International Journal of Wavelets, Multiresolution and Information Processing

A Novel Semi-supervised Learning Framework for Hyperspectral Image Classification

Zhijing Ye*, Hong Li†, Yalong Song‡
School of Mathematics and Statistics, Huazhong University of Science and Technology
Wuhan 430074, P. R. China
**xkinghust@163.com*
†*hongli@hust.edu.cn*
‡*ylsong@hust.edu.cn*

Jianzhong Wang
Department of Mathematics and Statistics, Sam Houston State University
Huntsville, Texas 77341, USA
jzwang@shsu.edu

Jon Atli Benediktsson
Faculty of Electrical and Computer Engineering, University of Iceland
107 Reykjavik, Iceland
benedikt@hi.is

In this paper, we propose a novel semi-supervised learning classification framework using box-based smooth ordering and Multiple 1D-embedding-based interpolation method in Ref. 25 for hyperspectral images. Due to the lack of labeled samples, conventional supervised approaches cannot generally perform efficient enough. On the other hand, obtaining labeled samples for hyperspectral image classification is difficult, expensive and time-consuming, while unlabeled samples are easily available. The proposed method can effectively overcome the lack of labeled samples by introducing new labeled samples from unlabeled samples in a label boosting framework. Furthermore, the proposed method uses of spatial information from the pixels in the neighborhood of the current pixel to better catch the features of hyperspectral image. The proposed idea is that, first, we extract the box (cube data) of each pixel from its neighborhood, then apply multiple 1D-embedding interpolation to construct the classifier. Experimental results on three hyperspectral data sets demonstrate the proposed method is efficient, and outperforms recent popular semi-supervised methods in terms of accuracies.

Keywords: Hyperspectral image classification; semi-supervised learning; boxed-based data smooth sorting; multiple 1D-embedding-based interpolation.

AMS Subject Classification: 68U10, 62H35

1. Introduction

Hyperspectral images can provide remote sensing analysts with very rich spatial, spectral, and temporal information.[7] Although the rich information is potentially

utilized for classification, several critical issues need to be addressed: 1) The small amount of available labeled samples; 2) the high dimensionality of hyperspectral data; 3) the spatial variability of spectral signatures; and 4) the high cost of sample labeling,[26] where the high dimensionality of hyperspectral data and small amount of available labeled samples pose the curse of dimensionality (Hughes phenomenon).[15] It is a very difficult task for supervised approaches due to the lack of labeled samples. In practice obtaining labeled samples is very difficult, expensive and time-consuming, while unlabeled samples are easily available. This observation has fostered the idea of adopting semi-supervised learning techniques in hyperspectral image classification. The main assumption of such techniques is that new labeled samples (the confident set) can be obtained from unlabeled samples in a self-learning framework without significant effort/cost.[10]

Recently, research on semi-supervised learning has been very active and has attracted many researchers devoting to hyperspectral image classification. Some semi-supervised methods have been presented to deal with hyperspectral image classification due to the lack of labeled samples. These semi-supervised methods mainly include two classes: generative models and discriminative models.[14] One of the popular generative models is the Expectation Maximization (EM) algorithm with finite-mixture model.[23] Compared with generative models, the discriminative models are more attractive, such as TSVM,[4] Graph-based methods,[5] LapSVM,[1] SSNNs,[22] MLR+AL,[19] and recent state-of-the-art Joint+Laplacian,[26] etc. All these methods have demonstrated a good performance in hyperspectral image classification in terms of accuracies. However, most of them are based on the assumption that similar spectral samples should share the same label. In fact, different spectral samples may share the same label, and similar spectral samples may also share different labels. Thus, the spatial information can be used for reducing the labeling uncertainty that exists when only spectral information is taken into account.[13]

So far, most hyperspectral image processing methods are based on single pixels. Conventional methods consider the image as an ensemble of spectral measurements without exploiting the spatial structure information.[12] However, hyperspectral images usually have large homogeneous regions where neighboring pixels within the regions consist of the same type of materials (same class).[8] In order to make use of the spatial organization, some spectral-spatial methods[2,3,6,17] are presented to improve the classification performance. However, these methods are also based on single pixels. In this paper, we proposed a box-based method for exploiting spatial information. The idea is that a fixed-size neighbourhood of each pixel is extracted as a basic processing unit. Each unit is a small cubic of data, called a box. The box can extract the local spatial information. The spatial smoothness assumption holds well for homogenous regions in hyperspectral images, which ensures the effectiveness of the box-based method.

In this paper, we propose a novel semi-supervised learning classification framework using box-based smooth-ordering and Multiple 1D-embedding -based interpo-

lation method[24,25] for hyperspectral image. The proposed method can effectively overcome the lack of labeled samples. Even in case of very few labeled samples, it can also exhibit a good performance in terms of accuracies. Different from general semi-supervised methods, the proposed method extracts a box-based feature, not a pixel-based feature. Compared with the pixel-based feature, the box-based feature can exploit the spatial arrangement of each pixel to reduce the labeling uncertainty of the pixel-based feature. One of the core idea of proposed framework is box-based smooth ordering, i.e., after extracting the box of each pixel, we order the boxes according to the proximity in their neighbourhoods. The box-based smooth ordering method is based on the assumption that proximity between two boxes implies proximity between their center pixels. The box-based permutation corresponds to a new 1D signal defined by the proximity in a 1D coordinate system, which is an isometric 1D embedding of the original data. Thus, the proposed method can overcome the high dimensionality of hyperspectral data. Then, we apply the multiple 1D-embedding-based interpolation on 1D signals to produce a new confident labeled subset of unlabeled samples, which is called a *newborn set* or *confident set*,[24,25] which is added to the labeled set. Repeating the above process till the updated labeled set reaches a certain size. Finally, we classify the remaining unlabeled samples by a classifier built on the nearest multi 1D-embedding. This part of the algorithm is very similar to M1DEI algorithm in Ref. 25, which provides moredetailed description for readers. In conclusion, the proposed framework provides the following contributions:

(1) **Box-based feature:** The proposed method is based on a box-based feature, while conventional hyperspectral image processing methods are based on a single pixel. The idea of box-based feature origins from patch-based processing in natural image (binary image, gray image and RGB image).[21] The boxed-based processing is a new technology in hyperspectral image. The box-based feature can extract the local spatial structure effectively. Due to many large homogeneous regions existing in hyperspectral image, the box-based feature can reduce the labeling uncertainty of a single pixel by exploiting the the local spatial structure information.
(2) **Box-based smooth ordering:** In fact, box ordering is a process to reorganized the high-dimensional data, giving the data a 1D representation while preserving the (local) proximity. Several version of such embedding at random provide a multiple 1D-embedding, which is the base for construct the proposed classifier classifiers by W1dEI.[25]
(3) **Label Boosting:** In this paper, we do not use the general classifiers, such as KNN,[20] SVM,[16],[11] MLR,[18] etc. We determine the class labels of unlabeled samples by a label boosting sub-algorithm in M1DEI. The boosting technique overcomes the difficulty in the classification of the data set with a small labeled subset.

The remainder of the paper is organized as follows. Section 2 describes the proposed semi-supervised learning classification framework and the related methodology. Furthermore, we give the algorithm of the proposed method. Section 3 shows the classification results on three hyperspectral image data sets and the comparisons between the proposed method and recent popular semi-supervised methods. Finally, Section 4 concludes this paper with some remarks.

2. Semi-supervised classification framework

First, we briefly define the notations used in this paper. Consider a hyperspectral image $\boldsymbol{X} \in R^{n\times m}$ of n pixels $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, in which $\boldsymbol{x}_i$ is an m-dimensional spectral vector and y_i is the label in the class label set $\boldsymbol{\Omega} = \{1, 2, \ldots, C\}$. $\boldsymbol{A}_i$ is a small cube data of the size $b \times b \times m$ (b is an odd number) centered at $\boldsymbol{x}_i$. Let each pixel correspond to a box. Thus there are n boxes in this hyperspectral image, which form an initial sequential box set $\boldsymbol{\Lambda} = \{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n\}$. In this paper, we consider reordering the initial sequential box set $\boldsymbol{\Lambda}$ based on the distance defined as follows:

Definition 2.1. *Let $\boldsymbol{A}_i, \boldsymbol{A}_j \in \boldsymbol{\Lambda}$ be any two boxes of size $b \times b \times m$ in hyperspectral image. The distance between boxes $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$ is defined as the average of the l_2-norm of the distances between corresponding pixels in $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$:*

$$Dis(\boldsymbol{A}_i, \boldsymbol{A}_j) = \frac{1}{b^2}\sum_{u=1}^{b}\sum_{v=1}^{b}\|\boldsymbol{A}_i(u, v, \cdot) - \boldsymbol{A}_j(u, v, \cdot)\|_2. \tag{2.1}$$

Our goal is to smoothly sort the boxes as a 1D sequence such that the boxes of similar structure are ordered in nearby positions.

2.1. *Box-based smooth ordering*

In this subsection, we wish to design a simple ordering operator $\boldsymbol{P}$ that is applied to the box set $\boldsymbol{\Lambda}$ to obtain a smooth 1D signal. Let $\boldsymbol{T} = \{t_1, t_2, \ldots, t_n\}$ is the index set of the new box-based permutation $\boldsymbol{\Lambda}_{\boldsymbol{T}} = \{\boldsymbol{A}_{t_1}, \boldsymbol{A}_{t_2}, \ldots, \boldsymbol{A}_{t_n}\}$. Thus, we have

$$\boldsymbol{A}_{t_i} = \boldsymbol{P}(\boldsymbol{A}_i), i = 1, 2, \ldots, n. \tag{2.2}$$

Obviously, the new permutation is related to the initial box. Different initial boxes correspond to different permutations. In this paper, the initial box is randomly chosen in each permutation. Based on new box-based permutation and the distance measure defined in (2.1), the proximity can be computed between two neighbouring boxes as following:

$$d_i = \begin{cases} 0 & i = 1 \\ Dis(\boldsymbol{A}_{t_{(i-1)}}, \boldsymbol{A}_{t_i}) & i = 2, 3, \ldots, n. \end{cases}. \tag{2.3}$$

Further, we define the path distance from the origin to some point as:

$$D_i = \sum_{j=1}^{i} d_j, i = 1, 2, \ldots, n. \tag{2.4}$$

Thus, according to the simple ordering operator $\boldsymbol{P}$ and distance measure Dis, we can find a projection g that maps the boxes set $\boldsymbol{\Lambda_T}$ into $\boldsymbol{D} = \{D_i\}_{i=1}^n$, i.e.

$$g(\boldsymbol{A}_{t_i}) = D_i, i = 1, 2, \ldots, n, \tag{2.5}$$

where g is called as box-based ordering projection operator, a composition operator of $\boldsymbol{P}$ and Dis. The projection g is also related to the initial box. Besides, $\boldsymbol{D}$ can be regarded as a sequential points set in a 1D coordinate system corresponding to the boxes set in the new permutation, where $D_1 = 0$ means that the initial box corresponds to the origin, and D_n denotes the total path distance of new permutation. Here, we call $\boldsymbol{D}$ as a 1D signal. The smoothness of this 1D signal can be measured by using D_n. So in order to improve the 1D signal smoothness, we need to minimize D_n. This can be regarded as the shortest path problem[9] of the box set $\boldsymbol{\Lambda}$ under the distance measure (2.1).

In general, the above problem can be solved by well-known algorithms like the Dijkstra algorithm.[9] However, in hyperspectral image analysis, due to large amounts of samples, such a solution becomes very difficult and time-consuming. On the other hand, a hyperspectral image usually has large homogeneous regions. Thus, we can take the approximation method by taking full advantage of spatial structure information, that is, in the ordering process, we first choose an initial box randomly in the boxes set $\boldsymbol{\Lambda}$ as the origin. Then, we search its next box in the neighbourhood of each box instead of in the whole image. It needs to be noted that each box should only appear once in the new permutation. As for the choice of next box at each step, the nearest one under the distance measure defined in (2.1) may not be the best choice for the whole permutation. Consequently, the front part of the permutation is very smooth, i.e., the proximity is small, while the latter part becomes rough, i.e., the proximity is large. In order to balance the difference and ensure that the whole permutation is more smooth, we first find the nearest box and the second nearest one, then we choose one of both according to a random rule, as shown in **Algorithm 1**. In the search process, if only one box remains in the neighborhood, we choose the box that is next in the permutation. If no box remains in the neighborhood, we find the nearest box and the second nearest one in the whole image. At each search step, we record the proximity between each box and its next box. This ordering method is very similar to that in Ref. 21. The box-based smooth ordering algorithm is summarized in **Algorithm 1**:

Remark: ϵ in (2.6) is a design parameter. Obviously, $Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i_1}}) \leq Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i_2}})$. Thus, we have $q_i \in [0.5, 1]$. In (2.6), it seems that next optimal box is randomly chosen between the nearest box and the second nearest box. But the strategy tends to the second nearest box when the difference of both is very small ($q_i \to 0.5$), while the nearest one when the difference is very large ($q_i \to 1$).

Algorithm 1 Box-based smooth ordering.

Input: the initial sequential box set $\mathbf{\Lambda}$, the distance measure Dis defined in (2.1), a random probability vector $\boldsymbol{p} = \{p_1, p_2, \cdots, p_n\} \subseteq [0.5, 1]$, and the search neighborhood size $B \times B$, $d_1 = D_1 = 0$.

Choose any box $\boldsymbol{A}_{t_1}$ as the origin and set $T(1) = \{t_1\}$.
for $i = 1 : n-1$ **do**
  set $\Omega(i)$ is the size $B \times B$ neighborhood of $\boldsymbol{A}_{t_i}$ in the image, and $\Omega^c(i)$ is the remaining boxes in the neighborhood of $\boldsymbol{A}_{t_i}$, which don't join in the new permutation.
  if $|\Omega^c(i)| = 1$ **then**
    set the only box in $\Omega^c(i)$ as $\boldsymbol{A}_{t_{i+1}}$, and $T(i+1) = \{t_{i+1}\}$. Compute $d_{i+1} = Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i+1}})$ and $D_{i+1} = D_i + d_{i+1}$.
  else
    if $|\Omega^c(i)| \geq 2$ **then**
      Find the the nearest box $\boldsymbol{A}_{t_{i_1}}$ and the second nearest one $\boldsymbol{A}_{t_{i_2}}$ in $\Omega^c(i)$.
    end if
    if $|\Omega^c(i)| = 0$ **then**
      Find the the nearest box $\boldsymbol{A}_{t_{i_1}}$ and the second nearest one $\boldsymbol{A}_{t_{i_2}}$ in the whole image, where $t_{i_1}, t_{i_2} \nsubseteq T$.
    end if
    First compute the proximity $Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i_1}})$ and $Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i_2}})$ using the distance measure in (2.1). Then compute q_i using the following formula:

$$q_i = \frac{1}{1 + exp(\frac{Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i_1}}) - Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i_2}})}{\epsilon})}, \tag{2.6}$$

    Set

$$t_{i+1} = \begin{cases} t_{i_1} & q_i > p_i \\ t_{i_2} & otherwise. \end{cases}. \tag{2.7}$$

    Finally, set $T(i+1) = \{t_{i+1}\}$, and compute $d_{i+1} = Dis(\boldsymbol{A}_{t_i}, \boldsymbol{A}_{t_{i+1}})$ and $D_{i+1} = D_i + d_{i+1}$.
  end if
end for

Output: the ordering index set T and 1D signal $\boldsymbol{D} = \{D_1, D_2, \cdots, D_n\}$.

This is very reasonable. We do not have to choose the second nearest one for balancing the difference when the difference of both is very large. Thus, the nearest one is the best choice in this extreme case.

2.2. *1D interpolation*

We adopt the method proposed in Ref. 25 to construct our classifier. According to the ordering index set $\boldsymbol{T}$ and 1D signal $\boldsymbol{D}$, we consider classifying all unlabeled samples using 1D interpolation. Let $\{\boldsymbol{x}_{j_i}, y_{j_i}\}_{i=1}^{l}$ is the training set including l labeled samples ($l << n$), which is a small subset of $\boldsymbol{X}$. The corresponding boxes set $\{\boldsymbol{A}_{j_i}\}_{i=1}^{l}$ is also a small subset of $\boldsymbol{\Lambda}$. Thus, $\{j_1, j_2, \cdots, j_l\} \subseteq \boldsymbol{T}$. Since $\{j_1, j_2, \cdots, j_l\}$ are the elements of $\boldsymbol{T}$, we can find their positions in $\boldsymbol{T}$ and the corresponding points $\{D_{j'_1}, D_{j'_2}, \cdots, D_{j'_l}\}$ in 1D signal $\boldsymbol{D}$. By Combining $\{D_{j'_1}, D_{j'_2}, \cdots, D_{j'_l}\}$ with the label set $\{y_{j_1}, y_{j_2}, \cdots, y_{j_l}\}$, we can construct l interpolation nodes $\{D_{j'_i}, y_{j_i}\}_{i=1}^{l}$. Next, we give the definition of 1D interpolation as follows:

Definition 2.2. *Let $\{x_i\}_{i=1}^{n}$ be n points in $\boldsymbol{R}$, $\{y_i\}_{i=1}^{n}$ be the corresponding function value. Then, we can construct an analytic function f through these points, i.e.,*

$$y_i = f(x_i), i = 1, 2, \cdots, n, \tag{2.8}$$

and compute the function value of all points among $\{x_i\}_{i=1}^{n}$. Such a process is called 1D interpolation.

On the other hand, hyperspectral image classification is usually a multi-class classification problem. We decompose this multi-class problem into multiple binary class problems with one against all strategy. Thus, for any class $c \in \boldsymbol{\Omega}$, we make the following processing:

$$\tilde{y}_{j_i} = \begin{cases} 1 & y_{j_i} = c \\ -1 & otherwise. \end{cases} i = 1, 2, \cdots, l. \tag{2.9}$$

Replacing y_{j_i} with $\tilde{y}_{j_i}$, we can get l new interpolation nodes $\{D_{j'_i}, \tilde{y}_{j_i}\}_{i=1}^{l}$. According to **Definition 2.2** and the interpolation nodes $\{D_{j'_i}, \tilde{y}_{j_i}\}_{i=1}^{l}$, we can find an analytic function f that makes

$$\tilde{y}_{j_i} = f(D_{j'_i}), i = 1, 2, \cdots, l, \tag{2.10}$$

hold. Further, we can compute all function values f in the set $\boldsymbol{D}$, then map them to $\{-1, 1\}$ according to the sign function. We give the 1D interpolation algorithm in **Algorithm 2**:

Remark: The classification results for one time interpolation are obviously not very accurate. 1D interpolation that is performed once is regarded as a weak classifier. In order to improve the classification accuracy, we take the mechanism like AdaBoost, that is, constructing a strong classifier by training some weak classifiers for the same training set.

Due to the randomness of the initial box in the new permutation, we easily obtain some permutations by choosing randomly the initial box for K times without a repetition in **Algorithm 1**. Then, we apply 1D interpolation to each permutation for getting K classification results of each unlabeled samples. If the classification results of a sample in each permutation are nearly the same, the sample is called

Algorithm 2 1D interpolation.

Input: The samples set $\boldsymbol{X}$, the class label set $\boldsymbol{\Omega} = \{1, 2, \ldots, C\}$, the initial sequential box set $\boldsymbol{\Lambda}$, the labeled samples set $\{\boldsymbol{x}_{j_i}, y_{j_i}\}_{i=1}^{l}$ and the corresponding boxes set $\{\boldsymbol{A}_{j_i}\}_{i=1}^{l}$, the ordering index set $\boldsymbol{T}^k$ and the corresponding 1D signal $\boldsymbol{D}^k (k = 1, 2, \cdots, K)$.

1: **for** $k = 1 : K$ **do**

2: Find $\{D^k_{j'_1}, D^k_{j'_2}, \cdots, D^k_{j'_l}\}$ in 1D signal $\boldsymbol{D}^k$ according to the positions of labeled samples index set $\{j_1, j_2, \cdots, j_l\}$ in $\boldsymbol{T}^k$, and construct the interpolation nodes $\{D^k_{j'_i}, y_{j_i}\}_{i=1}^{l}$.

3: **for** $c = 1 : C$ **do**

4: Construct the new interpolation nodes $\{D^k_{j'_i}, \tilde{y}_{j_i}\}_{i=1}^{l}$ using (2.9).

5: Find an analytic function f through these nodes, and compute all function values $f(\boldsymbol{D}^k)$.

6: Set $f(\boldsymbol{D}^k) \to \{-1, 1\}$ according to the sign function, then classify preliminary each sample $\boldsymbol{x}_i$ corresponding to the point in $\boldsymbol{D}^k$ as

$$y_i^c = \begin{cases} c & f(\boldsymbol{x}_i) = 1 \\ 0 & otherwise. \end{cases} \quad i = 1, 2, \cdots, n. \tag{2.11}$$

7: **end for**

8: Count the classification label number of each sample $\boldsymbol{x}_i$, then classify these samples with only one label (except 0) as the corresponding class, others as 0 according to the label matrix $\{y_i^c\}$.

9: **end for**

10: Find these samples $\{\boldsymbol{x}_i^c\}_{i=1}^{n_c}$, where $\boldsymbol{x}_i^c$ in each permutation nearly shares the same class c $(c = 1, 2, \cdots, C)$.

Output: The confident set $\{\boldsymbol{x}_i^c, c\}_{i=1}^{n_c} (c = 1, 2, \cdots, C)$.

a confident sample. The set consisting of these confident samples is called as the confident set. The confident set is used for updating the labeled set.

2.3. *Semi-supervised hyperspectral image classification*

Due to the lack of labeled samples in hyperspectral image analysis, conventional supervised approaches can not generally provide sufficient accuracies in classification. Although obtaining labeled samples for hyperspectral image classification is difficult, expensive and time-consuming, the unlabeled samples can be collected easily. In this paper, we apply the semi-supervised framework[25] for hyperspectral image classification based on a box feature, box-based smooth ordering and 1D interpolation. For each pixel, we first extract its box feature. Then we order the boxes according to **Algorithm 1**. After obtaining both the new permutations and the training set, we classify all the test samples, and potentially obtain a confident set according to **Algorithm 2**. Next, we add the confident set into the training set, and

repeat the training process until no new confident set (except for an empty set) to add. Finally, we continue repeating the training process, and classify the remaining test samples by majority voting. We summarize the semi-supervised hyperspectral image classification framework in the following flow chart (see Fig. 1):

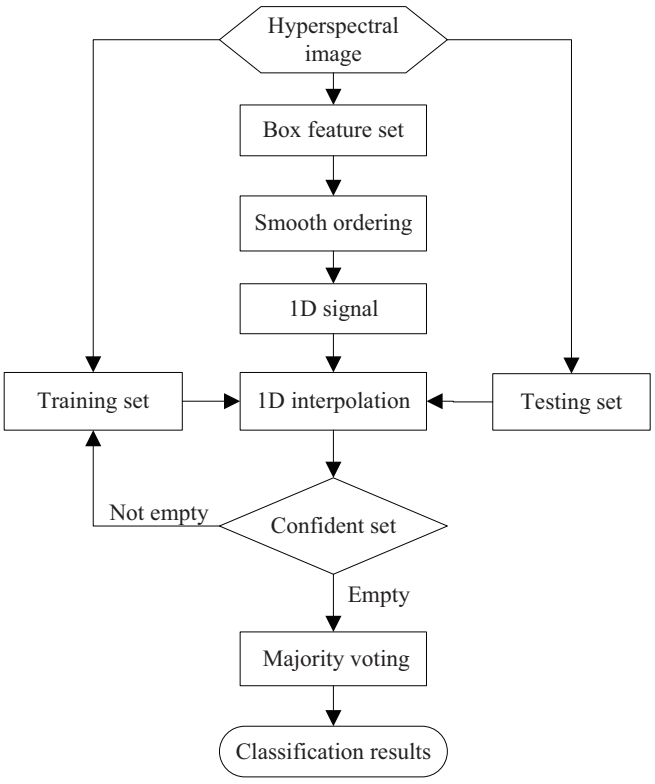


Fig. 1. The flow chart of semi-supervised hyperspectral image classification algorithm

3. Experiment results

The proposed framework is applied to three hyperspectral images: 1) AVIRIS Indian Pines scene, 2) AVIRIS Salinas scene and 3) ROSIS Pavia University scene. These data sets are available online from `http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`.

3.1. *Experimental Design and Parameter Selection*

In our experiments, we compare the proposed method with some popular semi-supervised methods, including LapSVM,[1] MLR+AL[19] and Joint+Laplacian.[26] In

order to ensure the comparison fair, we try our best to follow the settings in their original paper and choose the optimal parameters in these popular methods. As for the proposed method, there are three important parameters that need to tune. One is the box size b, which varies between 3×3 and 11×11 in an interval of two units. Another is the search neighborhood size B, which is related to b. The rest is the ordering times K, which directly affects the classification accuracy and its stability. We choose the linear method in 1D interpolation for efficiency. With $b = 1$ a pixel-based method is obtained, which is very different from the proposed method. We also compare the proposed method (utilizing the spatial structure and spectral information) with the pixel-based method (only the spectral information), and demonstrate that the spatial information can effectively improve the classification performance.

In order to show that the proposed semi-supervised method can effectively overcome the lack of labeled samples, we choose randomly five labeled samples for per class as the training set, the rest samples are regarded as the test set for all data sets. In order to ensure generality, the classification results including overall accuracies (OA [%]) and Kappa statistic (κ [%]) in our experiments are obtained by taking the mean of ten independent Monte Carlo runs.

3.2. *AVIRIS Indian Pines scene*

The scene was gathered by the AVIRIS sensor over the Indian Pines test site in North-western Indiana in 1992, and consists of 145×145 pixels and 16 classes ranging from 20 to 2455 in size, see Fig. 3 (a) for further details. In this image, 200 spectral bands are remained by removing 24 noisy bands covering the region of water absorption from 224 spectral reflectance bands. The scene represents a very challenging land-cover classification scenario. Since it is taken in June, some of the primary crops present, corn and soybeans, are in early stages of growth with less than 5% coverage. So discriminating among the primary crops can be very difficult (in particular, given the moderate spatial resolution of 20m).

We first evaluate the impact of the box size b and the corresponding search neighborhood size B on the classification results in the AVIRIS Indian Pines data set (see Table 1). The classification performance achieved the optimal performance with a box size $b = 5$ and search neighborhood size of $B = 5$. Furthermore it is observed that the classification accuracy first increases with the box size increasing, but decreases after achieving the maximum value. This demonstrates that the optimal box size is related to the spatial structure of the hyperspectral image. As for the search neighborhood size, it is affected by the box size. When the box size is no more than 5, the optimal search neighborhood size is 5. However, when the box size is larger than 5, the search neighborhood size varies with the box size. In the following experiments, we, respectively, choose the optimal values corresponding to the box sizes.

In order to explore the influence of the ordering times, we give the classifica-

Table 1. THE OA AND κ FOR DIFFERENT BOX SIZES AND SEARCH NEIGHBORHOOD SIZES FOR THE AVIRIS INDIAN PINES DATA SET. THE MEAN AND STANDARD DEVIATION OF TEN RUNS ARE SHOWED.

b	3	**5**	7	9	11
B	5	**5**	9	7	7
OA	85.41(±2.95)	**87.44**(±3.25)	85.43(±3.91)	85.77(±3.89)	85.06(±2.60)
κ	83.41(±3.13)	**85.71**(±3.68)	83.50(±4.39)	83.87(±4.37)	83.09(±2.90)

tion results for different ordering times in the case of the optimal box size $b = 5$ and search neighborhood $B = 5$ (see Fig. 2). Considering the instability of the classification results when the number of new permutations is very few, we choose the ordering times K in the interval [8,20]. When $K = 9$, the result achieves the optimal value. However, when the value of the ordering times is more than 9, the classification accuracy decreases slowly with the increase in the ordering times. This illustrates that with higher ordering times, the more demanding the classification is, and the worse the accuracy, which is similar to what is observed in over-fitting. Thus, the appropriate ordering times can allow the existence of fault tolerance and ease the problem 'different spectral samples share the same label' by utilizing the spatial structure information. Besides, it is easy to see that we obtain very good classification performance by using ordering times only 9 times. Thus, we choose $K = 9$ as the optimal value of the ordering times in this paper.

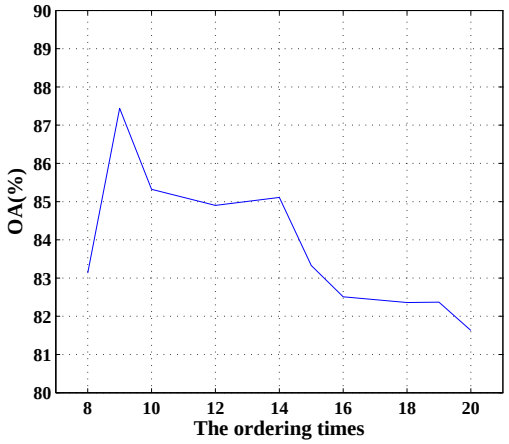


Fig. 2. The OA vs. the ordering times K for the AVIRIS Indian Pines data set. The mean in OA is obtained by ten runs ($b = 5$, $B = 5$).

Fig. 3 shows the classification results and maps of the pixel-based method and the proposed method under the conditions of $b = 5$, $B = 5$ and $K = 9$ in the AVIRIS Indian Pines data set with five labeled samples per class. These classifica-

(a) (b) (c)

Fig. 3. AVIRIS Indian Pines data set with the classification maps. (a) The ground reference map. (b) The classification map using the pixel-based method (OA=85.05%, κ=82.99%). (c) The classification map using the proposed method (OA=87.44%, κ=85.71%).

tion maps correspond to one of the ten Monte Carlo runs. It is easy to see that the proposed method has almost a same classification result for each region in the image, while the pixel-based method produces the confusing classification results for some regions (point-like). Obviously, the proposed method outperforms the pixel-based method, which illustrates that the image has large homogeneous regions, and different spectral samples may share the same label, and similar spectral samples may also share different labels. But the proposed method can make a full use of the spatial structure information to improve the classification.

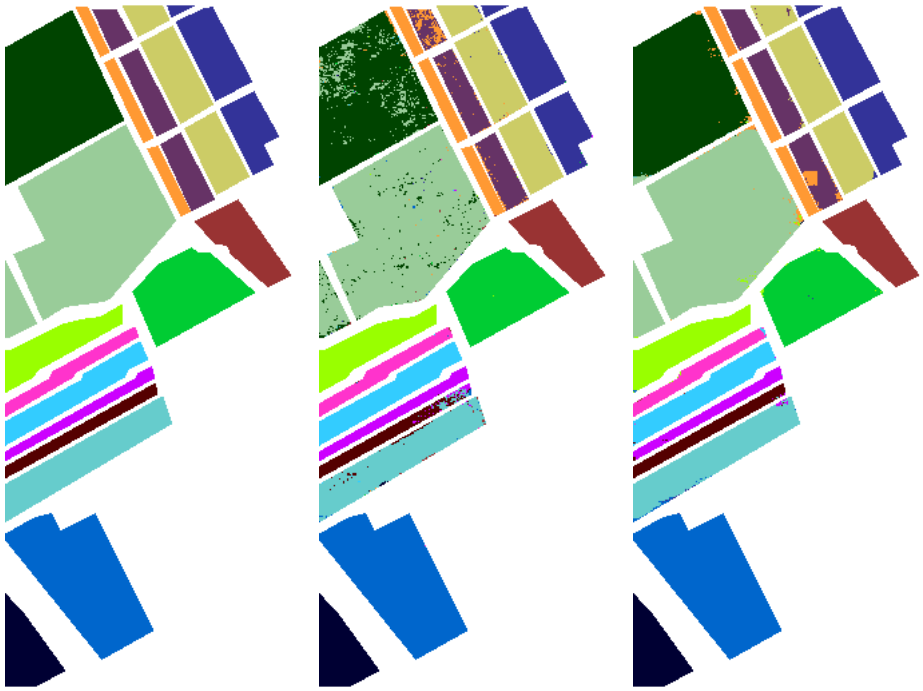
In order to show the good performance of the proposed method, we give the classification accuracy obtained by all the methods under the corresponding optimal conditions in Table 2. The proposed method achieves the best classification performance, which is much better than LapSVM,[1] MLR+AL,[19] even outperforms state-of-the-art Joint+Laplacian.[26] This confirms the effectiveness and advantage of the proposed method by compared with these popular semi-supervised methods.

Table 2. THE OA FOR ALL METHODS FOR THE AVIRIS INDIAN PINES DATA SET WITH FIVE LABELED SAMPLES PER CLASS. THE MEAN OF TEN RUNS IS SHOWED ($b = 5$, $B = 5$, $K = 9$).

Method	LapSVM[1]	MLR+AL[19]	Joint+Laplacian[26]	The proposed method
OA(%)	66.39	61.59	86.33	**87.44**

3.3. *AVIRIS Salinas scene*

This scene was collected by the 224-band AVIRIS sensor over Salinas Valley, California, and is characterized by high spatial resolution (3.7-meter pixel). The area covered comprises 512 lines by 217 samples. In this scene, we also discarded the



(a) (b) (c)

Fig. 4. AVIRIS Salinas data set with the classification maps. (a) The ground reference map. (b) The classification map using the pixel-based method (OA=98.83%, κ=98.70%). (c) The classification map using the proposed method (OA=98.46%, κ=98.29%).

20 water absorption band. It includes vegetables, bare soils, and vineyard fields. Salinas ground truth contains 16 classes, see Fig. 4 (a) for further details.

In order to compare the pixel-based method and the proposed method, we show their classification results and maps when $b = 11$, $B = 19$ and $K = 9$ in the AVIRIS Salinas data set with five labeled samples per class in Fig. 4. We find that the pixel-based method and the proposed method obtain similar high classification accuracy (even the pixel-based method does a little better in terms of accuracies). This illustrates that compared with the other two data sets, this data set is easier to process. The difference between different classes is very distinct and easy to distinguish, and there is almost no interference of different spectral information for the same class, and for different classes with similar spectral information. From Fig. 4, we can also see that the samples that are classified in error are point-like in the pixel-based method, while blocky in the proposed method. This illustrates that, in the case of little interference (same class with different spectral information and different classes with similar spectral information), the proposed method does not fully show the advantages of exploiting the spatial structure information.

Table 3 reports the classification accuracy of some popular semi-supervised methods and the proposed method in the AVIRIS Salinas data set. The proposed method obtains the best result, and shows very wonderful performance. LapSVM[1] and Joint+Laplacian[26] also perform well, a little worse than the proposed method. MLR+AL[19] is the worst in this data set. This experiment also demonstrates the proposed method can greatly improve the classification accuracy with very small labeled samples (only five labeled samples per class) in hyperspectral image classification.

Table 3. THE OA FOR ALL METHODS FOR THE AVIRIS SALINAS DATA SET WITH FIVE LABELED SAMPLES PER CLASS. THE MEAN OF TEN RUNS IS SHOWED ($b = 11$, $B = 19$, $K = 9$).

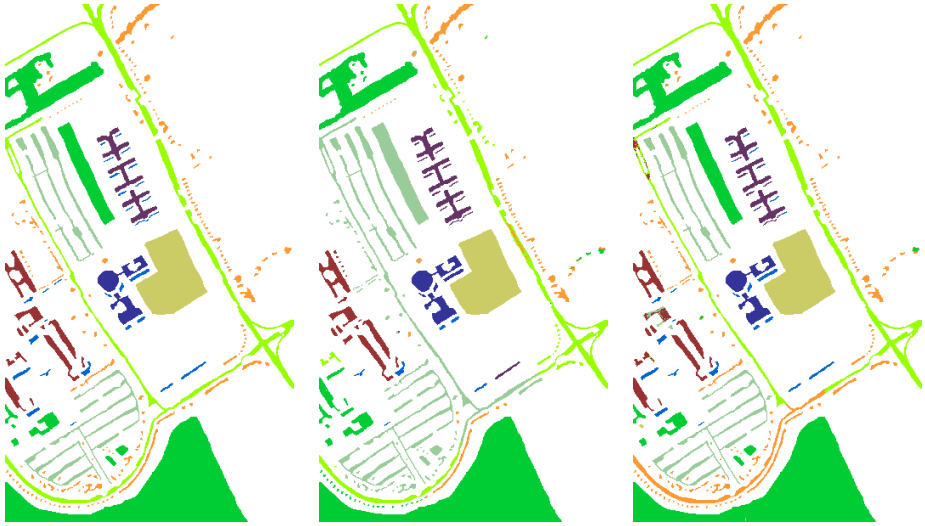
Method	LapSVM[1]	MLR+AL[19]	Joint+Laplacian[26]	The proposed method
OA(%)	90.36	87.21	91.14	**98.46**

3.4. *ROSIS Pavia University scene*

This scene acquired by the ROSIS sensor over the urban area of the university of Pavia, northern Italy. The flight was operated by DLR in the framework of the HySens project, managed and sponsored by the European Union. The image is 610×340 pixels. Due to noise 12 bands have been removed. The remaining 103 spectral bands are processed. Only 9 classes of interest are considered, see Fig. 5 (a) for further details.

Fig. 5 also gives the classification results of the pixel-based method and the proposed method when $b = 9$, $B = 9$ and $K = 9$ in the Pavia University data set with five labeled samples per class. The proposed method achieves better classification performance than the pixel-based method. From Fig. 3 (a), Fig. 4 (a), and Fig. 5 (a), we can see that the Pavia University scene is more complex than the AVIRIS Indian Pines scene and the AVIRIS Salinas scene in term of spatial structure information. Thus, the proposed method makes more full use of the advantages of box feature to improve the performance. On the other hand, in this data set, there also exists the phenomenon, that is, same class with different spectral information, and different classes with a similar spectral information. Thus, the proposed method can process the problem more effectively and reduce the error classification.

Further, we compared the proposed method with some popular semi-supervised methods (LapSVM,[1] MLR+AL,[19] and Joint+Laplacian[26]) in the Pavia University data set. The classification results are reported in Table 4. The proposed method is much better than other semi-supervised methods in terms of accuracies, even on average 9.23% higher in terms of accuracies than Joint+Laplacian.[26] Furthermore, Joint+Laplacian,[26] MLR+AL,[19] and LapSVM[1] perform all worse in terms of accuracies than the proposed approach, with LapSVM[1] performing worst on this data



(a) (b) (c)

Fig. 5. ROSIS Pavia University data set with the classification maps. (a) The ground reference map. (b) The classification map using the pixel-based method (OA=88.45%, κ=84.76%). (c) The classification map using the proposed method (OA=92.99%, κ=90.70%).

set. This verifies the effectiveness and advantage of the proposed method again.

Table 4. THE OA FOR ALL METHODS FOR THE PAVIA UNIVERSITY DATA SET WITH FIVE LABELED SAMPLES PER CLASS. THE MEAN OF TEN RUNS IS SHOWED ($b = 9$, $B = 9$, $K = 9$).

Method	LapSVM[1]	MLR+AL[19]	Joint+Laplacian[26]	The proposed method
OA(%)	71.05	80.22	83.76	**92.99**

4. Conclusion

In this paper, we adopt the semi-supervised learning classification framework introduced in Ref. 25 for the classification of hyperspectral images, using box-based smooth ordering as the initial step. The proposed method not only can effectively overcome the problem of high dimensionality in hyperspectral data, but also overcomes the lack of labeled samples. Furthermore, the proposed method is based on a box feature, not on a single pixel. The box feature can make use of the spatial organization in the image to improve the classification performance. Smooth ordering is used to project the box features into a 1D signal by ordering the boxes, which overcomes the problem of high dimensionality. This may seem a rough solution, but multiple times of ordering remedy for it. For the processing of 1D signals,

1D interpolation is both a simple and effective tool. Accurate results are obtained by combining multiple weak classification results. By constantly training and introducing new labeled samples, we finally classify all samples in a label boosting framework. Experimental results on three hyperspectral image data sets confirm the superior performance of the proposed method.

References

1. M. Belkin, P. Niyogi and V. Sindhwani, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, *J. Maching Learn. Res.* **7** (2006) 2399–2434.
2. J. A. Benediktsson, J. Chanussot, L. Najman and H. Talbot, Mathematical morphology and its applications to signal and image processing, *Proc. 12th International Symposium on Mathematical Morphology*, (Springer, 2015).
3. S. Bernabe, P. R. Marpu, A. Plaza, M. D. Mura and J. A. Benediktsson, Spectral-spatial classification of multispectral images using kernel feature space representation, *IEEE Geosci. Remote Sens. Lett.* **11**(1) (2014) 288–292.
4. L. Bruzzone, M. Chi and M. Marconcini, A novel transductive svm for semisupervised classification of remote sensing images, *IEEE Trans. Geosci. Remote Sens.* **44**(11) (2006) 3363–3373.
5. G. Camps-Valls, T. Bandos and D. Zhou, Semi-supervised graph-based hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **45**(10) (2007) 3044–3054.
6. G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés and J. Calpe-Maravilla, Composite kernels for hyperspectral image classification, *IEEE Geosci. Remote Sens. Lett.* **3**(1) (2006) 93–97.
7. G. Camps-Valls, D. Tuia, L. Bruzzone and J. A. Benediktsson, Advances in hyperspectral image classification: Earth monitoring with statistical learning methods, *IEEE Signal Proc. Mag.* **31**(1) (2014) 45–54.
8. Y. Chen, N. M. Nasrabadi and T. D. Tran, Hyperspectral image classification using dictionary-based sparse representation, *IEEE Trans. Geosci. Remote Sens.* **49**(10) (2011) 3973–3985.
9. T. H. Cormen, *Introduction to algorithms* (MIT Press, 2009).
10. I. Dopido, J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias and J. A. Benediktsson, Semisupervised self-learning for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **51**(7) (2013) 4032–4044.
11. M. Fauvel, J. Chanussot and J. A. Benediktsson, A spatial-spectral kernel-based approach for the classification of remote-sensing images, *Pattern Recognit.* **45**(1) (2012) 381–392.
12. P. Ghamisi, J. A. Benediktsson and J. R. Sveinsson, Automatic spectralcspatial classification framework based on attribute profiles and supervised feature extraction, *IEEE Trans. Geosci. Remote Sens.* **52**(9) (2014) 5771–5782.
13. P. Ghamisi, M. D. Mura and J. A. Benediktsson, A survey on spectralcspatial classification techniques based on attribute profiles, *IEEE Trans. Geosci. Remote Sens.* **53**(5) (2014) 2335–2353.
14. L. Gomez-Chova, G. Camps-Valls, J. Munoz-Mari and J. Calpe, Semisupervised image classification with laplacian support vector machines, *IEEE Geosci. Remote Sens. Lett.* **5**(3) (2008) 336–340.
15. G. F. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Trans. Inf. Theory* **14**(1) (1968) 55–63.

16. H. Li, G. Xiao, T. Xia, Y. Y. Tang and L. Li, Hyperspectral image classification using functional data analysis, *IEEE Trans. Cybern.* **44**(9) (2014) 1544–1555.
17. H. Li, Z. Ye and G. Xiao, Hyperspectral image classification using spectral-spatial composite kernels discriminant analysis, *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* **8**(6) (2014) 2341–2350.
18. J. Li, J. Bioucas-Dias and A. Plaza, Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields, *IEEE Trans. Geosci. Remote Sens.* **50**(3) (2012) 809–823.
19. J. Li, J. M. Bioucas-Dias and A. Plaza, Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning, *IEEE Trans. Geosci. Remote Sens.* **48**(11) (2010) 4085–4098.
20. L. Ma, M. M. Crawford and J. Tian, Local manifold learning-based k-nearest-neighbor for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **48**(11) (2010) 4099–4109.
21. I. Ram, M. Elad and I. Cohen, Image processing using smooth ordering of its patches, *IEEE Trans. Geosci. Remote Sens.* **22**(7) (2013) 2764–2774.
22. F. Ratle, G. Camps-Valls and J. Weston, Semisupervised neural networks for efficient hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **48**(5) (2010) 2271–2282.
23. B. Shahshahani and D. Landgrebe, The effect of unlabeled samples in reducing the small sample size problem and mitigating the hughes phenomenon, *IEEE Trans. Geosci. Remote Sens.* **32**(5) (1994) 1087–1095.
24. J. Wang, Semi-supervised learning using ensembles of multiple 1d-embedding-based label boosting, in this special issue of *International Journal of Wavelets, Multiresolution and Information Processing* (2015).
25. J. Wang, Semi-supervised learning using multiple one-dimensional embedding-based adaptive interpolation, in this special issue of *International Journal of Wavelets, Multiresolution and Information Processing* (2015).
26. Z. Wang, N. M. Nasrabadi and T. S. Huang, Semisupervised hyperspectral classification using task-driven dictionary learning with laplacian regularization, *IEEE Trans. Geosci. Remote Sens.* **53**(3) (2014) 1161–1173.