

Adaptive Distance based Band Hierarchy (ADBH) for Effective Hyperspectral Band Selection

He Sun, Jinchang Ren, *Senior Member, IEEE*, Huimin Zhao, Genyun Sun *Member, IEEE*, Wenzhi Liao, *Senior Member, IEEE*, Zhenyu Fang, and Jaime Zabalza

Abstract—Band selection has become a significant issue for the efficiency of hyperspectral image (HSI) processing. Although many unsupervised band selection (UBS) approaches have been developed in the last decades, a flexible and robust method is still lacking. The lack of proper understanding of the HSI data structure has resulted to the inconsistency in the outcome of UBS. Besides, most of UBS methods are either relying on complicated measurements or rather noise sensitive, which hinder the efficiency of the determined band subset. In this paper, an adaptive distance based band hierarchy (ADBH) clustering framework is proposed for unsupervised band selection in HSI, which can help to avoid the noisy bands whilst reflecting the hierarchical data structure of HSI. With a tree hierarchy-based framework, we can acquire any number of band subset. By introducing a novel adaptive distance into the hierarchy, the similarity between bands and band groups can be computed straightforward whilst reducing the effect of noisy bands. Experiments on four datasets acquired from two HSI systems have fully validated the superiority of the proposed framework.

Index Terms—Hyperspectral band selection, unsupervised learning, hierarchy clustering, adaptive distance.

I. INTRODUCTION

HYPERSPECTRAL images (HSI) contains spectral information in hundreds of contiguous bands. With the aid of a large number of spectral bands, hyperspectral image has been widely used in a range of applications [1]–[3], especially in the remote sensing area, such as precision agriculture [4]–[7], target detection [8], image enhancement [9], [10], object detection [11] and land cover analysis [12], [13], etc. Although numerous bands enable material identification and object detection, the processing of HSI suffers from the “curse of dimensionality” [14]. Besides, there are redundant bands in the HSI, which may lower the efficiency of data analysis. Moreover, due to the high dimensionality of the HSI, the

The authors thank the following funding support: The Dazhi Scholarship of the Guangdong Polytechnic Normal University (GPNU), Innovation Team Project of the Education Department of Guangdong Province(2017KCXTD021), Guangdong Provincial Key Laboratory of Intellectual Property and Big Data under Grant (2018B030322016), National Natural Science Foundation of China (41971292, 41801275, 41871270) and the ETF Scholarship from University of Strathclyde.

Corresponding authors: J. Ren (jinchang.ren@strath.ac.uk), H. Zhao are with School of Computer Sciences, GPNU, Guangzhou 510665, China; G. Sun (genyunsun@163.com) is with Qingdao National Laboratory for Marine Science and Technology and School of Geosciences, China university of Petroleum (East China), Qingdao, China.

H. Sun, J Ren, Z. Fang and J. Zabalza are with the Dept of Electronic & Electrical Engineering, Univ. of Strathclyde, Glasgow G1 1XQ, U.K.

W. Liao is Sustainable Materials Management, Flemish Institute for Technological Research (VITO), Boeretang 200, 2400, Mol, Belgium

Manuscript received September 10, 2019

computational burden is huge. To tackle these problems, it is crucial to reduce the dimensionality of the HSI data whilst preserving the useful spectral information.

Basically, there are two kinds of dimensionality reduction methods for HSI: feature extraction and feature selection. With the feature space transform, feature extraction can project the original data into a lower dimensional space, using approaches such as the principal component analysis (PCA) [15], [16], independent component analysis (ICA) [17], wavelet transform [18], the manifold learning [19], and the maximum noise fraction (MNF) [20], etc. The resulted data can be assumed to contain most of the spectral and spatial information from the original HSI data. Although the feature extraction methods successfully reduce the dimensionality of HSI whilst keeping the discriminative ability, the feature transform itself relies on the whole set of original data and often has poor correspondence to the process of optical acquisition of the data. In contrast, the feature selection method, which is also called band selection, can select an optimised subset from the HSI data, based on their dominant contributions to certain tasks. Since the band selection methods can maintain the physical acquisition characteristic of raw data and solve the high dimensionality problem simultaneously, an efficient band selection method is often preferred.

Generally, based on the availability of the class label information, existing band selection methods can be divided into two groups: i.e. supervised [21]–[23] and unsupervised ones [24]–[27], [29], [30], [32]–[36]. Supervised methods can construct a criterion with the label information of pixels aiming to improve the class separability. In [21], the desired band subset is chosen based on the class-based spectral signatures. By extracting two most distinctive bands whose dissimilarity is the largest among all bands, other bands can be chosen iteratively by minimizing the estimated abundance covariance from each pixel along with the class information. Cao *et al.* [22] proposed another wrapper-based supervised band selection method, where the chosen band subset is determined based on minimizing the defined local smoothness with the aid of the classification map from a Markov random field (MRF) classifier. To improve the reliability of the local smoothness generated from the classification map, the wrapper method is utilized to initialize the designed method. In [23], Patra *et al.* developed a rough-set-based supervised band selection method. The rough-set theory is applied to compute the relevance and significance of each band by using the class information as a prior knowledge, and bands with higher relevance and significance are chosen to form the band subset.

Although the band subset acquired by the supervised methods can achieve better classification performance, the selected bands are often affected by the chosen training samples where different training samples may lead different result of selected band subsets. Furthermore, these approaches can become less effective in practical applications if sufficient training samples with label information are not approachable. Even though some supervised band selection methods only choose few training samples, the classification performance with less band and less training samples are not reliable as a criterion for band selection. Therefore, we will focus on the unsupervised band selection (UBS) methods in this paper.

Based on certain searching strategies, UBS methods aim to select the most representative bands among the HSI data. Recently, many searching strategies have been developed for HSI band selection, which can be separated into two main groups: the ranking-based and the clustering-based methods. Various statistical metrics have been utilized to evaluate each band in the ranking-based methods, including mutual information [37], [38], variance [39] and local density [29], etc. After the band ranking, the desired band subset is determined by selecting bands with higher ranking values among all bands. Since the ranking process is only implemented once, the computational cost can be rather low. For the clustering-based methods [26], [27], [30], [31], spectrally continuous bands are grouped into desired clusters. Bands in each cluster are contiguous and with similar spectral information, where the most significant band in each cluster based on discriminative ability [27] or some ranking strategies [30] are selected to form the desired band subset. Due to the clustering procedure, this process can be lengthy whilst the selected bands are generally uncorrelated.

Although the aforementioned two groups of UBS methods have achieved certain success for band selection in HSI, both of them still suffer different drawbacks. For ranking-based approaches, the correlation between selected bands is usually quite high, where the data redundancy can be further reduced. On the contrary, the clustering-based methods usually select one band from each band cluster, thus the data redundancy is low. However, most of the clustering-based methods are very sensitive to the noisy bands because a noisy band can easily form a cluster due to low similarity to other bands thus affect the selection result. Meanwhile, the results of band selection depend on the clustering process, especially on the number of clusters. For example, a certain band can be selected when the number of clusters is three but it can then be deselected when the number of band clusters becomes five, where such inconsistency may lead to low robustness of UBS. Furthermore, the similarity metric between different bands plays a key role in clustering methods, including the efficacy and computational complexity. Some clustering-based methods may have a good performance, but their computational cost can be high due to the complicated metrics.

To tackle the aforementioned drawbacks, we propose a band hierarchy clustering UBS framework with adaptive distance (ADBH). The contributions can be summarized as follows:

- 1) A flexible tree hierarchy-based framework for HSI band selection is proposed. With the tree structure, the band subset with desired number of bands can be easily

chosen whilst minimising the data redundancy between selected bands. Moreover, the chosen bands with different numbers of clusters are more consistent when derived from the proposed tree hierarchy.

- 2) Targeting the noise sensitivity issue of clustering-based UBS methods, we design a novel distance measurement combined with the cluster density and the Euclidean distance. With the defined distance, the proposed ADBH framework is more robust to the noisy bands whilst the computational cost is acceptable.
- 3) We have applied our ADBH on four commonly used HSI datasets, the performance of our method has demonstrated the superiority comparing with the current state-of-the-art (SOTA) UBS methods.

The rest of this paper is organized as follows. Section II introduces related UBS methods. The proposed methodology is presented in Section III. In Section IV, experimental results and discussions are given on four HSI datasets. Finally, some concluding remarks are drawn in Section V.

II. RELATED WORK

In the last two decades, a number of approaches have been proposed for unsupervised band selection (UBS) in HSI. In this section, some typical UBS approaches from the aforementioned two groups, i.e. the ranking-based and the clustering-based methods, will be reviewed, and relevant analysis to motivate the proposed work is also given.

As mentioned in the last section, the goal of the ranking-based UBS methods is to find the most significant bands among the HSI data. To fulfil this purpose, an effective criterion for estimating the importance of each band is essential. With the aid of the designed criterion, most representative bands can be determined. In [24], a PCA-based band selection criterion was proposed. By applying the maximum-variance PCA (MVP-PCA), the band prioritization can be estimated according to the eigenanalysis. A defined load factor of each band can be obtained from the consolidation of eigenvalue and eigenvector. For each band, a variance-based band power ratio is utilized to represent its discriminative ability, which is accessed by using the variance of each band to divide that of all bands. By finding the bands with higher ratio, a band subset is determined. Although the chosen bands are more representative and more discriminative, the correlation between those bands are ignored in the MVP-PCA. The robustness of selected bands is not guaranteed as they are with higher variance. Chang and Wang [25] have presented a constraint band correlation strategy (CBS), which is derived from the idea of constrained energy minimization. By defining a finite impulse filter between each band and the whole dataset, the correlation can be represented by a minimized vector. After discarding bands with high correlation, the remaining bands are selected, which can be more robust to the noisy band.

Different from the ranking-based methods, clustering-based methods can naturally reduce the correlation between chosen bands. In these approaches, the HSI bands are sequentially grouped into different clusters by a defined criterion. Afterwards, typical bands from each cluster are selected to form

the desired band subset. Since the band subset comprises bands from different clusters, the high correlation between bands can be avoided. In [26], a hierarchical clustering (WaLuDi/WaLuMi) is applied to divide bands of whole dataset into segments. Two measurement, mutual information and K-L divergence, have been utilized to measure the distances between bands. In terms of the Ward's linkage theory [40], partitions with minimum variance can be achieved, and the band which is most identical to the rest is selected in each cluster. By considering the contextual information of the HSI dataset, Yuan *et al.* have proposed a novel clustering method, i.e., dual-clustering-based band selection by context analysis (DCCA), for UBS [27]. Along with the input raw HSI data, the DCCA has designed a new pairwise hyperspectral angle descriptor to exploit the contextual information of each pixel in HSI. With the dual clustering framework, the contextual feature of the HSI and the raw HSI are grouped simultaneously and the mutual effect of these two features determine the clustering result. Similar to other clustering-based methods, the most representative band from each cluster is selected based on a groupwise strategy.

Nowadays, it has become a trend to combine the ranking-based and the clustering-based methods. For the ranking-based methods, most representative bands can be easily found. Meanwhile, the clustering-based methods can restrict the correlation within the obtained subset of bands. Therefore, the merits from these two methods can enhance the performance of UBS. Inspired by the fast-peak-based clustering (FDPC) [28], Jia *et al.* have proposed the enhanced FDPC (E-FDPC) [29] where the characteristic of each band can be determined by its local density and its distance to the nearest high density band. The significance of each band can be determined by considering these two factors jointly. Based on the assumption that the band with a higher local density and maximum nearest neighbour distance is the cluster centre, top ranked bands are chosen to form the band subset, which is still similar to most ranking-based methods. Different from the E-FDPC which combines the clustering-based methods into the ranking-based methods, Wang *et al.* has further developed an optimal clustering framework (OCF) for HSI band selection [30]. With two defined objective functions, the normalized cut and top-rank cut have been used to partitioned the whole dataset into several clusters by an optimal way. Three ranking strategies, including E-FDPC, MVPCA, and Information Entropy, are utilized to find the most important band from each cluster. The performance of OCF has validated the successfully cooperation between ranking-based and clustering-based UBS methods. In [31], the adaptive subspace partition strategy (ASPS) has been proposed for UBS in HSI. By applying a coarse to fine strategy, the bands are grouped into different subcubes. By estimating the noise information for each band, the band with the minimum noise is considered as the most representative one for that subcube and added to the subset of selected bands. The experimental results have further emphasized the importance of removing the noisy band from the selected band subset.

Recently, in addition to the ranking-based and clustering-based methods, optimization based UBS methods have attracted increasing attention as the iterative process seems more

controllable to obtain the number of selected bands. The volume gradient band selection method (VGBS) is introduced by deriving the 'volume' information from the covariance matrix of all bands [32]. Instead of calculating any measurements between a single band and all other bands, VGBS removes the most redundant band by the assumption that it usually has the maximum gradient in the dataset. Different from the VGBS algorithm, the multitask sparsity pursuit (MTSP) [33] attempts to find an optimal solution by iteratively updating the chosen band subset. In MTSP, a constructed data descriptor based on the compressive sensing theory is firstly utilized to reduce the original HSI data, and a band subset with the desired number of bands can be obtained randomly. Afterwards, a multitask sparse representation based criterion is utilized to examine the potential band groups. By updating the preliminary band subset using the immune clonal strategy, the optimized result can be obtained. Under the consideration of structure information from both band informativeness and independence, Zhu *et al.* developed a greedy-search based UBS approach by tackling a graph-based clustering problem with dominant set extraction (DSEBS) [34]. The DSEBS takes the advantage of the first-order statistic of local spatial-spectral consistencies and structure correlation for quantifying band information and independence. After that, the band selection task is transformed to a dense subgraph discovery problem, where the dominant set extraction can provide an optimal solution. In DSEBS, the interdependencies between bands determine the reliability of each band and its contribution to the final result. By choosing the optimal band subset iteratively, the optimization-based UBS methods have comparable achievement. However, two major drawbacks restrict the performance of this kind of methods. Foremost, the iterative process usually focuses more on each individual bands, which fails to filter the contributions from noisy bands. Secondly, there is a trade-off between the computational complexity and performance in the iterative process, hence some valuable information may be compromised for reducing the complexity.

III. METHODOLOGY

In this section, our proposed ADBH framework for UBS will be presented in detail. First, we describe our tree hierarchy-based clustering strategy. Followed by the adaptive distance measurement within the ADBH framework, which is based on the fusion of the Euclidean distance and cluster density. Afterwards, the band evaluation and selection method is introduced. Finally, the advantages of our method are analysed. Fig. 1 illustrates the flowchart of our proposed ADBH framework. In the proposed framework, the raw HSI dataset is taken as input for both band clustering and band-based ranking. At first, each spectral band is considered as a cluster to form the initial similarity matrix, from which a tree-based band hierarchy can be constructed. Cluster-based adaptive distance (AD) is then calculated, and mutual neighbouring clusters are merged sequentially according to the determined AD. Afterwards, the similarity matrix will be updated, which actually becomes smaller due to the merged band clusters. The process above forms our proposed ADBH, where the

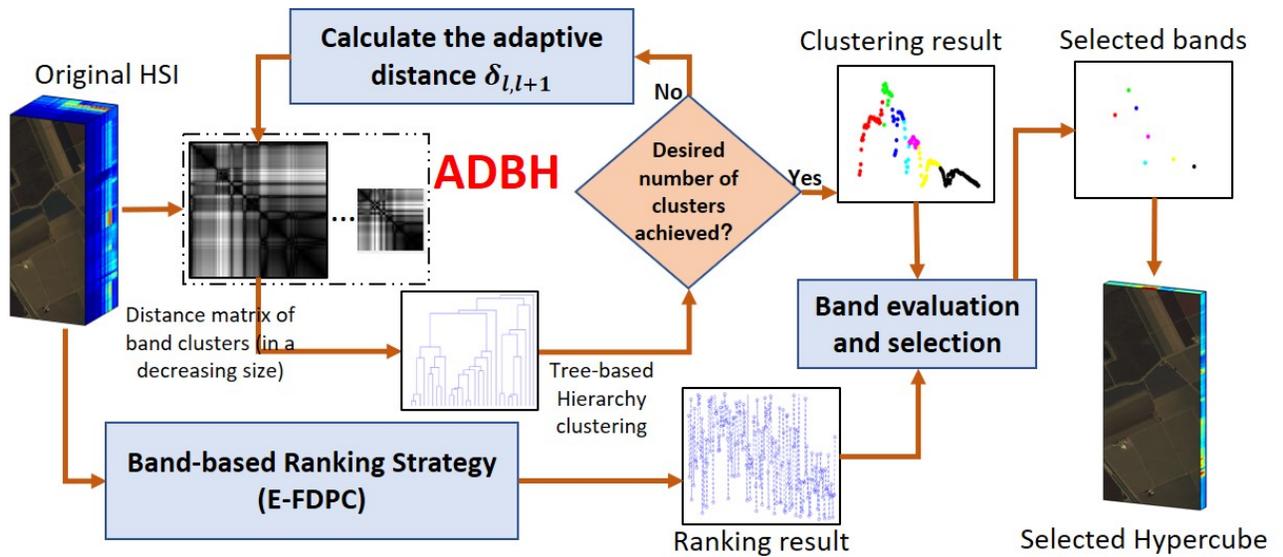


Fig. 1: The flowchart of the proposed ADBH framework

process continues until the number of band clusters reaches the desired number of selected bands. Relevant bands within the resulted band clusters will be ranked by the band-based ranking strategy (E-FDPC) before band selection. The band with the highest ranked value within each cluster is selected as the most representative band for that cluster, and all the selected bands are then grouped to form a dimension-reduced hypercube for following-on processing and analysis.

A. Band hierarchy

The clustering-based UBS methods aim to group similar bands into different clusters and select one most significant band from each cluster, which can reduce the data redundancy between selected bands. Due to the lack of ground truth, the number of band clusters and the exact indexes of bands for each cluster are actually unknown. As a result, the results of band clustering and the derived band subset become arbitrary, where the consistency of the results can hardly be maintained. To tackle this particular challenge, in this work, we propose a band hierarchy algorithm. Our method can construct a band hierarchy in a bottom-up manner and generate any number of band clusters (between one and the original amount of bands). As such, a better understanding of the HSI bands can be derived. Moreover, the clustering results can keep consistency despite of various number of bands are chosen. For instance, with desired k bands, our tree hierarchy can produce k clusters by an iterative way. When a band group with $k - 1$ groups is requested, the result will be adjusted in a flexible way by merging two clusters. Similarly, the result can be easily adjusted to $k + 1$ groups by cancelling the last merging operation. For iteration-based methods, the computational burden is a common challenge. For efficiency, complicated metrics or complex strategies are avoided in our ADBH framework as explained below.

Let us denote a HSI image as $Y \in R^{M \times N \times L}$, where the spatial size of this cube is $M \times N$ and L is the total number

of bands. The l th band can be represented as one vector $Y_l \in R^{1 \times M \times N}$ and the spectral signature of one pixel at the spatial location (m, n) can be denoted as $Y_{mn} \in R^L$. For reducing the computational cost, the spectral value of each pixel is normalized to the scale of $[0, 1]$. Let $G = (V, E)$ denote the HSI data in an undirectional graph, where the node set $V = [1, 2, \dots, l, \dots, L]$ represents the spectral bands in the HSI dataset. Considering the whole dataset as a forest, each band can be considered as a tree, i.e. each band is an individual cluster initially. E is the utilized similarity metrics to measure the connection of different clusters (bands). Due to the contiguous nature of the spectral bands in HSI, each band is assumed to be more closely linked to its neighbouring bands in the spectral domain. To this end, $E = [e_1, \dots, e_l, \dots, e_Z]$ represents the linkage between different clusters, where e_l represents the ‘edge’ between the l th cluster and the $(l + 1)$ th cluster with $1 \leq Z \leq (L - 1)$. Besides, for the first cluster and the last cluster, they only have one edge to connect with their neighbours according to our assumptions above. As a result, it is not necessary to estimate the similarity matrix between bands after each iteration instead of computing similarities between neighbouring band clusters. After that, we can start our tree hierarchy clustering in a bottom-up manner detailed as follows.

First of all, we define a ‘mutual nearest neighbouring’ according to the similarity between each cluster, which is very similar to the mutual nearest neighbours defined in [41]. By examining all connecting edges of each cluster, two clusters can become ‘nearest neighbour’ when they both have lighter edge with each other. For example, if $e_l < e_{l-1}$, the l th cluster is more close to the $(l + 1)$ th cluster, but the l th cluster and the $(l + 1)$ th cluster can be ‘nearest neighbours’ only if $e_l < e_{l+1}$ is also met. This criterion can identify similar clusters pairs and can be utilized in the following-on merging procedure.

After the ‘mutual nearest neighbouring’ search, we can start to merge the current clusters. To reflect and be consistent

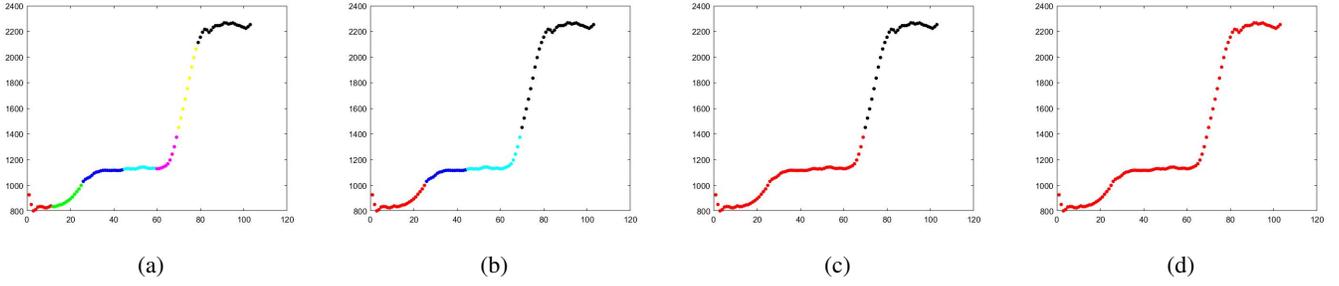


Fig. 2: The Clustering results with different desired number of clusters on the Pavia University dataset. In each figure, the horizontal axis represents the Band Index, and the vertical represents the mean spectral value. Different color represents different clusters (a) 7 clusters, (b) 5 clusters, (c) 2 clusters, (d) 1 cluster.

with the data structure of the HSI dataset, the merging is executed in a sequential way. With all the obtained pairs of clusters, the implementation starts from the pair with the smallest edge. Different from some clustering methods which merge the data sample points gradually [42] (i.e. one merging operation in one iteration), each iteration of our algorithms will not be completed until all the mutual neighbouring clusters are merged, i.e. merging all such band pairs simultaneously. For each new cluster, it is depicted by the mean spectral information of its comprised bands and the previous bands are removed while the spectral information is kept for after iterations. This is shown below:

$$\hat{Y}_i = \text{mean}(Y_{merged}) \quad (1)$$

the representation of the new l th cluster is the mean of all bands it contained. With new clusters, the defined E will also be updated before next iteration. The number of contained bands in each new cluster is also stored. For the circumstance that no nearest neighbour pairs exist, our framework will merge clusters gradually with only one merging in one iteration. In this situation, two clusters with the smallest edge will be merged. The whole clustering procedure will continue iteratively until the desired number of clusters has reached. As the purpose of the clustering step is to group similar bands together, the objective function can be transformed to minimize the cost function during clustering:

$$\min \sum_{t=1}^T e^t \quad (2)$$

where the $t = [1, \dots, T]$ is the evolution time and the e^t is the sum of merged e during the t th iteration.

In our clustering part, the bottom-up manner considers each band as an initial cluster, where the analogous bands can be determined via our 'mutual nearest neighbouring' approach. In each iteration, all the neighbouring pairs of bands can be merged simultaneously, and a stepping method is employed to combine clusters in case of such neighbouring pair of bands remaining in certain iteration. This iterative process will only stop after the requested number of clusters have been reached. An example of our clustering process is shown in Fig. 2.

B. Adaptive distance

Although the tree hierarchy method can help to understand the data structure of bands within HSI, noisy bands are still a serious problem in all hierarchy-based clustering methods. As the bands are clustered in a bottom-up manner, potential outlier of bands can be easily identified as a primary cluster in a similar way as other bands. The outlier is prone to forming a cluster even after numerous iterations because it is less correlated or similar to its neighbouring bands in the band hierarchy. Since the final result consists of bands selected from each cluster, it is inevitably that noisy bands may be added into the selected band subset. Besides, the distance measurement for inspecting the similarity between bands is another crucial issue in our hierarchy. As the distance measurement needs to be updated in each iteration, a complicated one may result in huge computational burden. Thus, an efficient yet robust distance is introduced in our band hierarchy as detailed below.

To estimate the differences between two variables, the Euclidean distance is regarded as one fundamental metric. In most of the clustering work, the Euclidean distance is widely used to assess the differences [43], [44]. In [28] and [29], distances of different bands in HSI are measured using the Euclidean distance to form a distance matrix $S \in R^{L \times L}$ as:

$$S_{ij} = \|Y_i - Y_j\|^2 = \sum_{m,n=1}^{M,N} (Y_{mni} - Y_{mnj})^2 \quad (3)$$

where the entry S_{ij} represents the difference between the i th band the j th band. According to the matrix S , a scaled distance can be obtained as [28], [29]:

$$D_{ij} = \sqrt{S_{ij}} * L^{-1} \quad (4)$$

In our proposed ADBH framework, we apply the aforementioned distance by setting $e_l = D_{l,l+1}$. However, the obtained result shows that the Euclidean distance is unstable for noisy datasets, for instance, the highly polluted KSC dataset. By only applying the Euclidean distance, it is likely to have the noisy bands as separate clusters because these noisy bands are usually sufficiently dissimilar to other neighbouring bands. To tackle this issue, we propose a novel adaptive distance (AD) for measuring the distance of bands by considering the number of bands within the associated cluster.

Basically, there are two motivations for designing the AD. The first is to restrict or even avoid a-single-band cluster

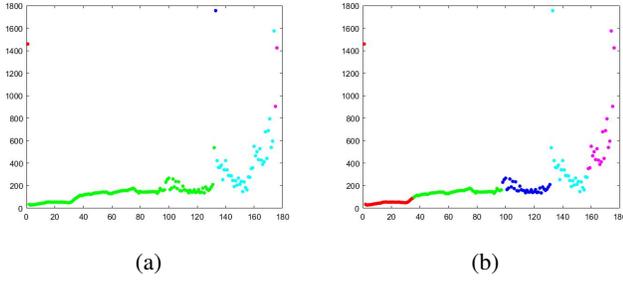


Fig. 3: The Clustering results (defined cluster number equals to 5) by Euclidean distance (a) and our AD (b) on the noisy KSC dataset. In each subfigure, the horizontal axis represents the band index and the vertical axis the mean spectral value. Different color represents different clusters.

formed by noisy bands as it will interfere the results of band selection. The second is to improve the computational efficiency especially during the iterative process of band clustering. Inspired by the above two motivations, we have designed a novel metric to estimate the distance between two adjacent clusters instead of adopting the Euclidean distance. As a regular cluster usually has more than one band, the number of contained bands is considered as a crucial metric to present the density of each cluster. To effectively represent the characteristic of each cluster, we have also estimated the Euclidean norm of each cluster. The Euclidean norm of one cluster \hat{Y}_l in (1) corresponds to the average magnitude of this cluster, which can be assumed as a simple data characteristic of \hat{Y}_l . Considering the representation of each cluster as a vector, we have found out that the product of its magnitude and contained bands can reflect its strength. In this way, the cluster density can be determined by both the number of contained bands and the data characteristics in each cluster. Accordingly, we define a novel measurement for estimating the cluster density I_l :

$$I_l = \text{norm}(\hat{Y}_l) * b_l \quad (5)$$

where b_l is the number of contained bands in the l th cluster, which has an initial value of $b_l = 1$. For a cluster with a single band, I_l is the Euclidean norm of that band. Otherwise, I_l is roughly the accumulated Euclidean norm of all the bands within the cluster. With more bands contained in a cluster in our proposed band hierarchy, the cluster density increases in nearly a linear way.

For two neighbouring clusters (the cluster can be a single band before the iterative process) l and $l + 1$, their densities are denoted as I_l and I_{l+1} according to (5). The defined AD $\delta_{l,l+1}$ is given by combining the Euclidean distance and cluster density as:

$$\delta_{l,l+1} = D_{l,l+1} * I_l * I_{l+1} \quad (6)$$

From this proposed distance, a cluster with a lower density will have shorter distance with its adjacent clusters comparing to other clusters with larger densities. As shown in Fig. 3, the first band of the KSC dataset has a distinct spectrum against its neighboring bands, thus it can be easily regarded as an

outlier in the dataset. In the Fig. 3 (a), this band is considered as a single-band cluster when the Euclidean distance is used to measure the distance between band clusters. Accordingly, this band will be selected because it is the only representative band within the cluster. However, in our proposed AD scheme, this band will be suppressed and grouped into other clusters. During the AD based clustering process, the density of a single band cluster will be relatively small due to the fact that it contains only one band. By applying $e_l = \delta_{l,l+1}$ into the band hierarchy, e_l will become quite small thus for the cluster with less bands can easily find its mutual nearest neighbour. As a result, noisy bands will be simply merged in our proposed ADBH hierarchy, which also meets the energy minimization principle according to (2). Compared to the commonly used Euclidean distance, our proposed ADBH combines the Euclidean distance with the cluster density, in which the cluster density is estimated by multiplying the Euclidean norm of the mean band and the number of bands contained in the associated cluster. In this way, the computational complexity of the proposed AD is further reduced for efficiency. In addition, for a cluster with noisy band being merged, the representative band can be selected by avoiding these noisy bands with the E-FDPC band ranking scheme, which is further detailed in the next subsection.

Algorithm 1 ADBH

- 1: **Input:** Raw HSI data Y , desired number of bands K .
 - 2: **Initialize:** Assume each band as a cluster.
 - 3: **BEGIN**
 - 4: **while** Number of clusters $> K$ **do**
 - 5: Update the AD among clusters by (5) and (6);
 - 6: **if** Mutual neighbouring clusters exist **then**
 - 7: Merging mutual neighbouring clusters pairs sequentially according to their edge;
 - 8: Update new cluster;
 - 9: **if** Current Number of clusters = K **then**
 - 10: Return clustering result C ;
 - 11: Break;
 - 12: **end if**
 - 13: **else**
 - 14: Merging two clusters with lightest edge;
 - 15: Update new cluster;
 - 16: **if** Current Number of clusters = K **then**
 - 17: Return clustering result C ;
 - 18: Break;
 - 19: **end if**
 - 20: **end if**
 - 21: **end while**
 - 22: Choose band subset X among clustering result C by (7).
 - 23: **Output:** Band subset X .
 - 24: **END**
-

C. Band evaluation and selection

In our proposed ADBH, the whole dataset can be grouped into several clusters of bands with similar characteristic. To select the most representative band from each band cluster, the ranking or priority of each band needs to be determined. Recently, many metrics [29], [37]–[39] have been utilized for this purpose. Among those criteria, E-FDPC is employed as it provides an efficient solution for determining bands with high discriminative ability. Due to the fact that a band which has large local density can be more easily chosen than others, E-FDPC is robust to the noisy bands. Although the E-FDPC is still substantially a ranking-based method, the combination of E-FDPC and the clustering process has proved to be effective [30]. Therefore, we have applied the E-FDPC algorithm after band clustering work, where the most vital band within each band cluster can be chosen to form the desired band subset. This ranking-based strategy is described as follows:

Denote the clustering result as $C = [c_1, \dots, c_k, \dots, c_K]$, where c_k is the k th cluster and $k = [1, \dots, K]$ is the cluster index with the desired number of bands equalling to K . As the band with the highest value in each cluster is the most vital one, the desired band X_k from the k th cluster can be determined as:

$$X_k = \arg \max_{\psi} \psi_{k_v} \quad (7)$$

where ψ is the rank values set for all bands and ψ_{k_v} is the rank value of the v th band in the k th cluster. The band with the highest rank value in the k th cluster is chosen as a band for the desired band subset X . Obviously, the band selection result can be decided with the aid of our proposed ADBH.

D. Merits of ADBH

With the designed adaptive distance, our ADBH helps to complete the UBS task in a bottom-up tree hierarchy. As the merging process starts from the smallest edge, the sequence can be recorded and the band clustering process can be visualized easily. In Fig. 2, partial of the clustering process from our ADBH of the Pavia University dataset is shown. We have chosen results from certain numbers of clusters to verify the consistency. This advantage may help to further understand the HSI dataset, where any desired number of bands can be easily determined. Secondly, the designed ADBH framework can be regarded as a parameter-free method, which means no other input parameters are needed except only the desired number of bands along with the raw data. Besides, the clustering result will not be affected by varying the requested number of clusters, where the consistency can always be kept. Finally, the clustering results can be improved with our defined similarity metric, i.e. the AD, which is verified on the KSC dataset in Fig. 3. It can be seen that the single band cluster is removed after applying the AD into the tree hierarchy, which has successfully suppressed the noisy band being chosen as part of the selected band subset. The proposed UBS framework is summarized in Algorithm 1, and some further experimental results are discussed in the next section to demonstrate the efficacy of the proposed ADBH method for UBS in HSI.

IV. EXPERIMENTAL RESULTS

Due to the lack of ground truth, the efficacy of band selection is often indirectly evaluated by using the classification accuracy with the selected bands. In our experiments, the proposed ADBH framework is benchmarked with several SOTA algorithms based on the classification results from four popular HSI datasets. Relevant details are presented as follows.

A. Datasets

To evaluate the performance of our proposed ADBH framework, four HSI datasets from two imaging systems have been used. The first one is the Indian pines dataset, which was collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the agricultural experimental field located at North-Western, Indian, USA in 1992. The original dataset has 224 spectral bands ranging from 0.4 to 2.5 μm with 16 manually labelled class, and its spatial size is 145×145 pixels with 10249 labelled pixels. After the removal of 24 water absorption bands, the rest 200 bands are utilized for band selection and data classification. The second dataset is the Pavia University (PaviaU), which was captured by the Reflective Optics System Imaging Spectrometer (ROSIS) system over the campaign of the university of Pavia, Italy in 2002. The PaviaU dataset has a spatial size of 610×610 pixels and 103 spectral reflectance bands with the spectral range from 0.43 to 0.86 μm . A cropped image of 610×340 pixels are employed after discarding pixels with no information. In the PaviaU dataset, 42776 pixels from 9 semantic classes are labelled. The third dataset is the Salinas scene (Salinas), which was also captured by the AVIRIS in Salinas Valley, California, USA in 1998. Same as the Indian pines dataset, the Salinas dataset collects spectral information within 0.4-2.5 μm in 224 bands. Its ground truth data also has 54129 labelled pixels from 16 classes and its image spatial size is 512×217 pixels. Similar to the Indian pines dataset, the Salinas dataset in our experiments also has 20 water absorption bands removed with the rest 204 bands for analysis. The last dataset is the Kennedy Space Center (KSC) dataset, which was obtained using the same AVIRIS sensor in Florida, USA, 1996. By removing the water absorption and low SNR bands, only 176 bands are used with 13 labelled classes, and the spatial size of this dataset is 512×614 pixels and 19035 pixels are manually labelled.

B. Experimental settings

To evaluate the performance of our ADBH framework in HSI classification, we have compared our framework with SOTA algorithms, including OCF (TRC-OC-FDPC) [30], VGBS [32], DSEBS [34], WaLuDi [26], WaLuMi [26], E-FDPC [29] and ASPs [31]. It is worth noting that our algorithm is parameter-free, only the HSI data and the desired number of bands are needed as input. Similarly, OCF does not have any determined parameters and experiments are implemented on code provided by authors. For other methods including VGBS, DSEBS, WaLuDi, WaLuMi, E-FDPC, and ASPs, experiments are tested on original codes with default parameters. For better investigating the effect of our proposed

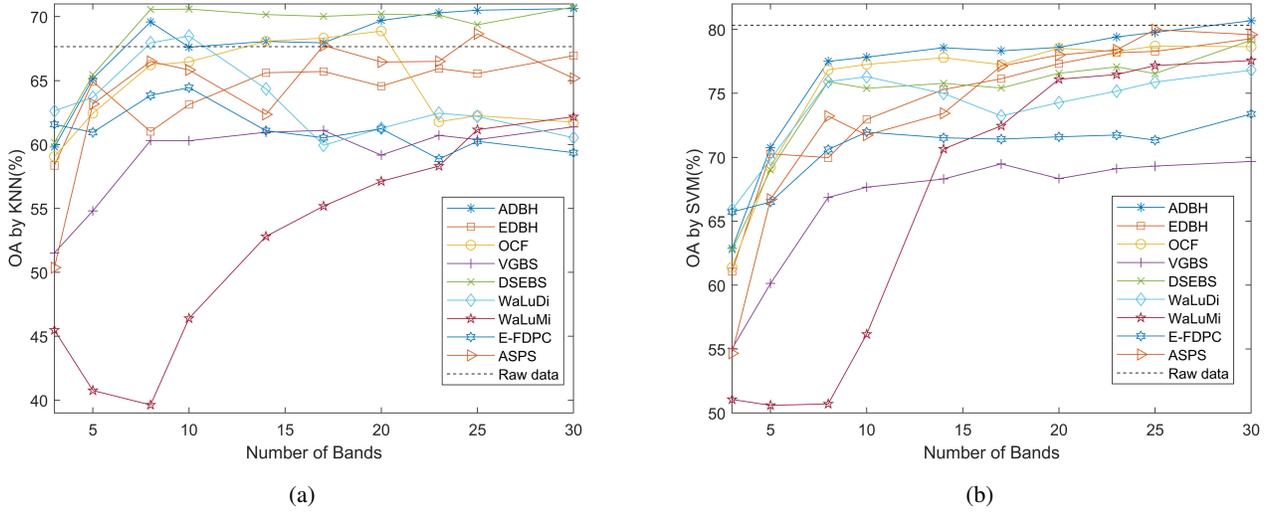


Fig. 4: OA curves on the Indian pines dataset with different UBS methods by using KNN (a) and SVM (b).

TABLE I: Classification results from different approaches for the Indian pines dataset.

| Classifier | OCF | VGBS | DSEBS | WaLuDi | WaLuMi | E-FDPC | ASPS | EDBH | ADBH | Raw data |
|--------------|------------|-------------|-------------------|------------|------------|------------|------------|------------|-------------------|------------|
| OA by KNN(%) | 68.07±0.01 | 60.96± 0.01 | 70.16±0.01 | 64.35±0.01 | 52.81±0.00 | 61.08±0.01 | 62.35±0.01 | 65.15±0.01 | 68.06±0.01 | 67.65±0.01 |
| AA by KNN(%) | 58.27±0.02 | 48.39±0.00 | 56.05±0.01 | 51.22±0.00 | 40.42±0.00 | 46.68±0.01 | 49.57±0.01 | 52.86±0.01 | 58.43±0.01 | 54.22±0.01 |
| Kappa by KNN | 0.63±0.01 | 0.55±0.01 | 0.66±0.01 | 0.59±0.01 | 0.45±0.00 | 0.55±0.01 | 0.57±0.01 | 0.60±0.01 | 0.63±0.01 | 0.63±0.01 |
| OA by SVM(%) | 77.79±0.01 | 68.30±0.01 | 75.78±0.01 | 74.99±0.01 | 70.65±0.01 | 71.52±0.01 | 73.44±0.01 | 75.3±0.01 | 78.52±0.01 | 80.33±0.01 |
| AA by SVM(%) | 76.82±0.01 | 64.53±0.02 | 74.96±0.01 | 75.58±0.01 | 67.35±0.01 | 70.57±0.02 | 73.50±0.01 | 73.30±0.01 | 77.75±0.01 | 72.09±0.01 |
| Kappa by SVM | 0.75±0.01 | 0.64±0.01 | 0.72±0.00 | 0.72±0.01 | 0.67±0.01 | 0.67±0.01 | 0.70±0.01 | 0.72±0.01 | 0.76±0.01 | 0.78±0.01 |

TABLE II: Classification results from different approaches for the PaviaU dataset.

| Classifier | OCF | VGBS | DSEBS | WaLuDi | WaLuMi | E-FDPC | ASPS | EDBH | ADBH | Raw data |
|--------------|------------|-------------|------------|-------------------|------------|------------|-------------------|-------------|-------------------|------------|
| OA by KNN(%) | 85.64±0.01 | 82.64± 0.01 | 84.58±0.01 | 86.43±0.01 | 85.77±0.01 | 84.73±0.01 | 86.41±0.01 | 81.38±0.00 | 85.57±0.01 | 85.73±0.01 |
| AA by KNN(%) | 82.55±0.00 | 77.08±0.00 | 81.32±0.00 | 82.31±0.00 | 83.11±0.00 | 81.24±0.00 | 83.47±0.01 | 75.61±0.00 | 81.76±0.00 | 82.02±0.01 |
| Kappa by KNN | 0.80±0.00 | 0.76±0.00 | 0.79±0.01 | 0.81±0.00 | 0.42±0.00 | 0.79±0.00 | 0.82±0.00 | 0.75±0.01 | 0.80±0.00 | 0.81±0.01 |
| OA by SVM(%) | 90.88±0.00 | 90.75±0.00 | 89.39±0.00 | 91.08±0.00 | 91.40±0.00 | 90.07±0.00 | 83.87±0.01 | 83.77±0.00 | 91.63±0.00 | 91.64±0.01 |
| AA by SVM(%) | 88.74±0.00 | 88.25±0.00 | 87.23±0.00 | 88.96±0.00 | 88.82±0.00 | 85.04±0.0 | 72.30±0.00 | 73.31±0.000 | 89.30±0.00 | 88.12±0.01 |
| Kappa by SVM | 0.88±0.00 | 0.88±0.00 | 0.86±0.00 | 0.88±0.00 | 0.88±0.00 | 0.84±0.00 | 0.78±0.01 | 0.78±0.00 | 0.89±0.00 | 0.89±0.00 |

TABLE III: Classification results from different approaches for the Salinas dataset.

| Classifier | OCF | VGBS | DSEBS | WaLuDi | WaLuMi | E-FDPC | ASPS | EDBH | ADBH | Raw data |
|--------------|------------------|-------------|-------------------|------------|------------|------------|------------|------------------|-------------------|------------|
| OA by KNN(%) | 88.82±0.00 | 82.74± 0.00 | 88.99±0.00 | 85.98±0.01 | 87.85±0.00 | 87.58±0.00 | 86.99±0.00 | 87.03±0.00 | 88.16±0.00 | 87.70±0.01 |
| AA by KNN(%) | 93.58±0.00 | 85.6±0.00 | 93.73±0.00 | 92.53±0.00 | 92.23±0.00 | 92.52±0.01 | 92.06±0.00 | 91.94±0.00 | 93.21±0.00 | 93.27±0.01 |
| Kappa by KNN | 0.88±0.00 | 0.81±0.01 | 0.88±0.01 | 0.86±0.01 | 0.84±0.00 | 0.86±0.01 | 0.86±0.01 | 0.86±0.00 | 0.87±0.01 | 0.86±0.01 |
| OA by SVM(%) | 92.28±0.00 | 92.04±0.00 | 92.28±0.00 | 91.03±0.00 | 92.31±0.00 | 92.45±0.00 | 91.90±0.00 | 92.40±0.01 | 92.72±0.00 | 92.87±0.00 |
| AA by SVM(%) | 95.67±0.00 | 95.28±0.00 | 95.93±0.00 | 95.68±0.00 | 95.18±0.00 | 95.91±0.00 | 95.63±0.00 | 95.79±0.00 | 96.04±0.00 | 96.42±0.00 |
| Kappa by SVM | 0.91±0.00 | 0.91±0.00 | 0.91±0.00 | 0.91±0.00 | 0.90±0.00 | 0.91±0.00 | 0.91±0.00 | 0.92±0.00 | 0.92±0.00 | 0.92±0.01 |

AD, we have compared the method employing the Euclidean distance instead of our proposed AD, which is represented as euclidean distance-based band hierarchy (EDBH). To better verify the effectiveness of the proposed ADBH framework, the classification results using all bands (shown as 'Raw data' in corresponding tables and figures) are also included.

For the classification part, two popular classifiers, K -Nearest Neighbourhood (KNN) [45] and Support Vector Machine (SVM) [46], are employed to validate the classification accuracy of chosen band subsets on classification of the aforementioned four HSI datasets. In our experiments, the

parameters in SVM and KNN are optimized through 10-fold cross-validation. In all four HSI datasets, 10% of the samples from each class are randomly selected as the training samples for both classifiers, whilst the rest of samples are used for testing. The experimental results are shown in the next subsection. All experiments are repeated 10 times, where the average metrics are reported for comparison. For hardware and software settings, all experiments are implemented on the MATLAB 2018b with a 16GB Intel i5-8400 CPU.

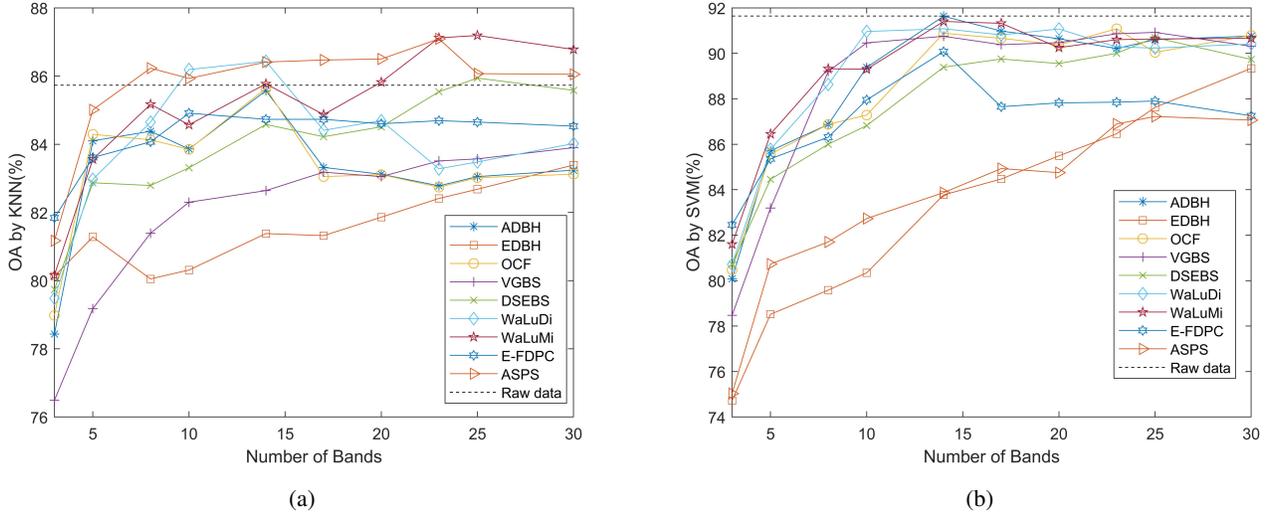


Fig. 5: OA curves on the PaviaU dataset with different UBS methods by using KNN (a) and SVM (b).

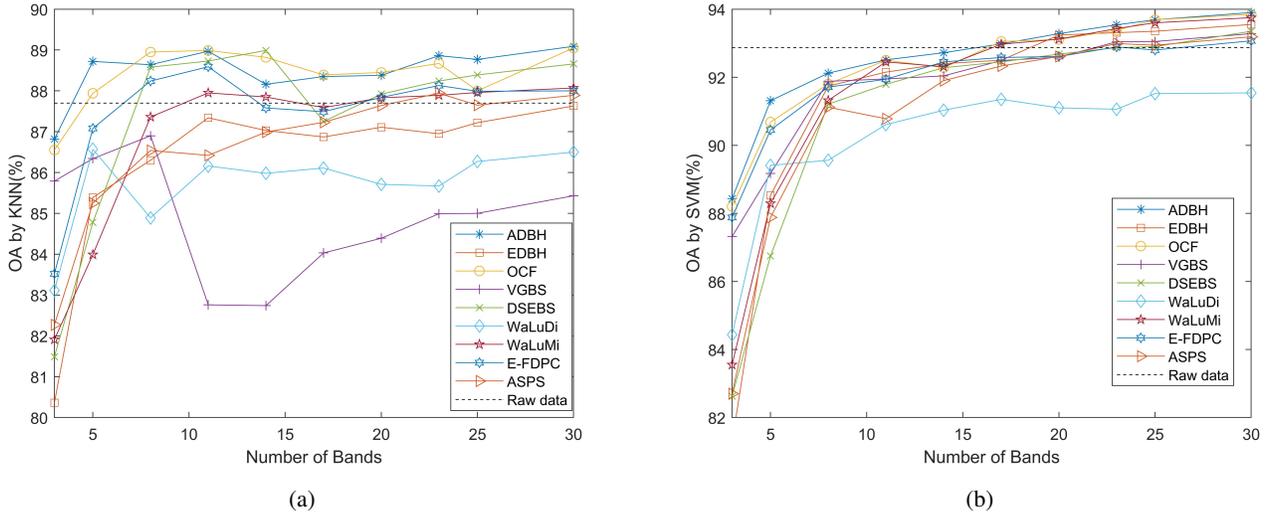


Fig. 6: OA curves on the Salinas dataset with different UBS methods by using KNN (a) and SVM (b).

C. Comparison experiments

In principle, the HSI classification results can be quantitatively evaluated by three common metrics from the confusion matrix, including the overall accuracy (OA), the average accuracy (AA) and the Kappa coefficient. The OA is the percentage of corrected classified pixels in total, and the AA reflects the mean classification accuracy over all classes. The Kappa coefficient is estimated for evaluating the reliability of the classification result. In this section, the compared results will be illustrated in two forms. Firstly, for all four HSI datasets, the OA curves are generated according to OAs against different chosen numbers of bands varying from 3 to 30. Also, we have compared the OA, AA and Kappa coefficient of different algorithms with certain determined numbers of bands. For the OA curves in most datasets, the performance of most approaches keep stable after the number of chosen band is around 10 to 15. Even when more bands are chosen, there is no

significant improvement for most of them. Therefore, detailed comparison with 14 selected bands on the four datasets, in terms of OA, AA and Kappa, is given in Tables I-IV. The best performance except the result with raw data are labelled bold.

Fig. 4 and Table I show the classification results for the Indian pines dataset. As seen in Fig. 4, our ADBH has the highest OA on the SVM classifier with 3 to 30 selected bands, which has also produced about the highest OA on the KNN classifier. Although the OA of ADBH on KNN is the second best when the number of chosen bands is no more than 20, it outperforms DSEBS after more bands are chosen. Despite of the best OA generated on the KNN classifier, DSEBS has quite poor performance on SVM, which shows a certain degree of lack of robustness or stability. The ASPS has a poor performance on both KNN and SVM classifiers. Table I actually shows as an example the classification results of all relevant methods with 14 selected bands. As can be seen, the

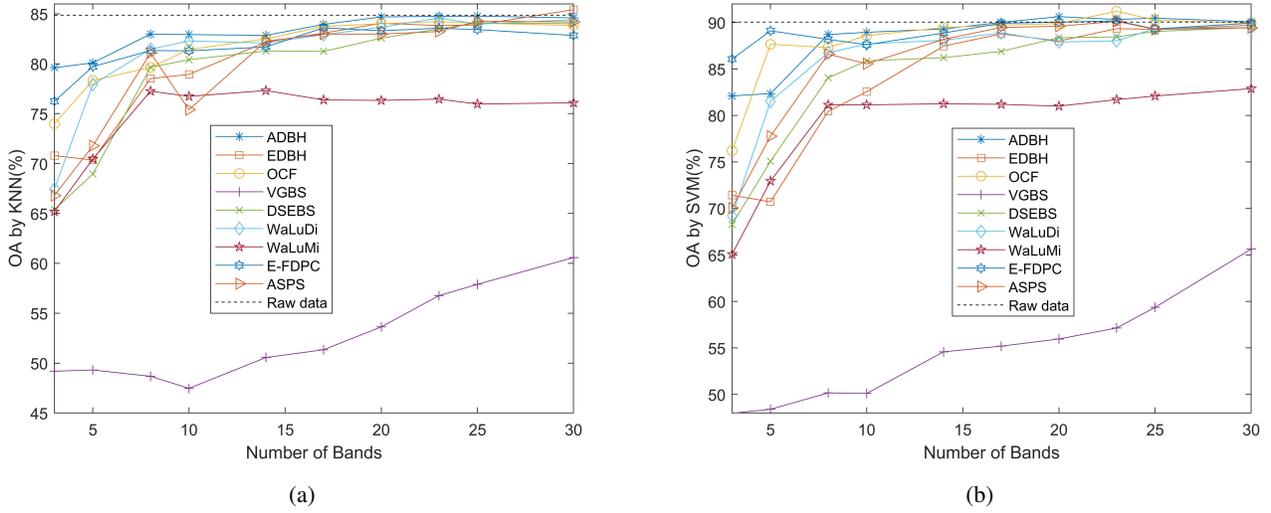


Fig. 7: OA curves on the KSC dataset with different UBS methods by using KNN (a) and SVM (b).

TABLE IV: Classification results from different approaches for the KSC dataset.

| Classifier | OCF | VGBS | DSEBS | WaLuDi | WaLuMi | E-FDPC | ASPS | EDBH | ADBH | Raw data |
|--------------|-------------------|------------|------------|------------|------------|------------|-------------------|-------------------|-------------------|------------|
| OA by KNN(%) | 82.49±0.01 | 50.56±0.02 | 81.28±0.01 | 82.07±0.01 | 77.31±0.01 | 81.68±0.01 | 82.26±0.01 | 82.18±0.01 | 82.84±0.01 | 85.86±0.01 |
| AA by KNN(%) | 74.59±0.01 | 37.08±0.03 | 71.30±0.01 | 67.32±0.01 | 73.08±0.01 | 73.45±0.01 | 74.33±0.01 | 74.88±0.01 | 73.85±0.01 | 79.15±0.01 |
| Kappa by KNN | 0.80±0.01 | 0.44±0.03 | 0.79±0.01 | 0.75±0.01 | 0.80±0.01 | 0.80±0.01 | 0.80±0.01 | 0.80±0.01 | 0.81±0.01 | 0.84±0.00 |
| OA by SVM(%) | 89.45±0.01 | 54.58±0.03 | 86.21±0.01 | 88.09±0.01 | 81.25±0.01 | 88.86±0.01 | 88.19±0.01 | 87.45±0.00 | 89.3±0.01 | 90.04±0.00 |
| AA by SVM(%) | 80.47±0.01 | 44.25±0.04 | 79.10±0.01 | 74.67±0.01 | 80.00±0.01 | 79.57±0.01 | 82.37±0.01 | 81.88±0.01 | 80.87±0.01 | 85.58±0.00 |
| Kappa by SVM | 0.88±0.01 | 0.48±0.01 | 0.85±0.01 | 0.79±0.01 | 0.87±0.01 | 0.87±0.01 | 0.86±0.01 | 0.86±0.00 | 0.88±0.01 | 0.89±0.01 |

proposed method has produced the best results in terms of OA, AA and Kappa on the SVM classifier, and the second best on the KNN classifier just after DSEBS. In addition, our ADBH framework can outperform the raw data without band selection when there are more than 10 selected bands on the KNN classifier or more than 30 on SVM, which further validates the superiority of the proposed approach.

Fig. 5 and Table II summarize the classification results of all methods on the PaviaU dataset. In Fig. 5 (a), with the KNN classifier, ASPS, WaLuDi and WaLuMi produce the best results with the number selected bands increasing from 5 to 30, and the proposed ADBH seems not ideal. However, the classification results from KNN only achieves about 87%, which is far less than those from SVM at nearly 92%. For the SVM classifier, the results from ADBH is among the best when 15 or more bands are selected, and other best ones include WaLuMi, WaLuDi and OCF. Surprisingly, ASPS and EDBH produce the worse results on SVM. Although WaLuMi, WaLuDi and OCF seem to produce the best results in this group of experiments, as shown in Fig. 4 and Table 1, they appear to among the worst with the Indian pines dataset on KNN and/or SVM classifiers under a certain range of selected bands. From Table II, we can find that WaLuDi, ASPS and WaLuMi have produced the best results with the KNN classifier with 14 selected bands. However, the results from ADBH is the best on the SVM classifier and outperform all these three approaches.

For the Salinas dataset, the related comparison is shown in

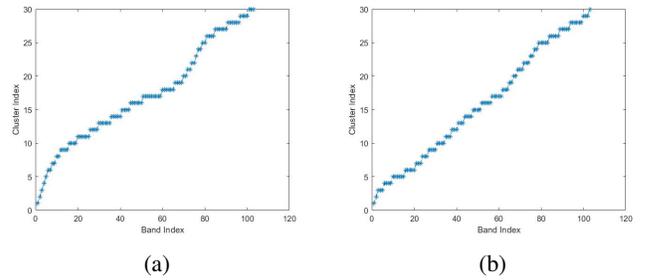


Fig. 8: The Clustering results (defined cluster number equals to 30).(a) ADBH, (b) OCF.

Fig. 6 and Table III. In Fig.6, ADBH algorithm achieves the most stable result on both classifiers, which is more robust than other methods. For the comparison between our ADBH method and the full dataset, it can be seen that the ADBH has an obvious advantage with the KNN classifier after 5 more bands are chosen. After more than 15 bands are chosen, the ADBH also achieves a better result than the raw dataset. The VGBS method does not perform well with the KNN classifier and the WaLuDi method has not achieved a good performance with the SVM classifier. Although the OCF method has the best performance with the KNN classifier when the number of chosen bands are around 10, its performance is not as robust as ours from any number of chosen bands. According to the Table III, the DSEBS is slightly better than ADBH on KNN while ADBH has the best performance on the SVM.

For the KSC dataset, ADBH has the best performance against all others with the KNN classifier. Although the E-FDPC method performs best with the SVM classifier when the number of selected bands is quite small, ADBH has better result after more bands are chosen. In general, our method has the best performance on both classifiers from Fig. 7, whilst the VGBS method performs quite poor. In Table IV, our ADBH algorithm has a satisfactory result when 14 bands are chosen with the KNN classifier. Although the OA of our ADBH method is not the best with the SVM, it achieves the second best with a small gap behind the first.

D. Extended discussions

In this subsection, extended analysis is carried out to compare the performance of different UBS methods over the four tested HSI datasets. Afterwards, the performance on the PaviaU dataset will be highlighted since our ADBH does not have the top robust performance on it. In addition, the computational time of each method will be compared to evaluate the efficiency of these UBS methods.

TABLE V: Number of parameters and computational time (s) of different UBS methods with 30 selected bands.

| Methods | No.Parm. | Indian pines | Pavia U | Salinas | KSC |
|---------|----------|--------------|-------------|-------------|-------------|
| ADBH | 0 | 0.35 | 2.69 | 2.23 | 5.91 |
| OCF | 0 | 0.7 | 0.65 | 1.13 | 1.66 |
| VGBS | 1 | 0.54 | 0.24 | 0.82 | 0.81 |
| DSEBS | 4 | 0.2 | 1.02 | 1.05 | 3.03 |
| WaLuDi | 4 | 41.95 | 99.7 | 198.51 | 408.68 |
| WaLuMi | 4 | 14.04 | 13.82 | 29.68 | 51.23 |
| E-FDPC | 1 | 0.97 | 6.85 | 3.11 | 27.97 |

From results of all compared methods, we can find that some methods have unstable performance on different datasets and different classifiers. For example, the WaLuMi method achieves better performance on the PaviaU dataset but ranks last on the Indian pines dataset. Moreover, the VGBS has the worst performance on the KSC dataset, especially the lower OA than those from other datasets. The ASPS has a robust performance with the KNN classifier in the PaviaU dataset, but its performance with the SVM is quite poor. From our point of view, this phenomenon may be explained by three reasons. Firstly, the four datasets are from two different HSI sensors, AVIRIS and ROSIS. The AVIRIS sensor seems to be noisy and usually heavily polluted, as seen in the KSC dataset. Therefore, poor results from some methods may indicate their lack of robustness in dealing with noisy datasets. Secondly, most of the compared methods have parameters which are usually set empirically. These fixed parameters may limit the stability of the associated algorithms when different datasets or different classifiers are applied. The inconsistency in performance is prevalent for most unsupervised methods when relying on different parameters. In addition, the UBS is an optimization task as we discussed before. If the algorithm focuses too much on local optimal solution, the instability will occur. However, thanks to the combination of AD and BH, our method can provide a parameter-free way for solving the inconsistency

problem and avoiding the effect of noisy bands. With the decent results of our method on all four datasets, especially the noisy KSC dataset, the robustness and stability of our proposed ADBH framework have been fully validated. We have noticed that the performance has been hugely improved with the utilization of AD, the poor performance of EDBH illustrates that the euclidean distance is not robust to the noisy band in our designed hierarchy. Besides, the proposed parameter-free framework can prevent from setting empirical parameters, where we compare the number of parameters in each method in Table V.

Although our method performs consistently well on four datasets, we have not achieved the best result on the PaviaU dataset, especially with the KNN classifier when more bands are chosen. According to Fig. 5, we have noticed that the OCF approach produces similar results to ours, especially after the desired number of band is above 15. Since both ADBH and OCF cluster bands into several groups and select the most significant one from each group, we conclude that this kind of strategy is sensitive to noisy bands when the desired number of bands is large. In Fig. 8, we can see that these two methods have clusters with only one band inside, where noisy bands have potential to be chosen. In Fig. 5, the OA curves of most approaches start to fall when 15 or more bands are chosen, which also infers that the proper number of selected bands might be around 15. More bands in the chosen subset may have few or even negative effect on the classification accuracy.

The computational complexity is a crucial issue for the efficiency of UBS algorithms. Hence, we compare this using the computational time of every method on various datasets on the same software/hardware platform. The Table V depicts the processing time of different methods when 30 bands are chosen on the four tested datasets. As seen in Table V, our ADBH framework has a fair computational time among all compared methods. Although the VGBS is the most efficient one, its performance on the classification is not good. For the WaLuMi and WaLuDi algorithms, their computational burdens are the heaviest, which reflects the drawback of the complicated distance measurement they used. Most of the existing band selection approaches fail to maintain the consistency of selected bands when the number of desired bands varies. As such, we aim to provide a band hierarchy to tackle this challenging problem. With the derived band hierarchy, any desired number of bands can be selected without re-running the whole process as most other approaches do, including OCF. Considering the fact that there is no prior information of the optimal number of bands for a given HSI dataset, in practice the process of band selection needs be repeated for quite a few times. As a result, the overall computational costs will be linearly accumulated for most other approaches. However, thanks to the ADBH, the overall computational cost of our approach remains almost unchanged as the additional costs in selecting different numbers of bands is minor and can be neglected. To this end, the computational cost of the proposed approach is actually far more efficient than conventional approaches including OCF. In summary, the proposed ADBH method seems to be a robust, effective and efficient solution for UBS of HSI.

V. CONCLUSION

During the last several years, various approaches have been proposed for the UBS of HSI. In this paper, we propose a band hierarchy clustering UBS framework for effective band selection. A flexible tree hierarchy-based algorithm ADBH is developed to explore the data structure within bands which can acquire any desired number of band subsets. To overcome the effect of noisy bands, we propose a novel adaptive distance AD, which is combined with our ADBH framework. Moreover, our approach is parameter-free hence easy for implementation. The satisfactory performance from experiments on four publicly available datasets have fully demonstrated the robustness and efficiency of our proposed ADBH method.

Although data classification is used to evaluate the efficacy of the selected bands, band selection can actually benefit many other applications of HSI, such as spectral unmixing, object detection and data visualisation et al [47]–[49]. However, applications in these fields are seldom selected, due mainly to the lack of available ground truth maps for quantitative evaluation. Further verification of the band selection approaches in these areas will be explored in the future.

ACKNOWLEDGMENT

The authors would also like to thank the authors of OCF, VGBS, DSEBS, WaLuDi/WaLuMi, E-FDPC, ASPS for providing the original codes.

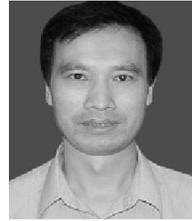
REFERENCES

- [1] J. Tschanerl, J. Ren, F. Jack, J. Krause, H. Zhao, W. Huang and S. Marshall, "Potential of UV and SWIR hyperspectral imaging for determination of levels of phenolic flavour compounds in peated barley malt," *Food. Chem.*, vol. 270, pp. 105-112, Jan. 2019.
- [2] J. Tschanerl, J. Ren, H. Zhao, F. Kao, S. Marshall and P. Yuen, "Hyperspectral image reconstruction using Multi-colour and Time-multiplexed LED illumination," *Opt. Laser. Eng.*, vol 121, pp. 352-357, Oct. 2019.
- [3] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol.47, no. 4, pp. 884-897, Apr. 2017.
- [4] Z. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial network and conditional random fields for hyperspectral image classification," *IEEE Trans. Cybern.*, to be published. doi: 10.1109/TCYB.2019.2915094.
- [5] H. Li, G. Xiao, T. Xia, Y. Y. Tang, and L. Li, "Hyperspectral image classification using functional data analysis," *IEEE Trans. Cybern.*, vol.44, no. 9, pp. 1544-1555, Sept. 2014.
- [6] Y. Zhou, and Y. Wei, "Learning hierarchical spectral-spatial features for hyperspectral image classification," *IEEE Trans. Cybern.*, vol.46, no. 7, pp. 1667-1678, Jul. 2016.
- [7] J. Zabalza, J. Ren, Z. Wang, S. Marshall and J. Wang, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1886-1890, Nov. 2014.
- [8] C. Zhao, X. Li, J. Ren and S. Marshall, "Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery," *Int J Remote Sens.*, vol. 34 no. 24, pp. 8669-8684, Dec. 2013.
- [9] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybern.*, to be published. doi: 10.1109/TCYB.2019.2951572.
- [10] Y. Chen, W. He, N. Yokoya, and T-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, to be published. doi: 10.1109/TCYB.2019.2936042.
- [11] J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li and S. Marshall, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in Hyperspectral Imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418-4433, Aug. 2015.
- [12] J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang, P. Du and S. Marshall, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging", *Neurocomputing*, Vol. 185, pp. 1-10, 2016.
- [13] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for Hyperspectral Images," *IEEE Trans. Cybern.*, vol.48, no. 1, pp. 16-28, Jan. 2018.
- [14] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55-63, Jan. 1968.
- [15] J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 112-122, Jul. 2014.
- [16] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140-7151, Dec. 2017.
- [17] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586-1600, Jun. 2006.
- [18] T. Qiao, J. Ren, Z. Wang, J. Zabalza, M. Sun, H. Zhao, S. Li, J. A. Benediktsson, Q. Dai, S. Marshall, "Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 119-133, Jan. 2017.
- [19] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning," *IEEE Trans. Cybern.*, to be published. doi:10.1109/TCYB.2019.2905793.
- [20] A. A. Green, M. Berman, P. Switzer and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65-74, Jan. 1988.
- [21] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138-142, Jan. 2011.
- [22] X. Cao, T. Xiong and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 329-333, Mar. 2016.
- [23] S. Patra, P. Modi and L. Bruzzone, "Hyperspectral band selection based on rough set," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495-5503, Oct. 2015.
- [24] C. I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631-2641, Nov. 1999.
- [25] C. I. Chang, and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575-1585, Jun. 2006.
- [26] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158-4171, Dec. 2007.
- [27] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1431-1445, Mar. 2016.
- [28] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492-1496, Jun. 2014.
- [29] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88-102, Jan. 2016.
- [30] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910-5922, Oct. 2018.
- [31] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published. doi:10.1109/JSTARS.2019.2941454.
- [32] 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.*, vol. 52, no. 11, pp. 7111-7119, Nov. 2014.
- [33] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631-644, Feb. 2015.
- [34] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 227-239, Jan. 2016.

- [35] J. Tschanneral, 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," *Inform Fusion.*, vol. 51, pp. 189-200, Jan. 2019.
- [36] Y. Yuan, X. Zhang, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 51-64, Jan. 2017.
- [37] F. Hong, H. Peng, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 1226-1338, Aug. 2005.
- [38] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no 5, pp. 2956-2969, May. 2015.
- [39] C.-I. Chang, S. Wang, K.-H. Liu, M.-L. Chang, and C. Lin, "Progressive band dimensionality expansion and reduction via band prioritization for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 591-614, Sep. 2010.
- [40] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 236-244, 1963.
- [41] X. Wei, Q. Yang, Y. Gong, N. Ahuja and M. Yang, "Supapixel Hierarchy," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4838-4849, Oct. 2018.
- [42] D. B. West *et al.*, *Introduction to Graph Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2001.
- [43] Z. Wang, Z. Yu, C. L. P. Chen, J. You, T. Gu, H-S. Wong, and J. Zhang, "Clustering by Local Gravitation," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1383-1396, May. 2018.
- [44] Y. Qian, F. Li, J. Liang, B. Liu and C. Dang, "Space structure and clustering of categorical data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 10, pp. 2047-2059, Oct. 2016.
- [45] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099-4109, Oct. 2010.
- [46] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, Aug. 2004.
- [47] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22-40, Jun. 2016.
- [48] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf.Sci.*, vol. 485, pp. 154-169, Jun. 2019.
- [49] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Compression of hyperspectral remote sensing images by tensor approach," *Neurocomputing*, vol. 147, pp. 358-363, 2015.



research interests focus mainly on hyperspectral imaging, computer vision, machine learning and big data analytics.

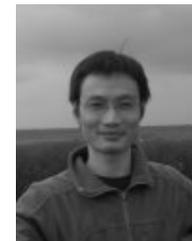


Jinchang Ren (M'05, SM'17) is currently a Reader with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K., and he is also a joint Professor with the School of Computer Sciences, Guangdong Polytechnic University, Guangzhou, China. He received his B.Eng., M.Eng. and D.Eng. all from the Northwestern Polytechnical University (NWPU), Xi'an, China. He also received a Ph.D. degree from the University of Bradford, Bradford, U.K. He has published about 300 peer-review journal and conference papers. His research interests focus mainly on hyperspectral imaging, computer vision, machine learning and big data analytics.

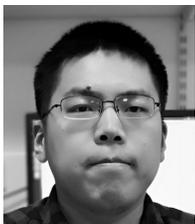
Huimin Zhao is currently a professor and Dean of the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China. He received the B.Sc. and the M.Sc. degrees in signal processing from Northwestern Polytechnical University, Xian, China, in 1992 and 1997, respectively, and the Ph.D. degree in electrical engineering from the Sun Yat-sen University, Guangzhou, China, 2001. His research interests include image and video analysis, and information security technology.



Genyun Sun (M'16) is currently a full professor with the School of Geoscience, China University of Petroleum (East China), Qingdao, China. He received the B.S. degree from Wuhan University, Wuhan, China, in 2003, and the Ph.D. degree from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2008. His research interests include remote sensing image processing, hyperspectral remote sensing, high resolution remote sensing, and intelligent optimization.



Wenzhi Liao (M'14, SM'16) is with the Flemish Institute for Technological Research (VITO), Belgium. He received the PhD Degree in Computer Science Engineering from Ghent University (Belgium) in 2012. His research interests include pattern recognition, remote sensing, and image processing, mathematical morphology, multitask feature learning, multisensor data fusion, and hyperspectral image restoration.



He Sun is currently working towards a PhD in Electronic and Electrical Engineering at the University of Strathclyde. His research interests focus mainly on computer vision and machine learning, especially the hyperspectral imaging.



Zhenyu Fang is currently working towards a PhD in Electronic and Electrical Engineering at the University of Strathclyde. His main interests are algorithm development for image classification, object detection and face detection.



Jaime Zabalza is currently a Research Associate with the Department of Electronic and Electrical Engineering, University of Strathclyde. He received a PhD in Hyperspectral imagin from the University of Strathclyde, Glasgow, U.K. His PhD work, supervised under Dr J Ren, was awarded the best PhD thesis from IET (Image and Vision Section) in 2016. His interests are related to SAR and remote sensing, hyperspectral imaging and DSP, including signal processing in a wide range of applications.