

Spectral-spatial classification for hyperspectral imagery: a novel combination method based on affinity scoring

Zhao CHEN^{1,2,3} & Bin WANG^{1,2,3*}

¹State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China;

²Key Laboratory for Information Science of Electromagnetic Waves (Ministry of Education), Fudan University, Shanghai 200433, China;

³Research Center of Smart Networks and Systems, School of Information Science and Technology, Fudan University, Shanghai 200433, China

Received September 13, 2015; accepted January 20, 2016; published online August 26, 2016

Abstract Recently, a general framework for spectral-spatial classification has caught the attention of the hyperspectral imagery (HSI) society. It consists of three parts: classification, segmentation and combination of the former results to make a refined labeled map. Seeing the potentials of the last part, we derive a novel combination rule based on affinity scoring (CRAS). The core of the system is affinity score (AS), which is derived from fuzzy logic. Every AS measures the degree, i.e., the affinity, by which a pixel belongs to a class. The score is essentially decided by three factors: local spatial consistency, spectral similarity, and prior knowledge. The method is compatible with basic classification and segmentation tools, thus saving the trouble of designing complex techniques for the other parts in the framework. Experimental results show that CRAS excels several basic techniques as well as various state-of-the-art methods in the area of spectral-spatial classification.

Keywords hyperspectral imagery, spectral-spatial classification, affinity score, local spatial consistency, fuzzy, superpixel

Citation Chen Z, Wang B. Spectral-spatial classification for hyperspectral imagery: a novel combination method based on affinity scoring. *Sci China Inf Sci*, 2016, 59(10): 102313, doi: 10.1007/s11432-016-5576-y

1 Introduction

Classification for hyperspectral imagery (HSI) is useful to many applications, such as crop mapping, environmental monitoring and military target detection. Hyperspectral remote sensor technology allows the simultaneous acquisition of hundreds of spectral wavelengths for each image pixel. This detailed spectral information increases the possibility to accurately identify materials of interest [1]. However, it may also cause problems classification [2]. One of them is intra-class heterogeneity, which means that pixels of the same class deviate from each other in the spectral domain. Most pixel-wise classifiers such as support vector machines (SVMs) [3] are incapable of solving the problem because they merely treat HSI as a list of spectral measurements with no spatial organization [1, 2]. Another challenge is the scarcity

*Corresponding author (email: wangbin@fudan.edu.cn)

of prior knowledge (pixels with their class labels known prior to classification). It may compromise the efficacy of supervised classifiers, which relies on the amount and the quality of training samples.

Fortunately, the use of spatial information is a solution to these problems. Therefore, spectral-spatial classification has attracted large attention [1, 2]. Using both spectral and spatial information of HSI, it aims at improving classification accuracy and reducing salt-and-pepper errors. There are three major kinds of the approach [2–4]: (1) constructing novel classifiers that simultaneously classify by spectral and spatial information, (2) concatenating the spectral and the spatial features and feeding them to existing classifiers and (3) processing the spectral and the spatial features separately then combining the two results to create one classification map. Sparsity-representation-based classifiers [5] belong to the first family. These classifiers perform well but require delicate parameter-tuning. The second kind of the approach often involves several features [1, 6, 7]. It leaves users a difficult job to choose the features which are the most helpful to classification among dozens of alternatives. A typical technique of the sort is denoted by EMP-KPCA, using Kernel Principal Component Analysis (KPCA) for spectral feature extraction and extended morphological profile (EMP) for spatial analysis [1, 8]. In the third category there are a variety of methods [1, 3, 9, 10]. Spectral information can be handled by one or multiple classifiers [11]. Meanwhile, spatial information can be exploited by segmentation [9, 12], marker selection [13, 14], Markov random field (MRF) [3] or edge-preserving filtering [15] etc. Among them, the segmentation methods such as simple linear iterative clustering (SLIC) [9] have been fully studied and largely practiced, with their toolboxes easily being accessed online.

In this paper, we adopt an existing framework in the third category, which consists of three main parts: spectral classification, spatial segmentation and combination. We use the framework because it is clearly laid out and widely accepted. In the literature, the classification and the segmentation have been often the focus of study. Many state-of-the-art algorithms share the same idea, e.g. the method of hierarchical segmentation plus majority voting (Hseg + MV) [1, 12, 16–20], consider them the most crucial parts of the framework [1, 2, 11]. Feature extraction, which is also an intensely studied area, is an optional pre-processing step for classification and segmentation.

Comparatively, the combination part lacks improvement. As one of the most frequently used combination method, MV is, however, not robust to errors caused by poor segmentation or classification [1, 6, 11, 21]. Weighted majority voting (WMV) is better than MV [10], but still, its accuracy is largely dependent on segmentation outcomes. One effective improvement is brought by marker selection. Typical examples of this kind include MSSC-MSF using multiple spectral-spatial classifier approach followed by minimum spanning forest segmentation (MSF) [11, 22] and SVMMSF + MV/SVMMSF using probabilistic SVMs followed by MSF with/without the optional MV step [13, 23] Although being viewed as a combination function sometimes [1], marker selection is actually more of a complex system embedded with segmentation. Moreover, it requires delicate parameter tuning.

Therefore, we focus on designing a combination method which is both simple and effective. By experience, we have found out that HSI samples show local spatial consistency in some scale, which means that a sample is highly likely to agree with its neighbors in class affiliation and pixel value. Moreover, as long as the local spatial consistency is maintained, even a rough over-segmentation map is qualified for combination. Therefore, we see the property as a key to an effective combination method, which is able to save the efforts spent on segmentation or classification.

Given that, we propose a combination rule based on affinity scoring (CRAS), which is also our main technical contribution. It takes care of local consistency along with other two factors, spectral similarity and prior knowledge, by affinity score (AS). AS is the core of CRAS¹). It measures how much a pixel belongs to a class by weighing on these factors. Every scoring takes place in the neighborhood of a superpixel, which is defined as a segment of HSI [9, 24]. In this way, AS considers the local spatial consistency maintained in the superpixels. Meanwhile, every score of a pixel is based on spectral information of all the other members within the neighborhood. Thus, AS takes care of spectral similarity. Unlike MV which only involves segmented and classified maps, AS also exploits prior knowledge. Misleading

1) For clarity, we denote the affinity score or affinity scoring by AS and the combination rule or method involving AS by CRAS throughout the paper.

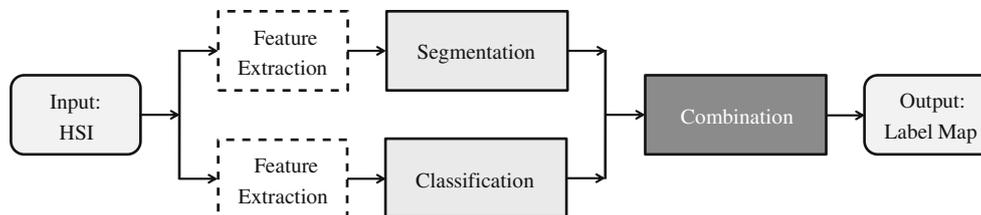


Figure 1 The general spectral-spatial classification framework.

information brought by the classification step therefore can be corrected by the re-use of training samples. Furthermore, AS is straightforward to be implemented since it only involves scalar computations.

On the whole, CRAS combines segmentation and classification by the fuzzy-logic-based AS. It exploits the fuzziness in the weights of the spectral and the spatial information, which are reflected by the classification and the segmentation outcomes, respectively, on the final classification map [25]. It also balances regional unity and pixel individuality by grading the impacts of the neighboring regions on a pixel while making decisions for every pixel individually. These features make CRAS a novel method.

The immediate contribution of CRAS to classification is the increase of accuracy. Moreover, CRAS guarantees its classification accuracy even when the accuracy of the preliminary procedures is not quite good. It can work with most existing classification and segmentation methods. Thus, we prefer using simple and efficient algorithms such as SVM or K-Nearest Neighbor (KNN) for preliminary classification and SLIC for over-segmentation. Based on two different definitions of superpixel neighborhood given in Section 3, we propose two effective modes of CRAS. Experiments show that either CRAS1/2 excels the basic MV/WMV as a sheer combination method when SVM/KNN and SLIC are engaged. They also prove that SVM/KNN + SLIC + CRAS1/2 is better as a whole classification-segmentation-combination system than the state-of-the-art approaches such as EMP-KPCA, Hseg + MV, SVMMSF, SVMMSF + MV and MSSC-MSF given the same condition. In addition, the analysis suggests that CRAS2 is more capable of dealing with poor preliminary classification results than CRAS1.

The rest of the paper is organized as follows. In Section 2, background knowledge about the general spectral-spatial classification framework, Principal Component Analysis (PCA), SLIC, SVM and KNN is briefly introduced. In Section 3, our method CRAS is proposed and elaborated. In Section 4, experimental results that demonstrate the efficacy of CRAS are analyzed and further insights of the proposed method are also provided. Finally, a conclusion is drawn in Section 5.

2 Theoretical background

2.1 The general spectral-spatial classification framework

This spectral-spatial classification framework for HSI is clearly laid out and widely accepted. It takes advantage of both spectral and spatial information by combining segmentation and classification results [1, 2, 6, 11]. It includes three major parts: segmentation, classification and combination, as illustrated in Figure 1. Feature extraction is an optional pre-processing step. In the combination procedure, segmented and classified maps are fused to produce a refined land cover map according to certain decision rules. The whole strategy is concise while specific methods in each step are user-dependent.

Most of the existing methods derived within the general framework emphasize on the improvements in feature extraction, classification or segmentation [1, 2, 6, 10, 11]. These modifications are often theoretically demanding and also practically complicated. However, we find that this complication can be avoided by simply improving the combination method. The key is to use local spatial consistency observed in small spatial structures wisely. It is significant to the correction of misclassifications. As long as these spatial patterns are captivated and maintained, a proper combination method can largely increase classification accuracy even by a roughly over-segmented map. Therefore, our adaptation of the framework focuses on the part of combination while simplifying the other steps.

2.2 Superpixel and SLIC

Recently, the concept of superpixel has been widely adopted [24]. A superpixel contains a bunch of natural pixels in an image that can be regarded as an entity holding some meaningful information. Suppose a hyperspectral data set $X \in \mathbb{R}^{I \times J \times Q}$ with $m = 1, 2, \dots, M$ superpixels $X_m \in \mathbb{R}^{B_m \times Q}$, whereas I, J, Q and B_m represent number of rows, columns, bands and the size of superpixel X_m (i.e., the number of members in the superpixel), respectively. In the area of segmentation, a superpixel X_m is constantly used to represent a segment of an image. Herein, M is determined by the image size $I \times J$, the superpixel size B_m and the segmentation algorithm being used.

Since our method CRAS relies on superpixels of small sizes to maintain local consistency, SLIC [9] is the exact choice for generation of such superpixels. We choose SLIC for two reasons. Firstly, it exhibits state-of-the-art boundary adherence, while being more memory efficient than most of the other existing methods. Secondly, SLIC can expose the true capacity of the follow-up combination method due to its ability to perform efficient over-segmentation instead of accurate segmentation for HSI. If highly accurate segmentation methods were adopted, the outcomes of most combination methods would be improved, of course. But it would also be difficult to tell apart the contributions of the segmentation tool from those of the combination algorithm.

Every superpixel generated by SLIC consists of approximately $R_s \times R_s$ pixels, whereas R_s ($B_m \approx R_s^2$) is fixed beforehand. The shapes of the superpixels are regularized by a given factor R_e . Since SLIC is only an alternative segmentation tool in the general framework, its computation details are omitted here, which are available in [9].

2.3 PCA, SVM and KNN

We choose SVM and KNN for classification and PCA for feature extraction, as they are all classical and useful algorithms. PCA can suppress the curse of dimensionality, ease the burden of computation and improve classification accuracy. SVMs perform supervised non-linear pixel-wise classification based on the full spectral information. KNN is also supervised, which assigns a pixel to the class occupying the majority in the k-nearest neighborhood of the pixel.

3 The proposed fuzzy decision system

3.1 Superpixel-wise similarity

Since the term of superpixel is the generalized form of pixel, the concept of pixel-wise (dis)similarity can be extended to superpixels. We use the superpixel-wise similarity to find the neighbors of a superpixel. To obtain good results from CRAS, we design the measurement in a way tightly related to AS, which is formulated by (1)

$$S_{mn} = \frac{\sum_{c=1}^C \sum_{1 \leq i \leq B_m} \sum_{1 \leq j \leq B_n} s_{ij} w_i^c w_j^c}{\sum_{c=1}^C \sum_{1 \leq i \leq B_m} \sum_{1 \leq j \leq B_n} w_i^c w_j^c}, \quad m, n = 1, 2, \dots, M, \quad (1)$$

where B_m and B_n represent the sizes of superpixel X_m and X_n , i.e., the numbers of pixels in X_m and X_n , respectively. M is the total number of superpixels in the image. $x_i^m \in \mathbb{R}^{Q \times 1}$, $i = 1, 2, \dots, B_m$ and $x_j^n \in \mathbb{R}^{Q \times 1}$, $j = 1, 2, \dots, B_n$ denote the members of X_m and X_n , respectively. $\forall c \in \{1, 2, \dots, C\}$ is the index for a class known beforehand. If pixel x_i is given the label of any class $c \in \{1, 2, \dots, C\}$ as prior knowledge, its weight w_i^c should be given a great value $W_1 \gg 1$; otherwise, w_i^c is 1. Throughout our experiments, W_1 is fixed at the same value for every pixel, whereas the value is chosen empirically. Moreover, $s_{ij} = d_{ij}^{-1}$, $d_{ij} \neq 0$. d_{ij} can be any spectral distance between two pixels, such as Euclidean Distance (ED) and Spectral Angle Metric (SAM) [4]. We define that $d_{ii} = 0$ and $s_{ii} = 0$ for simplicity. In our experiments, we prefer using natural exponential of pixel-wise correlation coefficient (CC) as s_{ij} ,

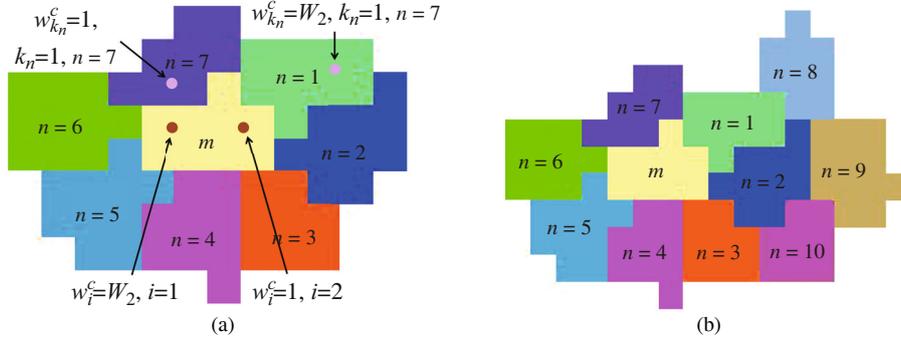


Figure 2 (Color online) Two definitions of superpixel neighborhood. (a) Natural neighborhood; (b) expanded neighborhood.

because it directly measures spectral similarity. The natural exponential is added so that s_{ij} can stay non-negative, which makes the CRAS to be proposed easy to be implemented

The superpixel-wise similarity above considers both pixel-wise similarity and prior knowledge (i.e., class labels known beforehand). A great value of the measurement does not just mean that the two superpixels being considered are spectrally close to each other. It also suggests that they may contain labeled samples of the same classes, thus being very likely to be consistent in class affiliation.

3.2 Neighborhood of a superpixel

There can be many alternative ways to define the superpixel neighborhood for CRAS. Here we propose two, which are simple and effective compared with all the other alternatives we have tested so far.

A pair of superpixels can be regarded neighbors if each of them has, at least, one pixel adjacent to another. We name this kind of neighbors as “natural”. However, if the natural neighbors of a superpixel are incorrectly labeled during a preliminary classification procedure, they can have little contribution to improving the classification accuracy of the superpixel. One solution is to include the neighbors of the most similar neighbor of the superpixel during AS. The similarity is evaluated by (1). Therefore, we define the “expanded” neighborhood as the union of the natural neighborhood of a superpixel and the natural neighborhood of the most similar neighbor to the superpixel.

These two definitions are illustrated by Figure 2. In Figure 2(a), superpixel X_m has 7 neighbors ($n = 1, 2, \dots, 7$). Meanwhile, by the second definition, there becomes 10 neighbors as presented in Figure 2(b), whereas the neighbors of the most similar neighbor ($n = 2$) of the superpixel X_m are introduced.

3.3 Affinity scoring

The aim of the combination step is to increase classification accuracy and reduce salt-and-pepper errors. While it is necessary to deal with both spectral similarity and local spatial consistency, the task can be difficult when conflicts exist between these two kinds of information. A superpixel with great spectral heterogeneity may result in inconsistency in its classification map. Yet in reality, its members are highly likely to fall into the same class as they are in the same region. It is uncertain to decide which of the segmented and the classified results contributes more than the other. In another word, it is in a fuzzy state that to what extent the class of a pixel should agree with its neighbors or merely the outcome of the spectral classifiers. Since the conflicts are unavoidable and unpredictable, we try to resolve them by combining the spectral and the spatial information, reflected by the classification and the segmentation, respectively

Therefore, we propose a combination rule, CRAS, based on fuzzy logic [25]. AS is the core of the method. It is determined by three factors simultaneously: local spatial consistency, spectral similarity and prior knowledge. AS measures the affinity of every member of a superpixel with one class by pixel-wise similarities of all the members and weights on the labeled samples within the neighborhood of this

superpixel. The higher the score is, the more likely the pixel does belong to the class. The superpixels and their neighborhoods are adopted for local consistency. Though primarily exploited by supervised classifiers, the prior knowledge is re-considered by AS for two reasons. First of all, the labeled samples can be rather valuable to correction of misclassifications. Second of all, some classifiers may fail to predict correct labels for a seriously spectrally heterogeneous superpixel even though it contains labeled samples. Thus, the training samples are heavily weighted during scoring.

By A_i^c we denote the AS which quantifies the degree (i.e., affinity) by which member x_i^m belongs to class c existing in the preliminary classification map of superpixel X_m . The detailed analysis of A_i^c is stated as follows. The affinity score contributed by the other members of X_m is

$$I_i^c = \sum_{1 \leq j \leq B_m} s_{ij} w_j^c, \quad i, j \in \{1, 2, \dots, B_m\}, \quad c \in \{1, 2, \dots, C_m\}, \quad i \neq j, \quad m = 1, 2, \dots, M, \quad (2)$$

which shares the definitions of s_{ij} , w_j^c , B_m and M with (1). w_j^c is the weight put on the training samples, which are provided as prior knowledge.

Suppose that superpixel X_m has N neighbors X_n , $n = 1, 2, \dots, N$ by an arbitrary definition of superpixel's neighborhood. Then the score contributed by the members of X_n is

$$O_i^c = \sum_{n=1}^N \sum_{1 \leq k_n \leq B_n^m} s_{ik_n} w_{k_n}^c, \quad k_n \in \{1, 2, \dots, B_n^m\}, \quad i \neq k_n, \quad n = 1, 2, \dots, N, \quad (3)$$

where B_n^m is the size of neighbor X_n and $w_{k_n}^c$ weighs on the training samples in this superpixel. Similarly, $w_{k_n}^c = W_2$ if k_n is already given a label of any class; otherwise, $w_{k_n}^c = 1$. It should be noted that $w_{k_n}^c$ is different from w_j^c . Since pixel $x_{k_n}^n$ is outside X_m while x_j^m is inside, the latter deserves a heavier weight than the former. Meanwhile, it is better to set W_1 and W_2 greater than the number classes according to the denominator of (4). Thus, $C_m \leq W_2 \leq W_1$ is recommended but not compulsive, whereas their values are chosen empirically. In Figure 2(a), the dots represent some members of superpixels in a neighborhood, with their weights placed at the end of the arrows.

In summary, the whole AS is given by (4)

$$A_i^c = \frac{I_i^c + O_i^c}{\sum_{c_m=1}^{C_m} I_i^{c_m} + \sum_{c_n=1}^{C_n} O_i^{c_n}}, \quad (4)$$

where $c_n = 1, 2, \dots, C_n$ present classes in the preliminary classification map of the neighboring superpixels. Obviously, A_i^c ranges from 0 to 1. The higher the score is, the greater the affinity is.

From the above, we can see that AS only considers one superpixel and its neighborhood at a time for spatial proximity and local consistency. It also explicitly involves spectral similarity and prior knowledge. Theoretically, the score can be embedded with any form of pixel-wise similarity. Furthermore, AS is easy to be computed, as it involves only scalar calculations.

According to Subsection 3.2, there are natural neighborhood and expanded neighborhood of a superpixel, from which we derive two types of scores, denoted by AS1 and AS2, respectively.

3.4 Procedure for implementation

Here we present a novel combination method CRAS, which can largely increase classification accuracy. After AS, the decision on class affiliation for a pixel takes place in the neighborhood of the superpixel containing the pixel. This is also to keep local spatial consistency. The pixel is assigned to a score for each class and then sorted into the only class with the highest score.

The whole procedure is outlined in Algorithm 1. It shows two features of CRAS: simplicity and compatibility. CRAS is easy to be implemented since it only involves scalar operations. Meanwhile, CRAS is compatible with various forms preliminary classification and segmentation algorithms. All it needs are basic tools such as PCA, SVM/KNN and SLIC. It can guarantee good performance even if the co-operating methods in the general framework have failed. From Algorithm 1, we can see that the computational

Algorithm 1 The adapted spectral-spatial classification framework based on CRAS

Input: $\mathbf{X} \in \mathbb{R}^{I \times J \times Q}$: a set of hyperspectral images I, J, Q : number of rows, columns and bands of the HSI, respectively, let $N_s = I \times J$ T : maximum iteration times**Output:** final map of labels**Step 1.** Classification

1a) Extract spectral features by PCA

1b) Classify by SVM or KNN with the spectral features and produce a roughly labeled map

Step 2. Segmentation2a) Segment \mathbf{X} by SLIC and create an over-segmented map2b) Number the superpixels $\mathbf{X}_m \in \mathbb{R}^{B_m \times Q}, m = 1, 2, \dots, M$, where M is the total amount of the superpixels and B_m is the size of each superpixel**Step 3.** Combination of Classification and Segmentation by CRAS3a) Compute pixel-wise similarity (CC) for each pair of the pixels in \mathbf{X} and set $t = 0$ while $t < T$ dofor $m = 1, 2, \dots, M$ 3b) Find the neighbors of \mathbf{X}_m by either of the definitions given in Subsection 3.2for $i = 1, 2, \dots, B_m$ 3c) Calculate AS1/2 for each superpixel member x_i^m according to (4)3d) Label x_i^m with the class that earns the highest score

end

end

end

3e) Output the refined classification map

complexity of Step 3 is $O(N_s(1 + R_s^4 + C \log C))$ if AS1 is engaged, or $O(N_s(2 + 2R_s^4 + C \log C))$ if AS2 is engaged.

In addition, there are four empirical recommendations. First, in Step 3d), if several classes earn the highest score, which is very rare, the class appears the most frequent in the whole neighborhood wins. If it still cannot be decided, then the pixel is assigned to one of these classes randomly. Second, CRAS2 embedded with the expanded neighborhood definition is better to be performed after, at least, an iteration of CRAS1. Since Step 1 only offers a roughly classified map, a large number of mistakes are expected in every superpixel and its neighbors. To emphasize on local consistency and avoid including too many questionable labels, CRAS should be confined to a rather small neighborhood during the first few iterations, whereas the natural neighborhood and CRAS1 are applied. Third of all, the labels of a superpixel can be regarded reliable and adopted as prior knowledge during the next iteration if all the superpixels in the neighborhood reach a unanimous decision about class affiliation at the present iteration. This move is to supplement the otherwise limited prior knowledge. Last but not least, after every superpixel in the image is checked, the whole process of AS and class assigning can be iterated so that classification accuracy may be further increased. The number of iterations is empirically determined.

4 Experimental results

4.1 Hyperspectral data

In our experiments, four hyperspectral data sets are used. Particularly, the well-known Indian Pines data set is analyzed in details. It is used as a benchmark to validate the accuracy of HSI analysis algorithms, especially for classification. There exist nonlinearity and serious intra-class heterogeneity in all available classes. It was captured in 1992 by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines region in Northwestern Indiana, a mixed agricultural/forest area²⁾. The image

2) <http://dynamo.ecn.purdue.edu/biehl/MultiSpec>

Table 1 Default setting of parameters and other components in CRAS

Item	Explanation	Choice
R_s	Approximate size ($R_s \times R_s$) of superpixels acquired by SLIC	3
R_e	Regularity of the shape of each superpixel	50
W_1	Weight on the member of the superpixel being scored when it is a training sample	800
W_2	Weight on the member in the neighborhood of the superpixel being scored when it is a training sample	50
T	Times of iterations of CRAS	1
P	Number of principal components kept by PCA	22

comprises 145×145 pixels and 220 spectral channels in the wavelength range from 0.4 to 2.5 μm . Bands 104–108, 150–163 and 220 have been removed due to noise or water absorption [26].

The other three are the University of Pavia data set³⁾, the Valley of Salinas data set⁴⁾, and the Kennedy Space Center data set⁵⁾. They are engaged to validate the conclusions drawn from the experiments on the first one.

4.2 Setup for experiments

Following the procedure in Algorithm 1, first we conduct classification and segmentation independently by applying SVM⁶⁾/KNN⁷⁾ and SLIC⁸⁾, respectively. As to suppress Hughes' effect and also to save time, PCA is performed prior to classification, whenever SVM or KNN is applied. SVM is tuned, only once per data set, by cross validation on one random selection of the training sets. The parameters of KNN adopt the default values provided by the MATLAB toolbox, whereas $K = 1$. Then we combine the classified map and the segmented map by CRAS. Due to the fact that SVM and KNN are supervised classifiers, a percentage of samples with given labels should be randomly selected and saved for training. The percentage is called training-to-total sample ratio (TTR). TTR should be kept low since real HSI generally lacks in prior knowledge. For statistical consistency, every numerical result is the average of 20 runs, whereas each run involves one set of the training samples being randomly selected.

Given two definitions of superpixel's neighborhood, two types of algorithms, CRAS1 and CRAS2, are derived. While SVM/KNN and SLIC are engaged, CRAS1/2 is compared with MV and WMV, the most frequently used decision rules. SVM/KNN + SLIC + CRAS1/2 is also compared with five state-of-the-art techniques for spectral-spatial classification, EMP-KPCA [8], Hseg + MV [12, 16–20], SVMMSF, SVMMSF + MV [13, 14, 23] and MSSC-MSF [11, 22], which are covered by a recently published review [1]. The indexes that evaluate the quality of the final classified maps are overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ).

In our experiments, all the parameters of CRAS are set by default, except when there are particular statements. When faced with unknown data in reality, we can hardly figure out the best parameter values. Therefore, we prefer the settings which can provide good results in general, though not the best. Still, the results can validate the efficacy of the proposed method. The default values are listed in Table 1. They are chosen based on several trials. All the algorithms are implemented in MATLAB R2013b running on a workstation with an Intel(R) Xeon(R) CPU X5667 @ 3.00GHz (dual core) and 48.0 GB of RAM.

4.3 Evaluation on the Indian Pines and the University of Pavia data sets

4.3.1 Comparison with basic combination rules

Table 2 exhibits the classification results of CRAS1/2 and some basic combination rules when SVM/KNN and SLIC are applied. It shows that CRAS1/2 obviously excels the basic MV/WMV for both test subjects

3) <http://www.ehu.es/ccwintco/uploads/e/ee/PaviaU.mat>

4) <http://www.ehu.es/ccwintco/uploads/f/fl/Salinas.mat>

5) <http://www.csr.utexas.edu/hyperspectral/data/KSC/>

6) Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

7) Available in MATLAB toolbox

8) Available online: <http://www.ivrg.epfl.ch/research/superpixels>.

Table 2 Comparison of CRAS with basic combination rules, MV and WMV

Data set	Indian Pines						University of Pavia					
	5.01%			9.99%			1.00%			5.00%		
	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ
SVM	78.96	70.58	0.76	84.45	79.83	0.82	85.74	80.34	0.81	93.02	90.50	0.91
SVM + SLIC + MV	85.79	75.48	0.84	91.61	86.66	0.90	89.04	83.96	0.85	96.25	94.20	0.95
SVM + SLIC + WMV	85.82	75.59	0.84	91.62	86.76	0.90	89.32	84.63	0.86	96.38	94.53	0.95
SVM + SLIC + CRAS1	95.62	89.49	0.95	97.76	94.37	0.97	94.21	91.27	0.92	99.14	98.16	0.99
SVM + SLIC + CRAS2	96.99	89.83	0.97	98.26	94.17	0.98	95.10	92.31	0.93	99.35	98.43	0.99
KNN	73.04	66.39	0.69	78.22	73.87	0.75	79.70	76.88	0.73	85.29	83.19	0.80
KNN + SLIC + MV	81.66	71.84	0.79	87.57	80.53	0.86	84.43	80.13	0.79	90.75	88.10	0.87
KNN + SLIC + WMV	81.56	72.28	0.79	87.58	80.90	0.86	84.60	80.79	0.79	88.67	91.02	0.88
KNN + SLIC + CRAS1	95.19	89.00	0.95	97.69	94.28	0.97	91.22	88.23	0.88	98.49	97.31	0.98
KNN + SLIC + CRAS2	96.75	89.55	0.96	98.22	94.13	0.98	92.33	89.39	0.90	98.94	97.83	0.99

when the same classifier and the segmentation tool are engaged. Moreover, it indicates that our algorithms are very robust to the inefficacy of the classifiers. For instance, as the OA of the Indian Pines images given by KNN drops to 73.04% (TTR = 5.01%), CRAS1/2 keeps the index stay above 95%. On the contrary, the accuracies of MV and WMV are seriously reduced.

MV and WMV perform poorly mainly because they neglect prior knowledge or pixel individuality. Completely relying on the preliminary outcomes, MV treats the votes of every pixel indifferently without considering any spectral information at all. WMV does not reuse the training data as CRAS does, although weighing on the vote of each pixel differently by its distance to the superpixel it resides in. Moreover, MV/WMV assigns a single class label to all the members in an entire superpixel, which can be unreasonable since it is possible that the superpixel can cover more than one class. This also can be a risky move as the preliminary classification or segmentation results are not 100% reliable.

It can also be inferred from Table 2 that CRAS is able to handle the aforementioned challenges in HSI classification: the spectral variability and the scarcity of prior knowledge. In the Indian Pines data set, there exists serious intra-class heterogeneity in all available classes (a widely perceived fact, no need to be justified in this paper). Still, CRAS gives it very high classification accuracies, all above 95%. What is more, CRAS performs well no matter when the TTR is about 5% or 10%. For the University of Pavia data set, CRAS can even stand the TTR of 1.00%. Although CRAS is engaged after a supervised classifier, it does not require abundant training samples in order to be effective. This is a useful advantage since the labeled samples are difficult to be acquired in reality.

Furthermore, it is interesting to discuss the unique value of CRAS2. It outperforms CRAS1 in most cases, though the advantage is not so distinct. Nevertheless, CRAS2 shows its potentials in dealing with unfavorable situations, whereas the preliminary classification results of a data set are rather poor. For example, when the preliminary OA of the Indian Pines data set is 78.96% (TTR = 5.01%, SVM employed), CRAS2 outperforms CRAS1 by 1.37%; however, this gap has grown to 1.56% as the preliminary OA drops to 73.04% (TTR = 5.01%, KNN employed). In this case, expanded neighbors of a superpixel with different labels and extra information are introduced to correct the poor judgment made by a few natural neighbors of the superpixel. With these merits, CRAS2 can be regarded as an updated version of CRAS1.

4.3.2 Comparison with state-of-the-art algorithms

To fully demonstrate the effectiveness of CRAS, we compare it with some state-of-the-art methods as a whole spectral-spatial classification process. When the test subject is the Indian Pines data set, SVM/KNN + SLIC + CRAS1/2 is compared with Hseg + MV, SVMMSF, SVMMSF + MV and MSSC-MSF, which are included by a recently published review [1]. The first one is dedicated to the improvement of segmentation and the last three versions of marker selection focused on classification. For the Univer-

Table 3 Comparison of SVM/KNN + SLIC + CRAS with some state-of-the-art algorithms

Data set	Indian Pines			University of Pavia		
Overall TTR	6.70%			8.93%		
Criterion	OA	AA	κ	OA	AA	κ
SVM + SLIC + CRAS1	93.42	96.30	0.93	98.99	99.05	0.99
SVM + SLIC + CRAS2	95.82	97.48	0.95	99.49	99.26	0.99
KNN + SLIC + CRAS1	92.99	96.08	0.92	99.48	99.26	0.99
KNN + SLIC + CRAS2	95.61	97.35	0.95	99.69	99.37	1.00
MSSC-MSF	92.3	94.2	0.91	97.9	98.6	0.97
SVM-MSF + MV	91.8	94.3	0.91	91.1	94.8	0.88
SVM-MSF	88.4	91.6	0.87	84.14	92.35	0.80
HSeg + MV	90.8	94.0	0.90	93.9	97.0	0.92
EMP-KPCA	-	-	-	96.3	95.7	0.95

sity of Pavia data set, one more algorithm is included in the comparison, since its results are available in reference [1]. It is denoted by EMP-KPCA, which emphasizes on feature extraction. Assuming that all the results of the state-of-the-art algorithms in [1] are reliable, we directly adopt them in Table 3. We regard it necessary to compare the results of CRAS with the best ones of the reference methods. It would give a piece of strong evidence for the efficacy of CRAS if CRAS can outperform those methods.

To be fair, the training samples are chosen exactly as the way did by the articles presenting those methods. Take the Indian Pines data set as an example. 15 samples are randomly selected for each of the classes named “alfalfa”, “grass/pasture-mowed” and “oats” and 50 samples per each of the rest classes. Table 3 shows that either of CRAS1/2 excels the state-of-the-art methods, therefore validating the efficacy of the proposed methods. It should be noted that KNN + SLIC + CRAS2 gives a Kappa coefficient of 1.00 for the University of Pavia data set, although the OA is not 100%. This happens as every numerical result in Tables 2–4 is accurate to the second decimal place.

4.3.3 Parameters

This set of experiments is designed to provide recommendations for the parameter setting for CRAS. Two parameters, W_1 and W_2 are varied. The rest are set as default according to Table 1. Overall TTR equals 5.01%. Figure 3 illustrates the relationship of OA with W_1 , W_2 and R_s , whereas KNN and SLIC are employed. Here we chose KNN over SVM to avoid any effect on CRAS brought by the parameter tuning of the classifiers, since all the parameters of KNN are set by default ($K = 1$)⁹.

In Figure 3(a), when W_2 is fixed, CRAS1/2 gives higher OA in general as W_1 gets larger ($W_1 = 10, 20, \dots, 100, 200, \dots, 1000$). Either method produces better results when $W_2 = 50$. In Figure 3(b), CRAS1/2 performs better when $W_1 = 800$ rather than $W_1 = 80$. For all the algorithms and the values of W_1 , OA reaches the peak as $W_2 = 20$. Therefore, we can obtain similarly good results by letting W_1 be any arbitrary value ranging from 100 to 1000 and W_2 from 30 to 100 while complying with $W_1 \gg W_2 \gg 10$. In addition, according to (4), we have $W_1 \gg W_2 \gg C_m$.

Furthermore, Figure 3 proves that the efficacy of CRAS is insensitive to parameter variance. Although AS seems to involve many unsettled parameters, they can be simply left by default as suggested in Table 1. “Nice” or “poor” parameter choice does not make a huge difference in the final result. For example, the gap between the best OA and the worst is less than 1.5% while OA is higher than 93.5% for CRAS1, whereas TTR = 5.01%. However, the very best OA can be reached by the state-of-the-art algorithms in Table 3 is only 92.3% with an even higher TTR equaling 6.70%. Following experimental results show that the default setting in Table 1 can satisfy all the data sets being tested in the paper.

9) Available from MATLAB toolbox

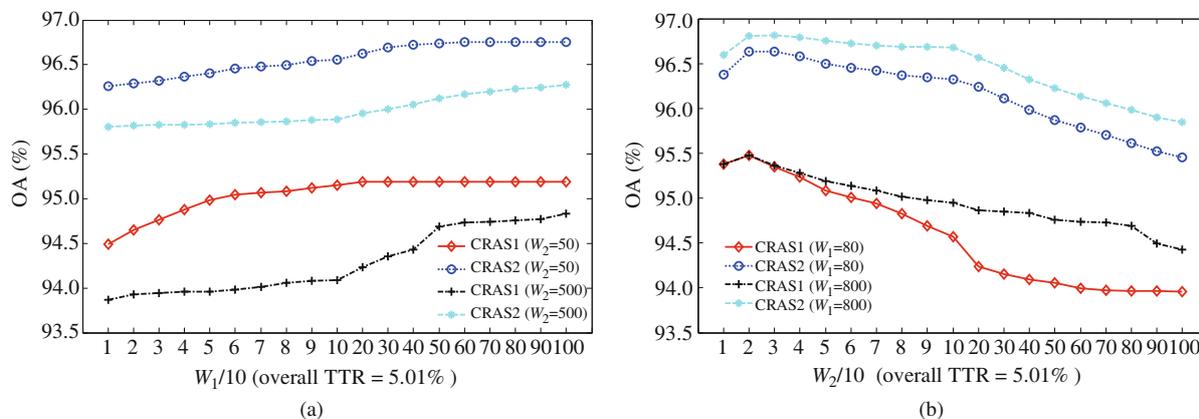


Figure 3 (Color online) Classification results (KNN + SLIC + CRAS1/CRAS2 applied on the Indian Pines data set, TTR = 5.01%) against parameters, whereas (a) W_1 and (b) W_2 .

Table 4 Classification accuracy corresponding to different algorithms

Data set	Salinas Valley						Kennedy Space Center					
	1.00%			5.00%			5.00%			9.99%		
	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ
SVM	88.47	92.21	0.87	91.75	95.02	0.91	90.41	92.21	0.89	93.36	95.02	0.93
SVM + SLIC + MV	90.41	93.73	0.89	93.36	96.18	0.93	90.41	93.73	0.89	93.36	96.18	0.93
SVM + SLIC + WMV	90.39	93.68	0.89	93.34	96.14	0.93	90.39	93.68	0.89	93.34	96.14	0.92
SVM + SLIC + CRAS1	95.31	96.76	0.95	99.19	98.98	0.99	97.27	96.39	0.97	99.49	99.32	0.99
SVM + SLIC + CRAS2	96.67	97.94	0.96	99.81	99.77	1.00	98.54	98.18	0.98	99.82	99.80	1.00
KNN	84.80	90.35	0.83	88.54	93.90	0.87	84.80	90.35	0.83	88.54	93.90	0.87
KNN + SLIC + MV	89.16	93.02	0.88	93.03	96.17	0.92	89.16	93.02	0.88	93.03	96.17	0.92
KNN + SLIC + WMV	88.82	92.87	0.88	92.75	96.04	0.92	88.82	92.87	0.88	92.75	96.04	0.92
KNN + SLIC + CRAS1	95.73	96.89	0.95	99.24	99.02	0.99	97.68	96.79	0.97	99.49	99.34	0.99
KNN + SLIC + CRAS2	97.37	98.15	0.97	99.82	99.77	1.00	98.93	98.56	0.99	99.88	99.90	1.00

4.4 Validation on the Salinas Valley and the Kennedy Space Center data sets

Experiments on the other two datasets validate the efficacy and the compatibility of the proposed algorithms. Apparently, CRAS1/2 outperforms MV/WMV, under all circumstances in Table 4. CRAS also gives robust performances, no matter when SVM or KNN is applied. In particular, the accuracy of the Kennedy Space Center data set is very high, due to the validation area provided by ground truth is limited.

Moreover, it must be emphasized that all the parameters are fixed according to Table 1 for every data set. Although the values of some parameters, such as P , R_s and R_e , are chosen empirically when the Indian Pines images are studied, they are obviously applicable to the other three different data sets. This point is made as CRAS in Table 2 and Table 4 excels the existing methods.

4.5 Summary of the experimental results

With the analysis of the experimental results, we can summarize the merits of CRAS as follows:

(1) CRAS is a very effective combination method as part of the general spectral-spatial classification framework for HSI. It excels the basic techniques as well as the state-of-the-art methods being compared with. The performance should be attributed to its ability to cope with two major challenges in the HSI classification area: the scarcity of prior knowledge and the serious inner-class spectral heterogeneity. CRAS can exploit local spatial consistency, finding a delicate balance between spatial homogeneity and spectral heterogeneity. Especially, CRAS2 is more advantageous in this aspect than CRAS1 is.

(2) CRAS can be combined with basic classification and segmentation algorithms. Since it adopts AS that exploits the fuzzy weights of both spectral and spatial information on the final classification accuracy, CRAS is robust to the errors in the preliminary classification and segmentation results. Consequently, simple and efficient techniques (e.g., SVM and SLIC) can be used as the other parts of the framework.

(3) Furthermore, CRAS can be considered implementation-friendly. Firstly, it is obvious that AS only involves scalar operations, as reflected by (2)–(4). Secondly, the protocol of the entire method is easy to follow, free of elusive calculation process and delicate parameter tuning, as exhibited in Algorithm 1. Thirdly, the computation complexity analysis shows that the number of multiplications involved is only the multiple of the total number of pixels to be classified. The heaviest computation step is the one involving pixel-wise similarity, which only needs to be performed once and stored for later use.

5 Conclusion

We have designed an accurate combination method (CRAS) based on fuzzy logic within the general framework of spectral-spatial classification. The core of the system is AS, which mainly considers three factors: local spatial consistency, spectral similarity and prior knowledge. We have offered two variables of the method, CRAS1 and CRAS2, according to different definitions of superpixel's neighborhood. Experiments have validated their efficacy on four different real hyperspectral data sets. The proposed algorithms outperform both of the basic and the state-of-art methods being compared with. Moreover, CRAS2 shows its advantage over CRAS1. In addition, the classification results have demonstrated the compatibility of CRAS. Therefore, we can use some default classifiers, over-segmentation tools and even the similarity measurement for most data sets.

Despite all the merits above, there is still room for improvement. As the preliminary classification engaged in the framework is essentially supervised, there needs to be a certain amount, though not great amount, of prior knowledge to ensure its accuracy. Hence, the classification results of CRAS would be compromised when the proportion of training samples was below the need. In the future, we will focus on developing semi-supervised classification methods based on AS, which is expected to require less training samples but produce even better results.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant No. 61572133), and Research Fund for State Key Laboratory of Earth Surface Processes and Resource Ecology (Grant No. 2015-KF-01).

Conflict of interest The authors declare that they have no conflict of interest.

References

- 1 Fauvel M, Tarabalka Y, Benediktsson J A, et al. Advances in spectral-spatial classification of hyperspectral images. *Proc IEEE*, 2013, 101: 652–675
- 2 Camps-Valls G, Tuia D, Bruzzone L, et al. Advances in hyperspectral image classification. *IEEE Signal Process Mag*, 2014, 31: 45–54
- 3 Khodadadzadeh M, Li J, Plaza A, et al. Spectral-spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization. *IEEE Trans Geosci Remote Sens*, 2014, 52: 6298–6314
- 4 Pu H Y, Chen Z, Wang B, et al. A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery. *IEEE Trans Geosci Remote Sens*, 2014, 52: 7008–7022
- 5 Chen Y, Nasrabadi N, Tran T. Hyperspectral image classification via kernel sparse representation. *IEEE Trans Geosci Remote Sens*, 2013, 51: 217–231
- 6 Huang X, Zhang L P. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Trans Geosci Remote Sens*, 2013, 51: 257–271
- 7 Pesaresi M, Benediktsson J A. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans Geosci Remote Sens*, 2001, 39: 309–320
- 8 Fauvel M, Chanussot J, Benediktsson J A. A spatial-spectral kernel-based approach for the classification of remote-sensing images. *Patt Recogn*, 2012, 45: 381–392
- 9 Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Patt Anal Mach Intell*, 2012, 34: 2274–2280

- 10 Yang H, Du Q, Ma B. Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery. *IEEE Geosci Remote Sens Lett*, 2010, 7: 875–879
- 11 Tarabalka Y, Benediktsson J A, Chanussot J, et al. Multiple spectral-spatial classification approach for hyperspectral data. *IEEE Trans Geosci Remote Sens*, 2010, 48: 4122–4132
- 12 Tilton J C, Tarabalka Y, Montesano P M, et al. Best merge region growing segmentation with integrated non-adjacent region object aggregation. *IEEE Trans Geosci Remote Sens*, 2012, 50: 4454–4467
- 13 Tarabalka Y, Chanussot J, Benediktsson J A. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Trans Syst Man Cybern B-Cybern*, 2010, 40: 1267–1279
- 14 Bernard K, Tarabalka Y, Angulo J, et al. Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach. *IEEE Trans Image Process*, 2012, 21: 2008–2021
- 15 Kang X, Li S, Benediktsson J A. Spectral-spatial hyperspectral image classification with edge-preserving filtering. *IEEE Trans Geosci Remote Sens*, 2014, 52: 2666–2677
- 16 Tilton J C. Image segmentation by region growing and spectral clustering with a natural convergence criterion. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Seattle, 1998*. 4: 1766–1768
- 17 Beaulieu J M, Goldberg M. Hierarchy in picture segmentation: a stepwise optimization approach. *IEEE Trans Patt Anal Mach Intell*, 1989, 11: 150–163
- 18 Plaza A J, Tilton J C. Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Seoul, 2005*. 7: 4946–4949
- 19 Tarabalka Y, Tilton J C, Benediktsson J A, et al. Marker-based hierarchical segmentation and classification approach for hyperspectral imagery. In: *Proceedings of International Conference Acoustics, Speech and Signal Processing (ICASSP), Prague, 2011*. 1089–1092
- 20 Tarabalka Y, Tilton J C, Benediktsson J A, et al. A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations. *IEEE J Sel Top Appl Earth Observ Remote Sens*, 2012, 5: 262–272
- 21 Widayati A, Verbist B, Meijerink A. Application of combined pixel-based and spatial-based approaches for improved mixed vegetation classification using IKONOS. In: *Proceedings of 23rd Asian Conference on Remote Sensing, Kathmandu, 2002*. 1–8
- 22 Briem G, Benediktsson J A, Sveinsson J R. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans Geosci Remote Sens*, 2002, 40: 2291–2299
- 23 Kittler J, Hatef M, Duin R P W, et al. On combining classifiers. *IEEE Trans Patt Anal Mach Intell*, 1998, 20, 226–239
- 24 Liu B, Hu H, Wang H. Superpixel-based classification with an adaptive number of classes for polarimetric SAR images. *IEEE Trans Geosci Remote Sens*, 2013, 51: 907–924
- 25 Jung M, Henkel K, Herold M, et al. Exploiting synergies of global land cover products for carbon cycle modeling. *Remot Sens Environ*, 2006, 101: 534–553
- 26 Dópido I, Villa A, Plaza A, et al. A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification. *IEEE J Sel Top Appl Earth Observ Remote Sens*, 2012, 5: 421–435