively. All experiments are repeated 10 times with the average classification results used for comparison. The results are quantitatively evaluated in terms of three metrics, i.e. overall accuracy (OA), average accuracy (AA) and kappa coefficient [35].

4.3 Analysis of the number of iterations α

The number of iterations, α, is an important parameter in the proposed MIMN-DPP method as it balances the diversity and classification performance of the selected band subset. For credible evaluation, we repeat 20 times for each α in band selection, where each of the selected band subsets is tested 10 times to obtain the average classification accuracy for comparison. The number of selected bands is set to 20 for the Indian Pines and 10 for both the Pavia University and Salinas datasets. Fig. 4 shows the obtained average classification accuracy with the parameter α varying from 2 to 20. When α is set to be 1, the criterion MIMN becomes void hence the model degrades to a double-DPP model. As seen from Fig. 4, for the Indian Pines dataset the optimal α is 3, whilst for both the Pavia University and Salinas datasets this becomes 5. Under this optimal α, the proposed MIMN-DPP is able to select the band subsets with high diversity and high accuracy of classification.

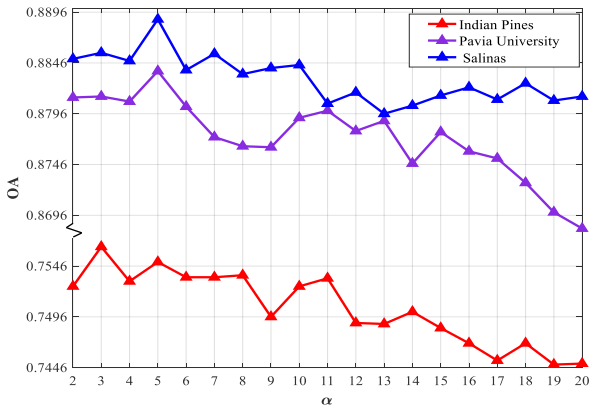
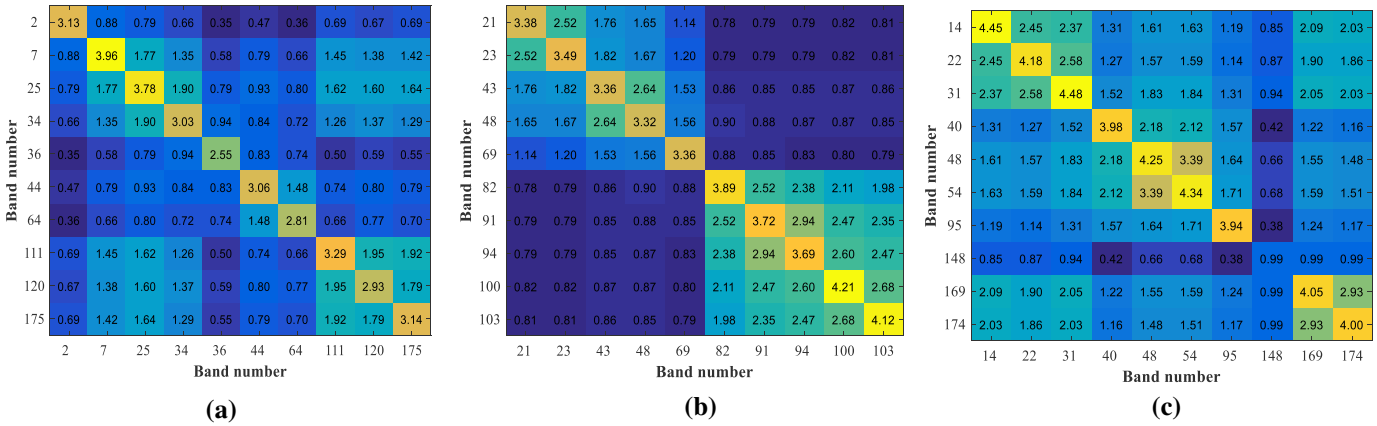


Figure 4: The parameter α analysis of MIMN-DPP on the datasets of Indian Pines, Pavia University and Salinas image.

4.4. Analysis of the selected band subset

In this section, ten bands on each dataset selected using the proposed MIMN-DPP algorithm are analyzed in detail. For the Indian Pines dataset, the indexes of selected bands are 2, 7, 25, 34, 36, 44, 64, 111, 120 and 175. For Pavia University, these are 21, 23, 43, 48, 69, 82, 91, 94, 100 and 103. For Salinas, these become 14, 22, 31, 40, 48, 54, 95, 148, 169 and 174.

Selecting a band subset with high diversity and low redundancy is the most important characteristic of our MIMN-DPP algorithm. We construct the similarity matrices based on minimum mutual information to ensure the diversity of selected band subsets as shown in Fig. 5. In HSI, the adjacent bands are often highly correlated and of high redundancy. Therefore, selected bands must be of low redundancy to provide supplementary information from each other.



(a) (b) (c)

Figure 5: Similarity matrix of selected bands with mutual information on (a) Indian Pines image (b) Pavia University image and (c) Salinas image.

As seen in Fig. 5, the selected bands have good coverage in terms of wide spectrum range and low similarity, thus high diversity information can be maintained from the selected bands. Besides, low mutual information with others means that the selected bands are less correlated to each other hence the low-redundancy among the selected bands.

According to our MIMN criterion, we calculate the MIMN value for each band on the three datasets and plot the results in Fig. 6. As seen, most of the selected bands have the highest MIMN value in the neighbouring bands. To this end, they carry more information and relatively lower noise. Bands indexed in ranges of 100-108,140-160 and180-200 in Fig. 6(a) seem to have lower MIMN values, and this indicates why they are not selected. For the Pavia University dataset, the band-based MIMN value is shown in Fig. 6(b). The bands with higher MIMN values are mainly in the ranges of 80-103, where half of bands are selected. On the other hand, bands 1-10 are seldom selected due to their lower MIMN values. For the Salinas in Fig. 6(c), the bands with higher MIMN values dominant mainly in the range of 16-76 and 160-180 where most selected bands fall within. Although a small number of bands with relatively low MIMN value are selected in Fig. 6, this is due to that our band selection considers both diversity and high-performance of the selected bands.

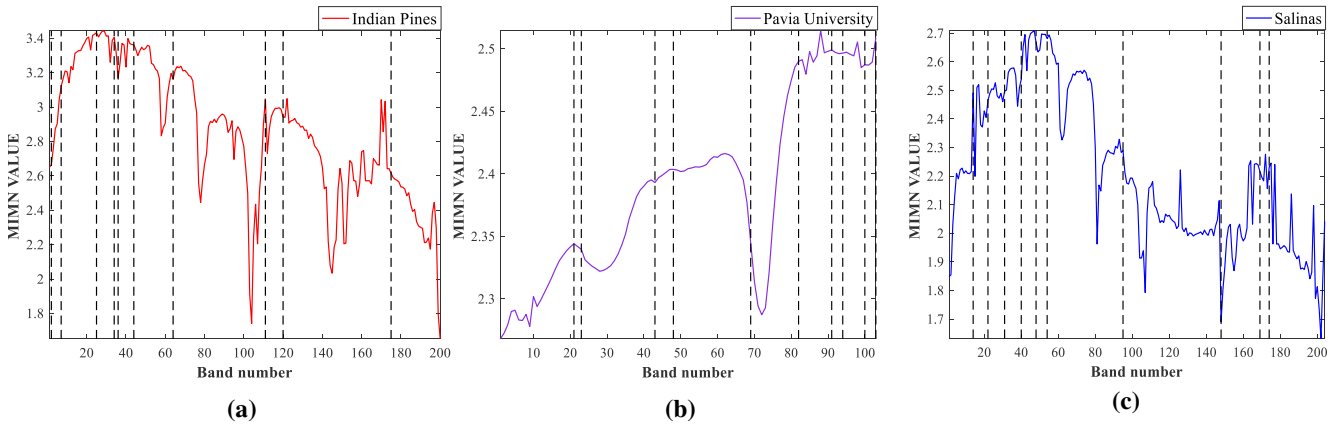


Figure 6: MIMN value of the selected subset of ten bands on (a) Indian Pines image (b) Pavia University image and (c) Salinas image.

In Fig. 7, we plot the average spectral profiles for each labelled class for the three datasets, along with the selected bands for analysis. According to [5], spectrums with large interval among classes mean that

they have highly discriminative ability for these classes. As shown in Fig. 7(a), the bands in 1-50, 60-70 and 110-130 for Indian Pines seem to have better discriminative ability as the intensity is large in these spectral ranges. As seen, most of selected bands cover these regions, indicating the good performance of data classification as well. Similar observations can also be found in Fig. 7(b) and Fig. 7(c) for the other two datasets.

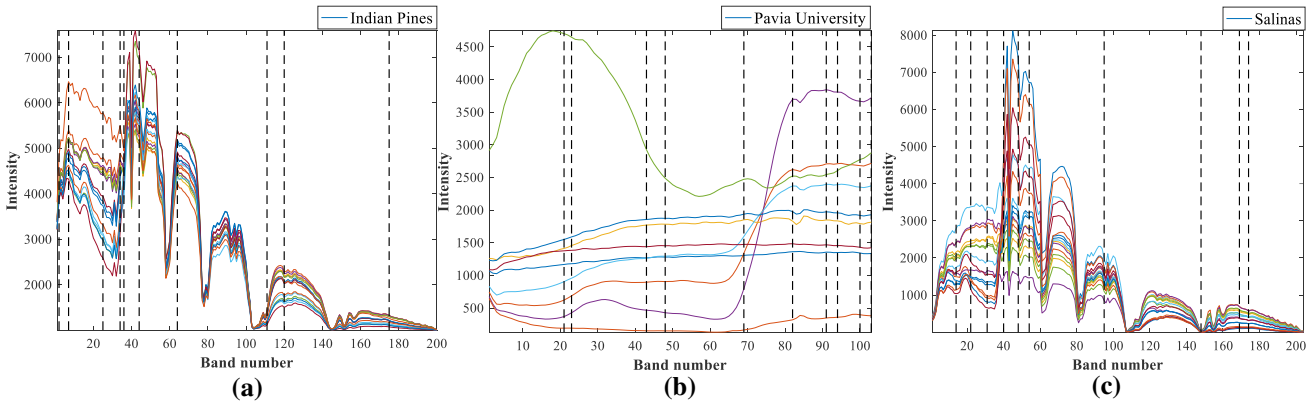


Figure 7: Spectral profiles of different land-cover classes for the three datasets of (a) Indian Pines, 16 classes, (b) Pavia University, 9 classes and (c) Salinas, 16 classes.

4.5. Computation time analysis and comparison

Let N, k and n denote respectively the number of total bands, the number of selected bands, and the number of similarity kernel matrices. In Algorithm #1, the computation complexity for Eigen-decomposition of n similarity kernel matrices is $O(nN^3)$. As suggested in [27], the computation complexity for the remaining part of Algorithm #1 is $O(nNk^2)$. In addition, Algorithm #2 also needs $O(\alpha Nk^3)$, where α is the number of iterations. In total, the computational complexity for the search part of our algorithm becomes $O(nN^3)+O(nNk^2)+O(\alpha Nk^3)$, where both n and α are scalars and their values are quite small (i.e. N is set to 2, and α is less than 5 in our paper). Therefore, the overall computation complexity can be simplified to $O(N^3)+O(Nk^3)$. As N is a fixed scalar indicating the total number of bands, only the parameter k may affect the search time for a given dataset.

For large k, the computational complexity of the algorithm will grow rapidly. However, in the band selection task, these number are generally small, especially for k. For the three datasets, the table of

MIMN evaluation criteria can be pre-established. To speed up the process, the similarity matrices with mutual information can also be pre-calculated and stored in lookup tables.

All our experiments were conducted on Matlab 2016b using a computer with Inter CPU E5-1620@3.5GHz and 32GB RAM. It should be noted that The MIMN-DPP algorithm obtains the value of φ_m from the pre-established lookup table. Therefore, when calculating the running time of the MIMN-DPP, the time for establishing the lookup table is not added.

In Fig. 8, the running times of MIMN-DPP and other algorithms on Indian Pines are shown with the number of selected spectral bands increasing from 10 to 150. As seen from Fig. 8, the computation complexity of our algorithm is only affected by the parameter k, which is the number of selected bands. The running time for MIMN-DPP and k-DPP is among the least, which increases slightly when selecting more bands. For MVPCA, MIC, Lscore, WaLuMI, SOP-SRL and DSEBS, however, the running time does not change significantly with the increasing number of selected bands. The running time of our algorithm is quite low when the parameter k is less than 70. Although it increases with the increasing number of selected bands k, it remains low even when $k>100$. Since the value of k in practical applications is far below 100, our algorithm is advantageous in time with a small k. MIMN-DPP takes more time than k-DPP due to the extra time needed to select band subset with MIMN criteria.

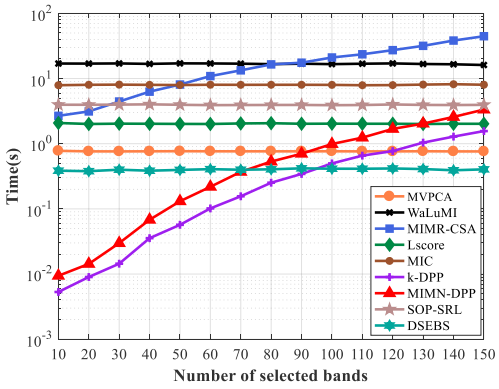


Figure 8: Running time comparisons of MIMN-DPP and other methods on Indian Pines.

4.6. Classification performance

In this subsection, the classification accuracy is used to evaluate the performance of the selected band subsets on Indian Pines, Pavia University and Salinas datasets. Due to the fact that SOP-SRL, k-DPP and

MIMN-DPP are random algorithm, to obtain credible result, we repeat 10 runs of band selection algorithm, where the number of selected bands increases from 10 to 100 on three datasets. For each selected band subset, the classification of HSI using SVM is implemented 10 times [34], where the average classification accuracy in terms of OA is obtained and compared in Fig. 9.

As seen in Fig. 9, for all the three datasets, MIMN-DPP have outperformed all other state-of-the-art approaches, which have shown the superiority of the proposed algorithm. For k-DPP, it also produces quite good results in all the three datasets, although not the best, this validates again the advantage of the diversity criterion in band selection. Although MVPCA produces quite good results in Pavia University and Salinas just after MIMN-DPP, the result on Indiana Pines is quite poor, indicating the instability of the approach. WaLuMI and MIMR-CSA seem to supplement each other to some extent, where good results from one come with relatively poor results from another. MIC and Lscore are among the worst in the group, though Lscore may have very good results even beat all others when the number of selected bands is extremely, i.e. over 80. Compared to the more recent methods, such as SOP-SRL and DSEBS, our method is more effective in low-dimensional situations. When the selected band exceeds 60, our method still surpasses other methods in most cases, although the improvement of performance becomes less significant. This is because, when the number of bands leeps increasing, the accuracy of the classification is close to saturation.

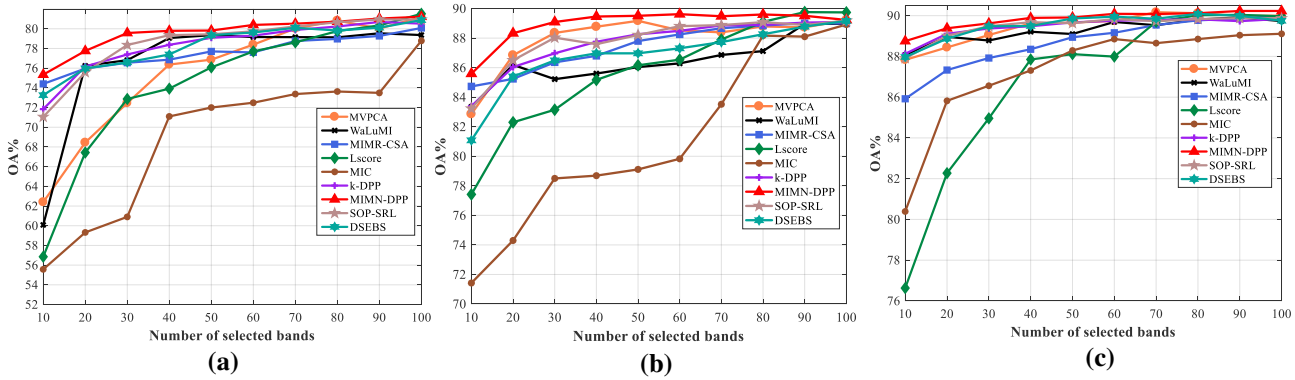


Figure 9: Classification accuracy comparisons of MIMN-DPP and k-DPP on three datasets including (**a**) Indian Pines, (**b**) Pavia University and (**c**) Salinas.

In Tables 4-6, MIMN-DPP is compared with the original k-DPP and the DPP-double model without the MIMN criterion. This is to validate how different key components and improvements can affect the classification accuracy on the three datasets. The number of selected bands is set to 20 for Indian Pines and 10 for both Pavia University and Salinas datasets.

As seen in Table 4, compared with k-DPP, DPP-double improves 0.65%, 1.18% and 0.64% in terms of OA, AA and Kappa, respectively. Meanwhile, the standard deviations for the three metrics all become lower, indicating a more robust solution from DPP-double than k-DPP. With the integration of the MIMN criterion, further improvements are achieved in terms of further increased OA, AA and Kappa of 1.58%, 1.84% and 1.82% respectively as well as reduced standard deviations. This shows again the superiority of the MIMN-DPP approach for producing the robust solution in band selection.

Table 4: Results compared on the Indian Pines dataset

method	k-DPP	DPP-double	MIMN-DPP
OA(%)	74.17±1.51	74.72±1.10	**75.75±0.95**
AA(%)	68.91±1.76	70.07±1.52	**70.75±1.47**
Kappa(%)	70.38±1.77	71.02±1.28	**72.20±1.10**

Table 5: Results compared on Pavia University dataset

method	k-DPP	DPP-double	MIMN-DPP
OA(%)	83.99±1.40	84.76±1.26	**85.59±0.97**
AA(%)	79.30±2.12	80.57±1.52	**81.00±1.09**
Kappa(%)	78.41±1.91	79.46±1.74	**80.59±1.34**

Table 6: Results compared on the Salinas dataset

method	k-DPP	DPP-double	MIMN-DPP
OA(%)	88.13±0.91	88.29±0.77	**88.75±0.40**
AA(%)	91.81±0.62	92.06±0.50	**92.34±0.26**
Kappa(%)	86.74±1.02	86.91±0.86	**87.43±0.45**

Similar observations can also be made from Table 5 and Table 6, which validates again the robustness of the proposed MIMN-DPP approach in another two datasets, i.e. Pavia University and Salinas. To intuitively show the advantages of MIMN-DPP in data classification and stableness, box-plots of the results are illustrated in Fig. 10 in comparison to k-DPP and MIMN-DPP. As seen, the median classification accuracy of MIMN-DPP is much higher than that of k-DPP which indicates MIMN-DPP is

generally better than k-DPP. In addition, MIMN-DPP has a smaller data fluctuation range, meaning that the result is more stable.

MIMN-DPP is superior to k-DPP for two main reasons: One is the neighbourhood information used to construct adjacency matrix as a measure of band redundancy, which makes redundancy measure more accurate. The other is that the MIMN criterion integrated in the sampling process, which can avoid selecting bands with low information and high noise. What's more, it can also effectively mitigate the randomness of the original k-DPP.

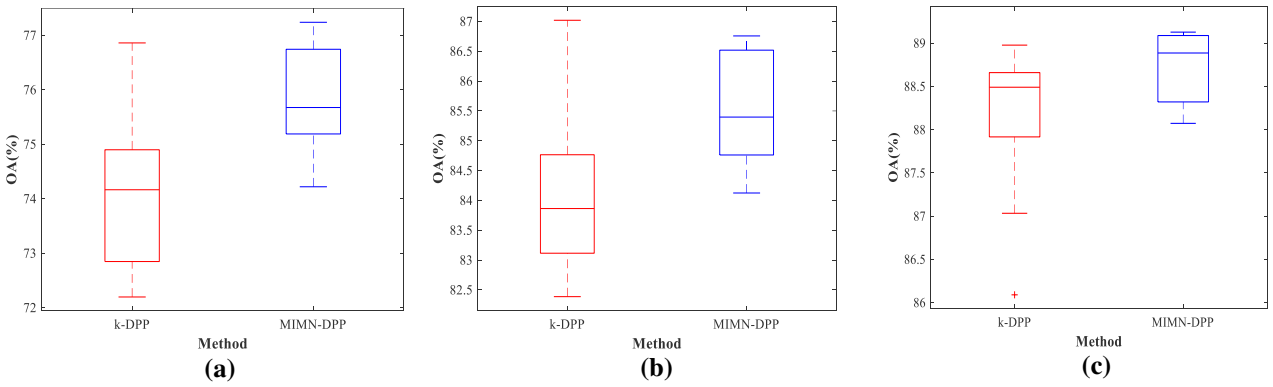


Figure 10: Comparing MIMN-DPP and k-DPP with OA on three datasets of (**a**) Indian Pines, (**b**) Pavia University and (**c**) Salinas.

To further demonstrate the advantages of MIMN-DPP in selecting low-dimension band subset, several band subsets selected by different methods are tested, as shown in Tables 7-9. Since SOP-SRL, k-DPP and MIMN-DPP are randomness method, we independently select ten band subsets and record a subset whose classification accuracy OA is closest to the mean. As shown in Tables 7-9, among all algorithms, Lscore and MIC obtain the worst results. MIMN-DPP and k-DPP generate better results as the selected band subsets are dispersed evenly and contain diverse information in different band wavelengths. Compared to other methods, the band subset selected by MIMN-DPP can obtain better classification in most regions of different classes.

In summary, MIMN-DPP can be effectively applied to band selection for the reason that it is able to select a band subset of high diversity, low redundancy, and good performance in data classification.

Table 7: Classification accuracy of the band set on the Indian Pines image selected by MVPCA, WaLuMI, MIMR-CSA, Lscore, MIC, DSEBS, SOP-SRL, k-DPP and the Proposed algorithm

Class	MVPCA [31]	WaLuMI [40]	MIMR-CSA [5]	Lscore [33]	MIC [32]	DSEBS [49]	SOP-SRL [48]	k-DPP [29]	Proposed
1.	26.15 ±19.34	33.46 ±21.70	40.38 ±15.44	26.54 ±20.83	4.62 ±7.60	**45.19 ±21.93**	25.77 ±19.51	34.62 ±19.92	36.54 ±22.06
2.	44.87 ±8.46	67.45 ±4.71	68.03 ±3.62	53.97 ±4.92	34.13 ±4.78	67.29 ±3.25	**74.97 ±2.93**	69.51 ±4.18	74.01 ±1.87
3.	38.60 ±4.87	53.33 ±5.84	56.67 ±7.34	31.72 ±6.28	44.52 ±4.16	57.88 ±5.31	47.68 ±6.71	55.28 ±5.10	**58.62 ±5.15**
4.	35.23 ±8.07	43.92 ±12.90	49.73 ±11.46	41.49 ±8.26	10.14 ±5.96	46.31 ±12.48	52.30 ±9.55	40.36 ±10.54	**54.14 ±13.74**
5.	87.99 ±3.90	84.19 ±3.59	**90.02 ±2.46**	78.73 ±5.37	31.06 ±16.11	85.05 ±3.04	85.64 ±3.25	82.52 ±5.74	83.32 ±4.01
6.	87.28 ±2.87	88.31 ±2.63	**92.77 ±2.81**	90.41 ±3.31	82.32 ±4.25	87.10 ±2.89	89.27 ±1.68	88.55 ±1.94	90.65 ±3.61
7.	68.75 ±24.00	**84.58 ±8.11**	75.00 ±15.47	41.67 ±22.05	2.92 ±5.57	75.42 ±18.16	73.75 ±15.35	77.92 ±19.15	74.17 ±12.08
8.	96.62 ±3.18	96.45 ±2.15	**97.08 ±1.70**	92.17 ±4.39	95.96 ±3.18	93.85 ±2.99	96.11 ±2.22	96.60 ±2.10	96.17 ±2.05
9.	18.89 ±13.91	48.89 ±23.25	61.67 ±27.27	21.67 ±16.45	7.78 ±11.17	48.89 ±13.81	48.33 ±13.62	52.22 ±19.98	**65.00 ±20.63**
10.	49.55 ±11.68	**67.12 ±4.14**	55.13 ±7.74	55.14 ±6.65	37.26 ±6.66	63.00 ±3.61	62.72 ±3.19	62.73 ±5.35	60.96 ±6.42
11.	**80.46 ±3.41**	79.13 ±3.99	77.81 ±2.91	73.37 ±4.39	70.75 ±4.40	77.58 ±3.68	78.62 ±2.87	78.37 ±4.04	78.96 ±2.90
12.	38.07 ±7.21	57.35 ±10.40	53.42 ±5.44	27.00 ±6.08	42.38 ±4.68	**63.12 ±6.38**	37.65 ±16.12	56.11 ±6.54	60.00 ±7.02
13.	94.90 ±4.19	95.74 ±2.38	96.88 ±1.48	90.15 ±4.51	67.77 ±15.93	96.73 ±1.93	95.50 ±3.51	**98.02 ±1.71**	96.83 ±1.78
14.	**96.01 ±1.54**	94.11 ±2.42	95.02 ±1.64	93.95 ±1.82	92.47 ±3.90	93.40 ±2.16	94.82 ±1.89	93.24 ±2.75	93.35 ±2.46
15.	17.67 ±6.18	33.88 ±5.77	37.12 ±4.24	28.25 ±5.29	14.93 ±6.31	32.11 ±6.50	38.48 ±7.36	37.31 ±5.12	**41.50 ±7.10**
16.	76.67 ±4.38	80.44 ±7.93	79.11 ±7.84	78.33 ±10.17	51.00 ±20.65	81.11 ±4.48	**83.33 ±5.88**	82.89 ±2.68	80.56 ±7.17
OA (%)	66.85 ±1.05	74.24 ±0.86	73.96 ±0.76	65.43 ±1.05	57.32 ±0.96	73.97 ±0.91	73.65 ±1.27	73.96 ±0.68	**75.69 ±1.05**
AA (%)	59.86 ±2.72	69.27 ±2.20	70.36 ±2.63	57.78 ±2.03	43.12 ±3.44	69.63 ±2.66	67.81 ±2.86	69.14 ±2.22	**71.55 ±1.91**
Kappa (%)	61.54 ±1.37	70.43 ±1.00	70.15 ±0.92	60.21 ±1.14	50.44 ±1.25	70.17 ±0.98	69.72 ±1.55	70.13 ±0.86	**72.15 ±1.21**

Table 8: Classification accuracy of the band set on the Pavia university image selected by MVPCA, WaLuMI, MIMR-CSA, Lscore, MIC, DSEBS, SOP-SRL, k-DPP and the Proposed algorithm

Class	MVPCA [31]	WaLuMI [40]	MIMR-CSA [5]	Lscore [33]	MIC [32]	DSEBS [49]	SOP-SRL [48]	k-DPP [29]	Proposed
1.	**88.51 ±4.31**	85.13 ±1.95	87.28 ±5.39	86.05 ±4.22	74.07 ±3.36	84.68 ±2.45	85.19 ±2.34	83.99 ±1.72	85.81 ±3.25
2.	95.27 ±1.15	95.19 ±1.61	95.27 ±1.62	**95.52 ±2.73**	94.79 ±2.09	93.82 ±2.56	95.06 ±2.41	93.96 ±1.17	94.56 ±2.53
3.	41.15 ±12.73	55.24 ±13.79	49.38 ±12.45	23.93 ±11.75	16.17 ±8.56	48.45 ±9.33	54.67 ±11.77	56.18 ±8.18	**64.56 ±7.48**
4.	79.63 ±4.00	**82.61 ±3.65**	80.16 ±6.68	63.70 ±8.33	57.80 ±6.10	74.86 ±4.91	80.56 ±5.29	80.76 ±4.03	78.85 ±3.87
5.	98.00 ±1.07	**99.15 ±0.19**	97.86 ±0.77	98.30 ±0.91	97.66 ±0.79	97.64 ±3.97	97.43 ±4.39	98.85 ±0.35	99.03 ±0.23

6.	59.40 ±8.08	49.86 ±8.16	**84.80 ±6.11**	27.44 ±9.34	14.03 ±1.95	43.67 ±9.97	46.53 ±4.39	55.25 ±7.38	67.72 ±3.82
7.	23.08 ±24.82	**65.96 ±11.50**	27.05 ±25.33	35.70 ±30.34	8.22 ±12.45	70.06 ±4.90	69.94 ±10.17	67.51 ±10.86	61.21 ±23.33
8.	80.03 ±8.88	78.71 ±5.04	77.63 ±6.15	**81.98 ±8.56**	79.94 ±7.86	78.12 ±5.14	79.30 ±5.03	80.00 ±5.89	80.65 ±5.21
9.	99.04 ±0.05	99.81 ±0.14	99.86 ±0.22	99.87 ±0.08	82.15 ±7.35	99.56 ±0.31	99.49 ±0.27	**99.88 ±0.09**	**99.88 ±0.11**
OA (%)	82.86 ±0.91	83.34 ±0.68	84.84 ±0.54	77.42 ±0.74	71.42 ±0.67	81.08 ±1.05	82.84 ±0.87	83.32 ±0.87	**85.47 ±0.99**
AA (%)	73.89 ±3.28	79.08 ±1.93	76.59 ±2.76	68.05 ±4.23	58.31 ±1.83	76.76 ±1.79	78.69 ±1.73	79.60 ±1.56	**81.36 ±2.35**
Kappa (%)	76.78 ±1.37	77.43 ±0.99	79.62 ±0.79	68.73 ±1.39	59.88 ±0.77	74.31 ±1.47	76.73 ±1.05	77.53 ±1.13	**80.50 ±1.24**

Table 9: Classification accuracy of the band set on the Salinas image selected by MVPCA, WaLuMI, MIMR-CSA, Lscore, MIC, DSEBS, SOP-SRL, k-DPP and the Proposed algorithm

Class	**MVPCA [31]**	**WaLuMI [40]**	**MIMR-CSA [5]**	**Lscore [33]**	**MIC [32]**	**DSEBS [49]**	**SOP-SRL [48]**	**k-DPP [29]**	**Proposed**
1.	95.87 ±2.18	**97.78 ±1.33**	97.33 ±1.03	93.12 ±4.04	94.20 ±2.35	94.09 ±5.06	97.22 ±1.34	95.48 ±3.95	96.23 ±1.83
2.	96.47 ±1.50	**99.51 ±0.36**	98.34 ±1.13	92.32 ±1.82	98.65 ±0.70	96.99 ±2.46	98.92 ±0.87	95.98 ±2.05	98.47 ±1.55
3.	**96.43 ±2.61**	93.79 ±3.43	90.82 ±4.29	77.18 ±8.15	63.06 ±7.28	90.31 ±5.34	92.03 ±5.79	96.23 ±3.40	95.99 ±2.32
4.	**99.03 ±1.11**	96.72 ±4.98	98.68 ±1.32	98.45 ±1.57	96.72 ±2.89	98.42 ±1.58	98.51 ±1.50	98.13 ±1.43	98.85 ±1.66
5.	95.51 ±2.98	95.94 ±1.48	92.88 ±3.70	91.60 ±3.83	78.64 ±4.81	96.95 ±1.17	95.42 ±1.85	**97.53 ±0.66**	96.34 ±1.41
6.	97.85 ±1.05	99.29 ±0.60	98.98 ±1.14	96.66 ±2.04	97.56 ±1.24	99.10 ±0.54	**99.41 ±0.25**	98.13 ±0.98	99.13 ±0.53
7.	98.97 ±0.37	**99.31 ±0.21**	99.00 ±0.27	94.41 ±1.72	98.06 ±1.43	98.81 ±1.11	99.14 ±0.27	98.55 ±1.11	99.01 ±0.72
8.	**86.64 ±2.42**	82.71 ±4.01	79.99 ±4.66	80.89 ±6.55	79.39 ±5.96	83.35 ±3.37	82.74 ±5.04	85.34 ±4.65	84.42 ±2.75
9.	97.89 ±0.82	97.66 ±0.99	96.28 ±1.59	94.44 ±2.02	94.57 ±4.76	**99.13 ±0.81**	**99.13 ±0.79**	98.89 ±1.03	97.47 ±1.40
10.	84.46 ±4.06	85.94 ±3.10	80.37 ±5.95	59.98 ±1.90	83.11 ±3.98	82.27 ±4.53	85.10 ±3.72	84.40 ±3.99	**87.08 ±3.38**
11.	79.11 ±3.62	78.61 ±6.94	68.12 ±11.38	56.68 ±21.30	21.97±17.28	85.93 ±4.57	**88.14 ±4.68**	85.04 ±3.93	87.26 ±5.42
12.	93.28 ±4.14	98.01 ±1.16	98.92 ±0.85	38.76 ±12.47	83.04 ±5.64	98.14 ±2.07	97.94 ±2.01	92.62 ±7.69	**99.56 ±0.46**
13.	**97.56 ±1.95**	96.41 ±2.99	97.29 ±1.27	87.68 ±5.98	88.51 ±4.27	93.74 ±4.65	95.95 ±4.70	94.31 ±6.94	97.54 ±1.30
14.	88.20 ±5.71	90.81 ±2.30	91.76 ±3.78	75.50 ±7.54	77.77 ±3.32	90.35 ±5.08	**91.61 ±2.01**	90.00 ±3.80	91.06 ±3.92
15.	56.24 ±5.65	58.75 ±6.77	54.86 ±6.00	35.19 ±13.92	45.57 ±8.93	59.89 ±5.05	55.66 ±8.72	57.10 ±6.78	**61.88 ±4.06**
16.	89.11 ±4.50	**95.18 ±1.67**	93.16 ±4.59	63.95 ±6.47	85.40±2.02	94.02 ±1.97	95.26 ±3.69	90.59 ±6.69	93.84 ±4.34
OA (%)	87.83 ±0.56	88.07 ±0.68	85.92 ±0.53	76.63 ±1.26	80.39±1.05	87.97 ±0.74	87.86 ±1.40	87.92 ±0.80	**89.09 ±0.54**
AA (%)	90.79 ±0.66	91.65 ±0.86	89.80 ±0.93	77.30 ±2.38	80.39±1.08	91.34 ±0.73	92.01 ±1.17	91.14 ±0.93	**92.76 ±0.54**
Kappa (%)	86.39 ±0.63	86.67 ±0.77	84.27 ±0.58	73.76 ±1.50	78.05±1.13	86.56 ±0.84	86.43 ±1.58	86.49 ±0.89	**87.82 ±0.60**

5. Conclusion

In this paper, a new unsupervised algorithm, MIMN-DPP, is proposed for band selection in hyperspectral images. First, by combining the hyperspectral image and its neighbour information, a double graph model is constructed to capture the correlation between bands. A new MIMN criterion is introduced to evaluate the spectral band performance, aiming to maximize the band information and minimize the local noise. Finally, combining with MIMN criterion, the original k-DPP is improved to select a band subset of high diversity and low-redundancy. Comparing with the original k-DPP, our approach has two advantages. First, MIMN-DPP contists of a double graph structure for better use of the spatial information and improved accuracy of the measurement. Second, combining the search criteria of MIMN has helped MIMN-DPP to balance the diversity of band subsets with performance and significantly improve the stability and robustness of data classification. Experimental results on several hyperspectral datasets show that the proposed algorithm outperforms many well-known methods in band selection of HSI.

Some further improvements for future study can be summarized as follows. First, the evaluation criteria for band performance assessment can be explored, such as using more integrated separability between classes for improved performance of band selection. In addition, some improved algorithms of DPP, such as Conditional DPPs and Structured DPPs [27], can also be applied for selecting a more diverse subset of bands.

Acknowledgements

This work is supported in part by the National Nature Science Foundation of China (nos. U1701266), Guangdong Graduate Education Innovation Project (no. 2020XSLT16), the Innovation Team Project of the Education Department of Guangdong Province (no. 2017KCXTD021, 2017KCXTD011), Science and Technology Program of Guangzhou, China (nos. 201803010065, 201802020010), and Guangdong Provincial Key Laboratory of Intellectual Property and Big Data under Grant (no. 2018B030322016).

References

[1] W. Li, C. Chen, H. Su, and Q. Du,Local binary patterns and extreme learning machine for hyperspectral imagery classification, IEEE Trans. Geosci. Remote Sens. 53(7) (2015) 3681-3693.

[2] M. Uzair, A. Mahmood, and A. Mian, Hyperspectral face recognition with spatiospectral information fusion and PLS regression, IEEE Trans. Image Process. 24(3) (2015) 1127-1137.

[3] Y. Yuan, X. Zheng, X. Lu, Discovering Diverse Subset for Unsupervised Hyperspectral Band Selection, IEEE Trans. Image Process. 26(1) (2017) 51-64.

[4] S. Jia, G. Tang, J. Zhu, and Q. Li, A novel ranking-based clustering approach for hyperspectral band selection, IEEE Trans. Geosci. Remote Sens. 54(1) (2015) 88-102.

[5] J. Feng, L. Jiao, F. Liu, T. Sun, X. Zhang, Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images, Pattern Recognit. 51 (2016) 295-309.

[6] T. Qiao, J. Ren, M. Sun, J. Zheng and S. Marshall, Effective compression of HSI using an improved 3D DCT approach for land cover analysis in remote sensing applications, Int. J. Remote Sensing. 35(20) (2014) 7316-7337

[7] J. Ren, Z. Zabalza, S. Marshall and J. Zheng, Effective feature extraction and data reduction with hyperspectral imaging in remote sensing, IEEE Signal Processing Magazine. 31(4) (2017) 149-154

[8] H. Wu, S. Prasad, Semi-supervised dimensionality reduction of hyperspectral imagery using pseudo-labels, Pattern Recognit. 74 (2018) 212-224.

[9] S. Li, Z. Zheng, Y. Wang, C. Chang, Y. Yu, A new hyperspectral band selection and classification framework based on combining multiple classifiers, Pattern Recognition Letters. 83 (2016) 152-159.

[10] Y. Li,W. Xie,H. Li, Hyperspectral image reconstruction by deep convolutional neural network for classification, Pattern Recognition. 63 (2017) 371-383.

[11] J. Tschannerl, J. Ren, P. Yuen, G. Sun, H. Zhao, Z. Yang, Z. Wang and S. Marshall, MIMR-DGSA: Unsupervised Hyperspectral Band Selection Based on Information Theory and a Modified Discrete Gravitational Search Algorithm, Information Fusion. 51 (2019) 189-200.

[12] L. Lei, S. Prasad, J.E. Fowler, L.M, Bruce, Locality-preserving dimensionality reduction and classification for hyperspectral image analysis, IEEE Trans. Geosci. Remote Sens. 50(4) (2012) 1185-1198.

[13] A. Agarwal, T. El-Ghazawi, H. El-Askary, and J. Le-Moigne, Efficient hierarchical-PCA dimension reduction for hyperspectral imagery, in Proc. IEEE ISSPIT. (2007) 353-356.

[14] M. Fauvel, J. Chanussot, and J. A. Benediktsson, Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas, EURASIP J. Adv. Signal Process. (2009) 2009 1-15.

[15] J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, J. Han, Novel Folded-PCA for Improved Feature Extraction and Data Reduction with Hyperspectral Imaging and SAR in Remote Sensing, ISPRS Journal of Photogrammetry and Remote Sensing. 93(7) (2014) 112-122.

[16] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, Mutual-information based semi-supervise hyperspectral band selection with high discrimination, high information, and low redundancy, IEEE Trans. Geosci. Remote Sens. 53(5) (2015) 2956-2969.

[17] X. Geng, K. Sun, L. Ji, and Y. Zhao, A fast volume-gradient-based band selection method for hyperspectral image, IEEE Trans. Geosci. Remote Sens. 52(11) (2014) 7111-7119.

[18] J. Feng, L. Jiao, X. Zhang, and T. Sun, Hyperspectral band selection based on trivariate mutual information and clonal selection, IEEE Trans. Geosci. Remote Sens. 52(7) (2014) 4092-4105.

[19] J.M. Sotoca, F. Pla, Supervised feature selection by clustering using conditional mutual information-based distances,Pattern Recognit. 43(6) (2010) 2068-2081

[20] H. Yang, Q. Du, H. Su, and Y. Sheng, An efficient method for supervised hyperspectral band selection, IEEE Geosci. Remote Sens. Lett. 8(1) (2011) 138-142.

[21] C. Sui, Y. Tian, Y. Xu, and Y. Xie, Unsupervised band selection by integrating the overall accuracy and redundancy, IEEE Geosci. Remote Sens. Lett. 2(1) (2015) 185-189.

[22] C. Wang, M. Gong, M. Zhang, and Y. Chan, unsupervised hyperspectral image band selection via column subset selection, IEEE Geosci. Remote Sens. Lett. 12(7) (2015) 1411-1415.

[23] M. Zhang, J. Ma, and M. Gong, Unsupervised Hyperspectral Band Selection by Fuzzy Clustering with Particle Swarm Optimization, IEEE Geosci. Remote Sens. Lett. 14(5) (2017) 773-777.

[24] W. Xie, Y. Shi, Y. Li, X. Jia , J. Lei, High-quality spectral-spatial reconstruction using saliency detection and deep feature enhancement, Pattern Recognit. 88 (2019) 139-152.

[25] C. Gong, D. Tao, S. J. Maybank, W. Liu, G. Kang, and J. Yang, Multimodal curriculum learning for semi-supervised image classification, IEEE Trans. Image Process. 25(7) (2016) 3249-3260.

[26] C. Chang, Spectral Inter-Band Discrimination Capacity of Hyperspectral Imagery, IEEE Trans. Image Process. 56(3) (2018) 1749-1766.

[27] A. Kulesza and B. Taskar, Determinantal Point Processes for Machine Learning, Foundations and Trends in Machine Learning. 5(2-3), (2012) 123-286.

[28] C. Shannon, A mathematical theory of communication, Bell Labs Technical Journal. 27(4) (1948) 379-423.

[29] A. Kulesza and B. Taskar, k-DPPs: Fixed-size determinantal point Processes, In Proc. Int. Conf. on Machine Learning. (2011) 1193-1200.

[30] O. Macchi, The coincidence approach to stochastic point processes, Advances in Applied Probability. 7(1) (1975) 83-122.

[31] C. Chang, Q. Du, T. Sun, M. Althouse.: 'A Joint Band Prioritization and Band Decorrelation Approach to Band Selection for Hyperspectral Image Classification'. IEEE Trans. Geosci. Remote Sens., 1999, 37(6), pp. 2631-2641.

[32] P. Mitra, C.A. Murthy, S.K. Pal, Unsupervised feature selection using feature similarity, IEEE Trans. Pattern Anal. Mach. Intell. 24(3) (2002) 301-312.

[33] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, NIPS. (2006) 507-514.

[34] Y. Bazi, F. Melgani, Toward an optimal SVM classification system for hyper-spectral remote sensing images, IEEE Trans. Geosci. Remote Sens. 44(11) (2006) 3374-3385.

[35] G.M. Foody, Status of land cover classification accuracy assessment, Remote Sens. Environ. 80 (2002) 185-201.

[36] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, O. Williams, Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Remote. Sens. Environ. 65(3), 1998, 227-248.

[37] S. Holzwarth, A. Muller, M. Habermeyer, R. Richter, A. Hausold, S. Thiemann,P. Strobl,Sens-DAIS 7915 / ROSIS Imaging Spectrometers at DLR, in: 3rd EARSeL Workshop on Imaging Spectroscopy, Herrsching, (2003) 2003 12.

[38] C. Chang, S. Wang, Constrained Band Selection for Hyperspectral Imagery, IEEE Trans. Geosci. Remote Sens. 44(6) (2006) 1575-1585.

[39] S. Jia, Z. Ji, Y. Qian, and L. Shen, Unsupervised band selection for hyperspectral imagery classification without manual band removal, IEEE J. Sel. Topics Appl. Earth Observ. Remote Ses. 5(2) (2012) 531-543.

[40] A. MartÍnez-usó, F. Pla, J.M. Sotoca, P. GarcÍa-sevilla, Clustering-based hyper-spectral feature selection using information measures, IEEE Trans. Geosci. Remote Sens. 45(12) (2007) 4158-4171.

[41] Q. Peng, A. Gentles, and S. Plevritis, Fast Calculation of Pairwise Mutual Information for Gene Regulatory Network Reconstruction, Computer Methods and Programs in Biomedicine. 94(2) (2009) 177-180.

[42] H. Shimazaki, S. Shinomoto, Kernel bandwidth optimization in spike rate estimation, J. Comput Neurosci. 29(1-2) (2010) 171-182.

[43] B. Gao, An Operational Method for Estimating Signal to Noise Ratios from Data Acquired with Imaging Spectrometers, Remote Sensing of Environment. 43(1) (1993) 23-33.

[44] B. Qin, B. Hong, Z. Zhang, X. Yang and Z. Li, A generally applicable noise-estimating method for remote sensing images, Remote Sensing Letters.5(5) (2014) 481-490

[45] M. Yang, C. Deng, F. Nie, Adaptive-weighting discriminative regression for multi-view classification, Pattern Recognition.88 (2019) 236-245.

[46] C. Deng, X. Liu, C. Li, D. Tao, Active multi-kernel domain adaptation for hyperspectral image classification, Pattern Recognition.77 (2018) 306-315.

[47] C. Deng, Y. Xue, X. Liu, C. Li and D. Tao, Active Transfer Learning Network: A Unified Deep Joint Spectral-Spatial Feature Learning Model for Hyperspectral Image Classification, IEEE Trans. Geosci. Remote Sens. 57(3) (2019) 1741-1754.

[48] X. Wei, W. Zhu, B. Liao and L. Cai, Scalable One-Pass Self-Representation Learning for Hyperspectral Band Selection, IEEE Trans. Geosci. Remote Sens. 57(7) (2019) 4360-4374.

[49] G. Zhu, Y. Huang, J. Lei, Z. Bi and F. Xu, Unsupervised Hyperspectral Band Selection by Dominant Set Extraction, IEEE Trans. Geosci. Remote Sens. 54(1) (2016) 227-239.