Pixel classification of large-size hyperspectral images by affinity propagation

Kacem Chehdi
Mariem Soltani
Claude Cariou

Pixel classification of large-size hyperspectral images by affinity propagation

Kacem Chehdi,* Mariem Soltani, and Claude Cariou
University of Rennes 1, IETR, TSI2M, Enssat - 6, rue de Kerampont, 22300 Lannion, France

Abstract. Affinity propagation (AP) is now among the most used methods for unsupervised classification. However, it has two major drawbacks: (1) the number of classes (NCs) is over-estimated when the preference parameter value is initialized as the median value of the similarity matrix; and (2) the partitioning of large-size hyperspectral images is hampered by its quadratic computational complexity. To overcome these two drawbacks, we propose an approach which consists of reducing the number of pixels to be classified before the application of AP. To reduce the number of pixels, the hyperspectral image is divided into blocks, and the reduction step is then independently applied within each block. This step requires less memory storage, since the calculation of the full similarity matrix is no longer required. AP is applied on the new set of pixels, which is then set up from the representatives of each previously formed cluster and nonaggregated pixels. To correctly estimate the NCs, we introduced a bisection method which aims to assess intermediate classification results using a criterion based on pixel interclass variance. The application of this approach on hyperspectral images shows that our results are efficient and independent of the block size. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JRS.8.083567]

1 Introduction

The interest in hyperspectral image data has been constantly increasing during the last years. Indeed, hyperspectral images provide more detailed information about the spectral properties of a scene and allow a more precise discrimination of objects than traditional RGB images or even multispectral images. High spatial and spectral resolutions of hyperspectral images enable one to precisely characterize the information pixel content. Though the potentialities of hyperspectral technology appear to be relatively wide, the analysis and the treatment of these data remain complex. In fact, exploiting such large datasets presents a great challenge and most methods of image analysis and interpretation require further development to tackle this situation. To be able to exploit hyperspectral data, classification is an essential step. This step can be done in a supervised or unsupervised manner. Unsupervised classification presents many advantages with respect to supervised classification: (1) the user does not need to state the classes to be discriminated nor any training samples. The classification algorithm automatically detects the distinct classes in an objective way, thus considerably reducing the risk of classification error; and (2) the access to training samples for some applications is very difficult. Consequently, supervised methods are not appropriate in this case.

Thus, in this paper, we are mainly interested in the unsupervised classification approach for the images' partitioning. The methods belonging to this type of approach can be divided into two groups: parametric and nonparametric methods. Parametric methods present many disadvantages compared with the latter. Among the disadvantages, we include the difficulty of accurate

---

*Address all correspondence to: Kacem Chehdi, E-mail: kacem.chehdi@univ-rennes1.fr

estimation of class-conditional parameters, which can highly distort the classification, especially in the case of large-size data as hyperspectral images. Nonparametric classification methods are based on some direct measure of similarity (or dissimilarity) between data points. The most well-known semi-supervised methods are K-means,[1] fuzzy C-means (FCM),[2] and ISODATA.[3] We consider that these three methods are semi-supervised since they require, at least, the number of classes (NCs) as a prior knowledge. In 2007, Frey and Dueck[4] proposed a new clustering algorithm named affinity propagation (AP), which can be used either in a semisupervised mode or a completely unsupervised mode. Its objective is to detect exemplars among data points and to form clusters of data points around these exemplars.

The AP algorithm has received much attention due to its three main advantages: (1) it is efficient, (2) it can be used in two modes: unsupervised or semi-supervised, and (3) it is insensitive to initialization. With these advantages, it has become widely used in many application areas such as remote sensing,[5–13] multimedia data management and pattern recognition, image categorization,[14,15] handwritten digits recognition,[16] extraction of key sentences,[17] visual query suggestion,[18] similarities of protein sets,[19] and gene data analysis.[20]

In the remote sensing domain, the AP clustering algorithm is used for feature selection or pixels clustering. It is mainly used to reduce the number of spectral bands or attributes in remote sensing and, more specifically, hyperspectral imaging. This is motivated by the fact that the number of spectral bands in hyperspectral imagery is limited to a few hundreds, which AP can easily handle despite its quadratic complexity. The first application of the standard AP to hyperspectral images for band selection was suggested in Ref. 13. In Ref. 5, Yao and Qian introduced a band selection method based on Gaussian processes. This method combines the standard AP (for band selection only) with a Gaussian processes classifier, which is a class of supervised probabilistic kernel-based learning algorithms. In Ref. 6, Qian et al. proposed an improvement of the method described in Ref. 13, by introducing the Kullback–Leibler divergence as the similarity between any two different bands as well as the kurtosis for the preference parameter (similarity between two identical bands or objects). In Ref. 7, Jia et al. introduced a feature extraction step using discrete wavelet transform before the feature selection step by AP. They also developed an unsupervised band selection method for hyperspectral imagery classification without manual band removal.[8] This method first uses wavelet shrinkage to remove the spatial noise in hyperspectral data, then applies AP in order to choose representative bands, and finally uses a classification algorithm (k-nearest neighbors or support vector machine) to assess the relevance of the selected bands. In Ref. 9, a semi-supervised method introducing a feature metric into AP as the criterion for spectral band selection was proposed.

In all these publications, AP is shown to provide the best results in band selection on a set of various hyperspectral images (from AVIRIS, HYDICE, and HYPERION sensors) with respect to several other approaches including maximum-variance principal component analysis, information divergence, and mutual information.

However, AP has been less used for pixel clustering in the remote sensing domain. This is due to the complexity of the algorithm, which cannot handle the huge amount of pixels provided by remote sensing sensors. AP has been used for the detection of regions of interest and the selection of representative landscapes using remote sensing, but only on small spatial size images. In Ref. 10, it is proposed to replace the conventional Euclidean distance by a fuzzy statistical (FS) similarity as the input to the AP algorithm (FS-AP). The method has been compared versus K-means, FCM, and the standard AP (with Euclidean distance metric) on three types of multispectral remote sensing images with a small size. The method FS-AP provides better results, improving the quality of classification with the standard AP, and with a lower computational load. In Ref. 11, Yang et al. proposed a semi-supervised AP clustering approach based on Incremental Decremental learning. It is applied on three different types of multispectral images for land cover classification and successfully compared with semi-supervised clustering algorithms: constrained K-means, incremental AP, but also maximum likelihood and Mahalanobis distance.

In Refs. 10 and 11, the proposed extensions of the AP algorithm are applied on the same set of multispectral images. It is worth mentioning that these images are of very small sizes (less than 100×100 pixels), which do not meet the requirements of the very large size images encountered with modern remote sensing sensors.

AP was also successfully used recently with another application related to the identification of representative landscapes of forest ecosystems in 10 forest areas of Canada from Landsat satellite imagery.[12]

Besides its application in remote sensing, the introduction of the AP concept in classification methods still has the potential to lead to better performances. In Ref. 16, an extension of the single-exemplar model of AP to a multiexemplar has been recently developed named ME-AP. This ME-AP determines the number of exemplars in each cluster associated with a super exemplar to approximate the subclasses in the category. It has been applied to three image databases for image categorization and two databases for handwritten digits. The results outperform exemplar-based clustering, kernel-based clustering, spectral clustering, multiprototype clustering, and hierarchical AP clustering. The results of this last algorithm are close to those of ME-AP. AP has also been combined with semi-supervised learning through Seeds AP (SAP).[21] This method is a semi-supervised version of the AP using an asymmetric similarity measurement that captures the structural information of texts and introduces a semi-supervised learning approach, where the knowledge from a few labelled objects versus a large number of unlabelled ones is exploited. SAP has been compared with K-means and AP on text data (with adapted similarity definitions) and has shown to perform better in terms of classification accuracy.

All studies cited in this section show the superiority of the AP concept compared with conventional methods, either for reducing attributes (or spectral bands) or for the classification of pixels. The superiority of the AP compared with other conventional clustering methods has already been shown in a comparative study on face recognition,[4] where AP has provided better classification rates than K-means, Kcenters, and hierarchical ascending classification. This superiority has also been confirmed, thanks to a comparative study which we have carried out with K-means, FCM, and ISODATA methods for the partitioning of synthetic images of data constructed by Rosenberger and Chehdi[22] from the Brodatz album[23] [average correct classification rate (ACCR): sum of correct classification rates per-class divided by the NCs: 95%].

Despite its high-correct classification rate, the application of AP on real, large-size image data like aerial images issued from a multispectral or hyperspectral imager remains intractable. For instance, under a regular PC (MATLAB environment), the number of data points which can be clustered by the AP algorithm is generally limited to less than 3000 pixels or objects. Furthermore, the full unsupervised version gives an over-estimation of the NCs.

That is why we suggest the solution to allow its application to the partitioning of large-size images, along with the estimation of the NCs. Our unsupervised method, namely a large spatial size-AP (LSS-AP), is composed of two steps: pixels' reduction to enable the use of AP in the case of LSS images and estimation of the correct NCs via optimization of the preference parameter of AP.

The remainder of the paper is divided into three sections. In Sec. 2, we will describe the unsupervised classification approach developed after an overview of the main steps of the AP algorithm. In Sec. 3, we show the experimental results obtained by our proposed method on two hyperspectral datasets. The first one is a synthetic image, and the second one illustrates a real application for the identification of invasive and noninvasive vegetation species from aerial hyperspectral images. The last section concludes this work and gives some perspectives.

2 Developed Classification Approach

In this section, we first present the AP classification method and highlight its two main disadvantages. Then, we describe the approach developed in order to overcome these drawbacks.

2.1 Classification with AP

Recently, Frey and Dueck[4] have suggested a new approach for classification called AP. This method has shown great success in different fields as we mentioned in Sec. 1. Its exploitation in different application fields is justified by the fact that it is totally unsupervised. In addition, it is deterministic and can use similarities that are not symmetric or do not satisfy the triangular inequality.[24]

2.1.1 *Reminder of the main stages of AP*

The classification via AP first requires the calculation of a similarity matrix S. The element $S(i,k)$ of this matrix indicates the similarity between pixels or objects (data points) i and k to be classified. Note that the diagonal elements $S(k,k)$ of matrix S are not computed in the same way as elements $S(i,k)$ for $i \neq k$. More precisely, $S(k,k) = p$ for all k (preference parameter) is initialized to the median value of the elements of S for $i \neq k$. Then, AP calculates for each pixel its degrees of availability and responsibility to the other pixels in an iterative way. Any type of similarities can be used, e.g., the opposite of the squared Euclidean distance[4] or the Manhattan distance [see Eq. 9] or other distances selected according to the application domain.

Initially, all pixels are considered as potential exemplars, though for each one, a preference parameter p value is allocated so that it can be chosen as an exemplar. Two procedures of message transmission called responsibility and availability are used to exchange messages between pixel i and a candidate exemplar k. The responsibility $R(i,k)$ [Eq. (2)] is the message sent from pixel i to candidate exemplar k, indicating how well-suited pixel k would be as the exemplar for pixel i. Alternatively, the availability $A(i,k)$ [Eq. (4)] is the message sent from candidate exemplar k to pixel i, indicating how likely pixel i would choose candidate k as its exemplar. This procedure identifies for each pixel the exemplar that maximizes the sum of responsibility and availability denoted by $E^*(i)$ [see Eq. (6)]. For pixel i, the value of k that maximizes $[R(i,k) + A(i,k)]$ either identifies i as an exemplar if $k = i$ or identifies the pixel that is the exemplar for pixel i. The message-passing procedure may be terminated either after a fixed number of iterations, after the changes in the messages fall below some threshold value, or when local decisions remain constant during some number of iterations. The updated messages [Eqs. (7) and (8)]) are damped by a constant factor λ, to avoid numerical oscillations that may arise under some circumstances. The value of this damping factor is defined in the interval [0,1].

Each iteration of AP consists of (1) updating all responsibilities given the availabilities, (2) updating all availabilities given the responsibilities, and (3) combining availabilities and responsibilities to monitor the exemplar decisions and to terminate the classification process.

The main steps of the algorithm are:

- *Initialization:*

    For N pixels to be classified, R, A, and S are the responsibility, availability, and similarity matrices of size $N \times N$, respectively.

$$R(i,k) = 0, \quad A(i,k) = 0 \quad \text{for all } i,k. \tag{1}$$

- *Responsibility updates:*

$$R(i,k) = S(i,k) - \max_{j,j \neq k}[A(i,j) + S(i,j)] \quad \text{for } i \neq k \tag{2}$$

$$R(k,k) = S(k,k) - \max_{j,j \neq k}[A(k,j) + S(k,j)]; \quad S(k,k) = p \quad \forall\, k. \tag{3}$$

- *Availability updates*:

$$A(i,k) = \min\left\{0, R(k,k) + \sum_{j,j \neq \{i,k\}} \max\{0, R(j,k)\}\right\} \quad \text{for } i \neq k \tag{4}$$

$$A(k,k) = \sum_{(j,j \neq k)} \max\{0, R(j,k)\}. \tag{5}$$

- *Making assignments*:

$$E^*(i) = \arg\max_k\{(R(i,k) + A(i,k)\}, \tag{6}$$

where $E^*(i)$ is the exemplar attributed to pixel i.

Updated messages are sent iteratively after regularization of responsibilities $\widehat{R}_l$ and availabilities $\widehat{A}_l$ as follows:

$$\widehat{R}_l = \lambda \widehat{R}_{l-1} + (1 - \lambda)R_l \tag{7}$$

$$\widehat{A}_l = \lambda \widehat{A}_{l-1} + (1 - \lambda)A_l, \tag{8}$$

where l is the current iteration.

2.1.2 *Main disadvantages of AP*

Despite the success of AP and its application to different fields as we mentioned in Sec. 1, it does present two major disadvantages which seriously hamper its usage.

- *The first drawback* is the limited number N of data points it can handle; this number cannot exceed some given value due to the necessary computation of the similarity matrix which has size $N \times N$. Practically, we observed that the image size allowed by the algorithm for an execution on a regular PC (MATLAB environment) should not exceed 3000 pixels. Wang and Zang[25] suggested a solution to reduce the image size by introducing a sampling step in the spatial dimension by choosing one pixel over two. Yet this solution is nonoptimal since the choice of pixels is made in an arbitrary way.
- *The second drawback* is the overestimation of the NCs given by the original unsupervised version of the algorithm.

The approach that we suggest in the following section allows the usage of AP in the case of large-size data and allows one to correctly estimate the NCs.

2.2 *Proposed Approach*

The proposed image partitioning approach using AP (namely LSS-AP) is composed of two steps:

Step 1: Reduction of the number of pixels or data points to be classified.
Step 2: Application of AP on the data retained after Step 1, with automatic estimation of the NCs via estimation of preference parameter p, which will be referred as to AP-modified in the following.

2.2.1 *Reduction of the number of pixels to be classified*

To be able to apply the AP method on images with large spatial dimensions, we have introduced a preliminary step prior to the classification in order to reduce the size of the similarity matrix. It consists of automatically grouping data points that are highly similar so as to not affect the data to be classified by AP and replacing each homogeneous group formed by a single representative. The criterion of aggregation used is the Manhattan (l_1) distance [see Eq. (9)] between each pair of pixels. To reduce the calculation time, the image is divided into blocks of size $N_B \times N_B$ pixels, and the search for most similar pixels is achieved in a parallel way on each block. This approach does not require the construction of the similarity matrix between all pixels of an image, contrary to the original AP. The procedure of reduction in each block is carried out in an iterative way. More precisely, the pixels which are spectrally identical are grouped during the first iteration. At this level, each subgroup of pixels formed is represented by a single pixel chosen randomly among them, since the spectral signatures are identical; then, from the second iteration, the pixel having the smallest distance from the center of gravity of its subclass is selected. Then at each iteration, matching is achieved for the set of pixels kept from the previous iteration by releasing the constraint made for the similarity criterion. For N pixels to be classified, this step groups a pixel i with a pixel k, presenting a minimum distance with respect to the set of remaining pixels. Then, if going through the remaining data, there exists some pixel j

at a distance to pixel k smaller than the distance to pixel i, so that the link between i and k is broken (i remains alone), and k is eventually grouped with j.

The reduction procedure is automated, since at the end of each reduction level the algorithm checks the maximum distance of aggregation that should not be exceeded. The value of the adaptive aggregation threshold δ_B for each block is estimated as follows:

Let N_B be the number of data points to be classified in a block B, and Sim_B be a set of size $(N_B^2 - N_B)/2$ representing the similarity values S_m between each pair of pixels within this block.

The similarity index between the pair of pixels i and k is given by

$$S_m(i, k) = \sum_{u=1}^{n} |i_u - k_u|, \tag{9}$$

where n is the number of spectral components, and i_u (respectively k_u) are the spectral values of pixel i (respectively k). Then, we proceed as follows:

- Calculation of the average value m_B and the standard deviation value σ_B of the set Sim_B.
- Calculation of the new set $Sim_B\prime \subset Sim_B$ having values within the interval: $[m_B - \sigma_B, m_B + \sigma_B]$.
- Estimation of the standard deviation of the new set $Sim_B\prime$, thus representing the threshold of aggregation δ_B of block B.

At the end of this step, in each block only one representative pixel for a group of pixels having a very high similarity with this representative and the pixels that have not yet been grouped are preserved. The clustering of the remaining pixels of all blocks by AP gives a global classification result for the full image, which is independent of the blocks' size, and more particularly, for the estimation of NCs. This is confirmed by the results shown later in Sec. 3.2.2.

2.2.2 *Classification with estimation of the NCs via AP*

As stated above, the NCs determined by AP are generally over-estimated. The estimated value is implicitly controlled by the preference parameter p (diagonal elements of the similarity matrix), which corresponds to the "self-attractive force" of an object. The smaller the preference value, the more the object loses its priority to be selected as an exemplar, whereas the higher its value (close to 0), the more likely it is to become an exemplar. In the original algorithm,[4] this parameter is set to the median value of the off-diagonal elements of the similarity matrix S. Under this setting, AP generates high NCs and consequently the correct classification rate on the data is decreased.

To correctly estimate the NCs, we introduced an evaluation criterion (EC) for the partitions obtained by AP (named AP-modified). This step involves the estimation of the preference parameter p. The best partition is the one which maximizes the EC based on the interclass variance defined by Levine and Nazif.[26]

To summarize, the sequential steps of AP-modified for partitioning are

- Execution of AP with $p =$ median value of the off-diagonal elements of S.
- Calculation of the responsibility matrix R at last iteration (R_f).
- Search for the value of p maximizing the EC in the range $[p_{\inf}, p_{\sup}]$ of the matrix R_f, using a bisection approach, with $p_{\inf} = \min R_f$ and $p_{\sup} = \operatorname{median} R_f$.

3 Assessment of Proposed Method

In this section, we first assess the effectiveness of each step of the proposed LSS-AP method (reduction process and classification by AP-modified) on a synthetic hyperspectral image constructed from the ground truth of a real-hyperspectral image. Afterward, we show the performance of the proposed method on a real application concerning the detection of invasive and noninvasive plant species. The data images used are first presented.

3.1 Presentation of Experimental Data

3.1.1 Synthetic hyperspectral image

A first evaluation is performed on a hyperspectral test image constructed from regions of interest of a real image for which the ground truth (spectral measurements and observations) is available. The original aerial image has a spatial size of 8075×9748 pixels and 62 spectral bands covering the visible–near infrared spectrum ([400,970] nm). It was acquired in the region of Murcia, Spain, on October 1, 2010, by means of the hyperspectral imager AISA Eagle available in our laboratory. The spatial ground resolution of this image is 0.5 m.

The size chosen for the test image is limited to only 64×64 pixels for two reasons: the first reason is the possibility of applying AP directly without any reduction step with the aim of analyzing the impact of reduction on the quality of the classification results. The second reason is the availability of ground truth data. Therefore, we selected five classes to build a patchwork image using the ground truth, as shown in Fig. 1. According to the ground truth class labels (masks) defined in Fig. 1(b), the data used for the construction of the test image [Fig. 1(c)] were randomly taken from areas 1, 2, 3, 4, and 5 of the original image [Fig. 1(a)]. These areas correspond, respectively, to the classes: river, *Pinus halepensis*, peach trees, *Arundo donax*, and buildings.

3.1.2 Real application data

To assess our approach on a real application, we selected regions of interest from the image of Murcia, Spain (8075×9748 pixels) for which the ground truth was available in six classes, among which three classes are invasive vegetation species (*Phragmites australis*, *Tamarix* and *A. donax*) as shown in Fig. 2. The objective of this application is to detect invasive plant species at an early stage in order to undertake appropriate subsequent management actions to limit their further development.

We selected a test area that contains the maximum of ground truth data to allow validation of our method. The size of the area is 1000×1000 pixels.

3.2 Assessment on Synthetic Data

It must be noticed that, in all experiments, the value of the damping parameter λ is set by default to 0.9 to avoid oscillations in the AP algorithm. When standard AP is used, the value of the preference parameter p is set to the median similarity value. The metric used in AP is the opposite of the Manhattan distance.
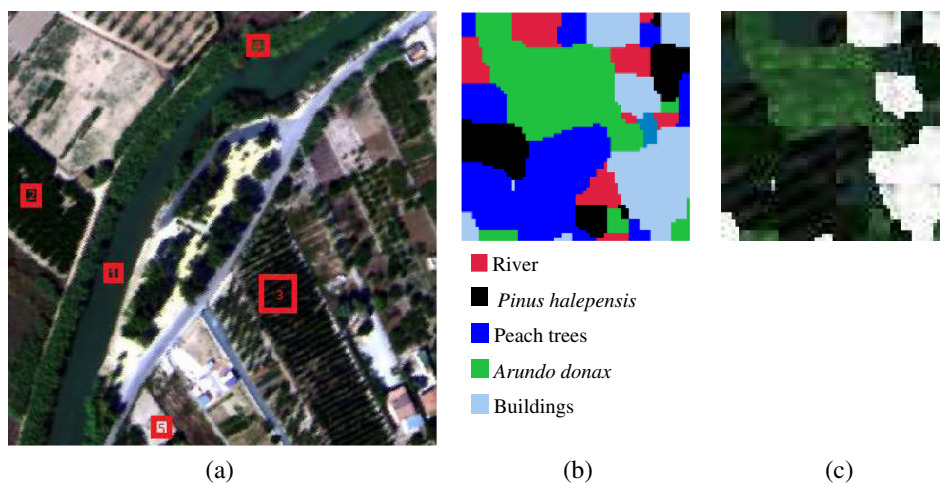

■ River
■ *Pinus halepensis*
■ Peach trees
■ *Arundo donax*
■ Buildings

(a)   (b)   (c)

Fig. 1 Original image and constructed image test. (a) Original hyperspectral image (62 bands) displayed in RGB (650, 550, 450) nm, 400×400 pixels. (b) Image of ground truth class labels, 64×64 pixels. (c) Hyperspectral test image displayed in RGB (650, 550, 450) nm, 64×64 pixels.

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)

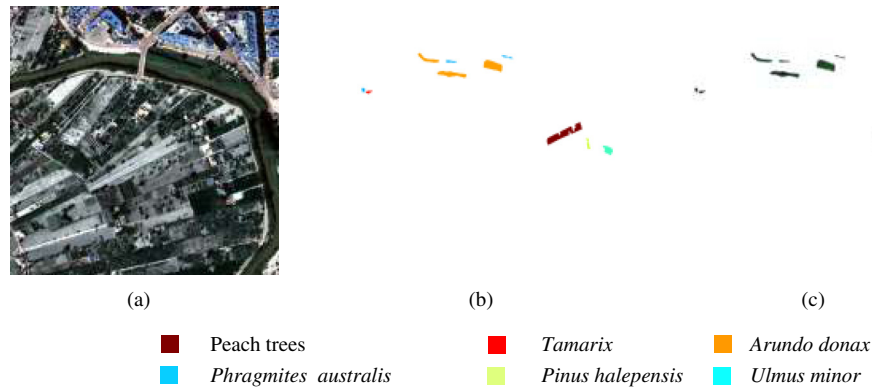|	
■ Peach trees	■ *Tamarix*	■ *Arundo donax*
■ *Phragmites australis*	■ *Pinus halepensis*	■ *Ulmus minor*

Fig. 2 Hyperspectral image (region of Murcia, Spain) acquired by the AISA Eagle sensor and mask of ground truth classes. (a) Original hyperspectral image (62 bands) displayed in RGB (650, 550, 450) nm (1000 × 1000 pixels). (b) Mask of ground truth classes. (c) Areas of the original image corresponding to the mask. Image displayed in RGB (650, 550, 450) nm.

3.2.1 *Analysis of the reduction step results*

The application of the proposed reduction step, associated with standard AP in the supervised mode (by setting the NCs to five), on the synthetic image, divided into four nonoverlapping blocks (32 × 32 pixels), gives an ACCR of 97.51%. This rate is superior to that obtained by the direct application of AP (without reduction) to the whole image (ACCR: 90.72%). In fact, in the reduction step, substituting a single representative to each group of strictly similar pixels disturbs the calculation of weights assigned to pixels by the AP less and, therefore, distorts the partitioning results less.

We compared our reduction process with the neuronal algorithm self-organizing maps (SOM),[27] which are often used as a reduction step prior to classification. Figure 3 shows the classification results of AP, without the preliminary reduction step [Fig. 3(a)], with the reduction by our approach [Fig. 3(b)], and with the reduction by SOM [Fig. 3(c)]. This experiment shows that our reduction approach provides a better ACCR (97.51% versus 95.36% with a reduction by SOM). The SOM algorithm is used here in the supervised mode, where the number of pixels is set the same as the one given by our reduction step. The result is even worse if SOM is used in an unsupervised mode (74.32%).

3.2.2 *Analysis of the AP-modified result*

Before applying our method of estimation of the NCs via AP, in this section we analyze the consequences of the choice of the preference parameter p value on the quality of the estimate of the NCs and, of course, on the quality of the classification result.

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)

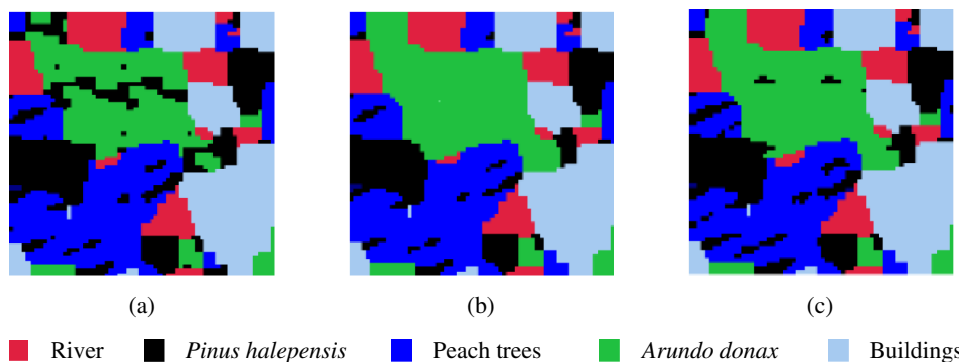■ River　■ *Pinus halepensis*　■ Peach trees　■ *Arundo donax*　■ Buildings

Fig. 3 Results of classification via standard affinity propagation (AP) by fixing the number of classes (NCs) to five. (a) Result without reduction (ACCR: 90.72%). (b) Result with reduction by the proposed approach (ACCR: 97.51%). (c) Result with reduction by SOM (ACCR: 95.36%).
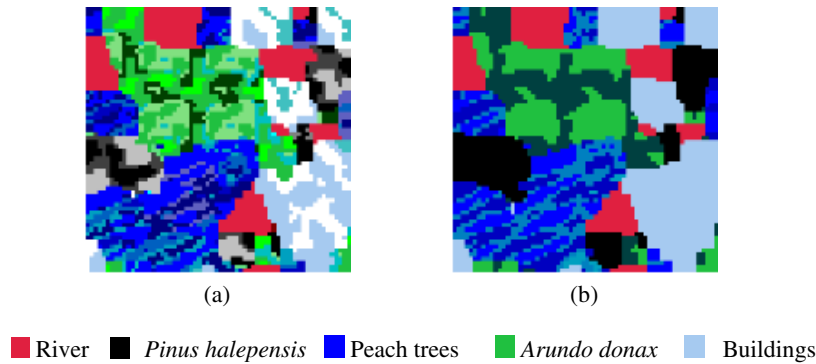
River   *Pinus halepensis*   Peach trees   *Arundo donax*   Buildings

Fig. 4 Classification results via standard AP. (a) Result without data reduction (p = median value of S). NC estimated by AP: 19. (b) Result with data reduction (p = median value of S). NC estimated by AP: 9.

Figure 4 shows the results of AP classification (unsupervised mode) with and without the preliminary reduction stage on the synthetic image. By fixing p to the median similarity value, the estimated NCs by the unsupervised AP is 19 without reduction, whereas it is reduced to 9 after application of the proposed reduction step on four blocks. After this step, the NCs were reduced, but are still biased.

When applying our approach of class number estimation after the reduction step, the correct number of five classes has been identified. Figure 5 shows the intermediate results of the estimation of the NCs, according to the evolution of the EC (search of the optimal value of p). Table 1 shows the confusion matrix of the optimal partitioning result, where the NCs were correctly estimated (five classes). Referring to the ground truth data, four classes (*P. halepensis*, *A. donax*, river, and buildings) were identified at a 100% rate and only one class (peach trees) was identified at an 87.55% rate.

These first tests show the relevance of the proposed approach (LSS-AP): a low classification rate error and an estimation of the NCs independent of block sizes. For example, the results obtained on the synthetic image, following three configurations (16 blocks of 16×16 pixels, 4 blocks of 32×32 pixels, and full image), confirms the exact NCs with a very small variation of ACCR (0.07% in average) every time. We also note that the number of pixels retained after the reduction step for the full image varies by the same order.

3.3 *Assessment on Real Application: Identification of Invasive and Noninvasive Vegetation Species*

This section presents the validation of our approach on the detection of three invasive plant species (*P. australis*, *Tamarix*, and *A. donax*). The test image of Fig. 2(a) (size
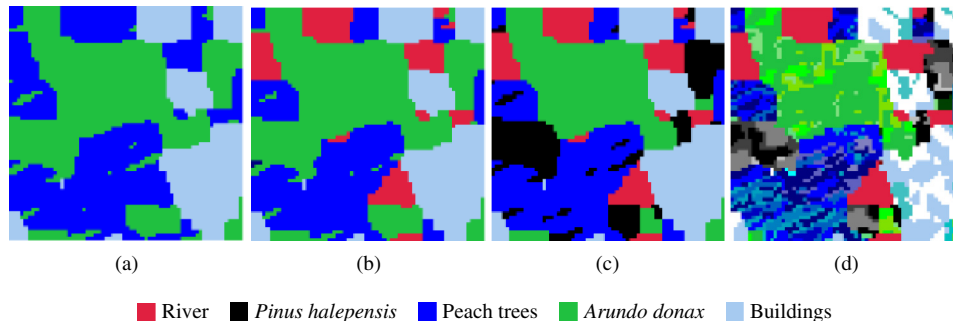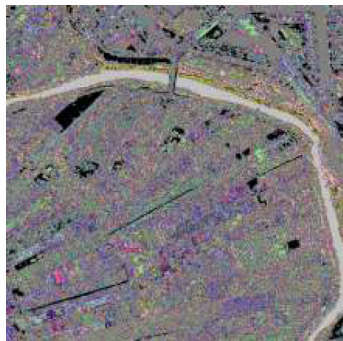

River   *Pinus halepensis*   Peach trees   *Arundo donax*   Buildings

Fig. 5 Partitioning of the hyperspectral image of Fig. 1(c) via LSS-AP method: estimation of the NC depending on the value of the parameter p and corresponding value of the EC. (a) $p = -6066\,300$ (minimum value of R_f matrix). NC estimated by AP: 3. EC = 0.59. (b) $p = -4563\,400$ (intermediate value). NC estimated by AP: 4. EC = 0.65. (c) $p = -1557\,700$ (intermediate value). NC estimated by AP: 5. EC = 0.69. (d) $p = -54\,806$ (median value of R_f matrix). NC estimated by AP: 23. EC = 0.29.

Table 1 Confusion matrix (five classes) corresponding to the result of classification in Fig. 5(c). %: CCR; (.): number of pixels (ACCR: 97.51%).

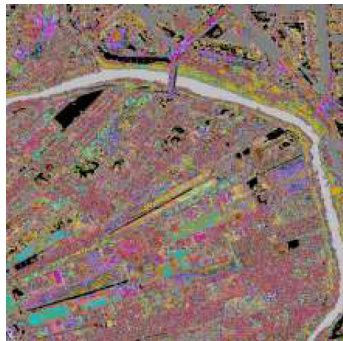| Classes predicted by the proposed approach (number of pixels)	Classes of ground truth (number of pixels)				
	River (452)	*Pinus halepensis* (500)	Peach trees (1189)	*Arundo donax* (1068)	Buildings (887)
River (468)	100% (452)	0	1.35% (16)	0	0
Pinus halepensis (630)	0	100% (500)	10.93% (130)	0	0
Peach trees (1041)	0	0	87.55% (1041)	0	0
Arundo donax (1070)	0	0	0.17% (2)	100% (1068)	0
Buildings (887)	0	0	0	0	100% (887)

p = -52 200 (minimum value of R_f matrix)
NC estimated by AP : 313   EC=0.31   NC estimated by AP : 68   EC= 0.25

p = - 28 500 (intermediate value)
NC estimated by AP : 41   EC=0.48   NC estimated by AP : 26   EC= 0.50

Arundo donax
Tamarix
Pinus halepensis
Ulmus minor
Phragmites australis
Peach trees

p = **- 66** (median value of R_f matrix)
NC estimated by AP : **35**   **EC= 0.71**   NC estimated by AP : **8**   **EC= 0.68**

Fig. 6 Partitioning the hyperspectral image of Fig. 2(a) based on the parameter p.

1000×1000 pixels, 62 spectral bands) has been divided into 400 blocks of size 50×50 pixels. Figure 6 shows the intermediate results of the estimation of the NCs, according to the evolution of the EC (search for the optimal value of p) and the corresponding partitioning results. The first column gives the partitioning result of the entire image, whereas for easier readability, the second column shows the result corresponding to the areas of interest only.

The optimal result which maximizes the EC is given for eight classes. Table 2 gives the corresponding confusion matrix. The results of the application of the proposed approach in areas of interest provided an ACCR of 97.66%. These results show that two of the invasive vegetation species (*Tamarix* and *A. donax*) have an ACCR of 100%, which confirms the result of Table 1. The ACCR for *P. australis* is 86.69%.

We also compared the performances of the proposed reduction step with the SOM approach on the same real hyperspectral image with an LSS. The selected pixels were classified by

Table 2 Confusion matrix corresponding to the results of optimal partitioning of areas of interest (eight classes). (%): CCR; (.): number of pixels (ACCR for six classes: 97.66%).

Classes predicted by the proposed approach (number of pixels)	Classes of ground truth (number of pixels)					
	Phragmites australis (556)	*Arundo donax* (4305)	*Tamarix* (162)	Peach trees (3115)	*Ulmus minor* (795)	*Pinus halepensis* (274)
Phragmites australis (482)	86.69% (482)	0	0	0	0	0
Arundo donax (4305)	0	100% (4305)	0	0	0	0
Tamarix (162)	0	0	100% (162)	0	0	0
Peach trees (3115)	0	0	0	100% (3115)	0	0
Ulmus minor (795)	0	0	0	0	100% (795)	0
Pinus halepensis (272)	0	0	0	0	0	99.27% (272)
Anonymous class 1 (8)	1.43% (8)	0	0	0	0	0
Anonymous class 2 (68)	11.87% (66)	0	0	0	0	0.72% (2)

Table 3 Performances comparison of three classification configurations on hyperspectral real image—Cieza site (1000×1000 pixels, 62 spectral bands).

	Proposed algorithm (step reduction + AP-modified)	SOM (supervised) + AP-modified	SOM (unsupervised) + AP-modified
Initial number of pixels to classify		1 000 000	
Number of pixels retained after step reduction	45 518	45 518	4943
NC estimated for all image	35	35	24
NC estimated on masks	8	21	11
ACCR	97.66%	47.20%	45.60%

AP-modified. Table 3 shows that the results of the reduction by SOM, either in supervised or in unsupervised mode, are not satisfactory. Indeed, in both cases, the results of classification by AP-modified give, respectively, an ACCR of 47.20% and 45.60%, against 97.66% with the LSS-AP. These results confirm those obtained on the synthetic image and show that the reduction with SOM gives very bad results when the image has an LSS.

Finally, we compared the classification results of the unsupervised AP-modified with classical methods often used in remote sensing: namely ISODATA and K-means. To be able to use these semi-supervised methods, a prior knowledge of the three following data is required: the NCs, a threshold value, and an iteration number. We specify that these two methods were applied on the data after the reduction step in the same conditions as the AP-modified. We first estimated the NCs with our approach, and then introduced it as *a priori* knowledge for the other two. The best correct classification rate for each method was obtained with threshold values fixed to 5% and three iterations. The performances (ACCR) of LSS-AP, K-means, and ISODATA are 97.66, 65.03, and 62.81% respectively.

This last experiment confirms the efficiency of our approach compared with the others for both the reduction step (with respect to SOM) and classification step (with respect to K-means and ISODATA).

4 Conclusion

In the present work, we have addressed the two main problems of AP, i.e., (1) the difficulty in handling datasets (or images) with a high number of points (or pixels), and (2) the difficulty in linking the preference parameter to the final NCs provided by AP. A preliminary step of the number of pixels reduced was proposed to answer the first problem. For the second problem, we have proposed using an automatic search of the optimal preference parameter to estimate the correct NCs present in an image. More precisely, the proposed nonparametric and unsupervised classification approach (LSS-AP) consists of first reducing the number of pixels to be classified on a metric-based procedure and second applying AP with a bisection method based on pixel interclass variance EC.

Our method was successfully applied to various hyperspectral images. The results show the robustness of our approach, which exploits the spectral signatures in a rather direct manner.

Experiments conducted on a synthetic hyperspectral image, and on a real application to identify invasive and noninvasive plant species from an LSS hyperspectral image, showed the effectiveness of our approach.

The proposed unsupervised reduction step compared with SOM in a supervised mode or unsupervised mode, both associated with AP-modified, showed its superiority. In addition, it is completely unsupervised. Also, compared with the traditional clustering ISODATA and K-means algorithms, which were applied on the data after our reduction step, experimental results show that the ACCR related to AP-modified is always more effective. These results confirm that the LSS-AP is effective for pixel classification of large-size hyperspectral remote sensing images.

The contribution of our approach is twofold: the possibility of applying AP on an LSS hyperspectral image and performance improvement with respect to classical AP in terms of the ACCR. Yet, in the case of strongly textured images, the exploitation of spectral and spatial information remains to be investigated.

Acknowledgments

References

1. J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, University of California Press, Berkeley (1967).

2. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell, Massachusetts (1981).
3. G. H. Ball and D. J. Hall, "Some fundamental concepts and synthesis procedures for pattern recognition preprocessors," in *Proc. Int. Conf. on Microwaves, Circuit Theory, and Information Theory*, Tokyo, Japan, pp. 281–297 (1964).
4. J. F. Frey and D. Dueck, "Clustering by passing messages between data points," *Science* **315**(5814), 972–976 (2007).
5. F. Yao and Y. Qian, "Band selection based gaussian processes for hyperspectral remote sensing images classification," in *Proc. 16th IEEE Int. Conf. on Image Processing (ICIP)*, pp. 2845–2848 (2009).
6. Y. Qian, F. Yao, and S. Jia "Band selection for hyperspectral imagery using affinity propagation," *IET Comput. Vision* **3**(4), 213–222 (2009).
7. S. Jia et al., "Feature extraction and selection hybrid algorithm for hyperspectral imagery classification," in *Proc. IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 72–75 (2010).
8. S. Jia et al., "Unsupervised band selection for hyperspectral imagery classification without manual band removal," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **5**(2), 531–534 (2012).
9. C. Yang et al., "A feature-metric-based affinity propagation technique for feature selection in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.* **10**(5), 1152–1156 (2013).
10. C. Yang et al., "A fuzzy-statistics-based affinity propagation technique for clustering in multispectral images," *IEEE Trans. Geosci. Remote Sens.* **48**(6), 2647–2659 (2010).
11. C. Yang et al., "Incremental and decremental affinity propagation for semisupervised clustering in multispectral images," *IEEE Trans. Geosci. Remote Sens.* **51**(3), 1666–1679 (2013).
12. J. A. Cardille et al., "Representative landscapes in the forested area of Canada," *Environ. Manage.* **49**(1), 163–173 (2012).
13. S. Jia, Y. Qian, and Z. J. Jia, "Band selection for hyperspectral imagery using affinity propagation," in *Proc. IEEE Digital Image Computer: Techniques and Applications*, pp. 137–141 (2008).
14. Y. Yimin et al., "A multimedia semantic retrieval mobile system based on HCFGs," *IEEE Multimedia* **21**(1), 36–46 (2014).
15. H. Xiao and P. Guo, "Iris image analysis based on affinity propagation algorithm," *Lect. Notes Comput. Sci.* **5552**, 943–949 (2009).
16. C.-D. Wang et al., "Multi-exemplar affinity propagation," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(9), 2223–2237 (2013).
17. Z. Liu et al., "Clustering to find exemplar terms for keyphrase extraction," in *Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Vol. 1, pp. 257–266 (2009).
18. Z. J. Zha et al., "Visual query suggestion," in *Proc. Association for Computing Machinery (ACM) Int. Conf. on Multimedia (MM)*, pp. 15–24 (2009).
19. K. Lindoff-Larsen and J. Ferkinghoff-Borg, "Similarity measures for protein ensembles," *PLoS ONE* **4**(1), e4203 (2009).
20. Z. Zhu, S. Jia, and Z. Ji, "Towards a memetic feature selection paradigm," *IEEE Comp. Int. Mag.* **5**(2), 41–53 (2010).
21. R. Guan et al., "Text clustering with seeds affinity propagation," *IEEE Trans. Knowl. Data Eng.* **23**(4), 627–637 (2011).
22. Ch. Rosenberger and K. Chehdi, "Unsupervised segmentation of multi-spectral images," in *Proc. Int. Conf. on Advanced Concepts for Intelligent Vision Systems*, Ghent, Belgium (2003).
23. P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover, New York (1966).
24. Y. Fujiwara, G. Irie, and T. Kitahara, "Fast algorithm for affinity propagation," in *Proc. of the Twenty-Second Int. Joint Conf. on Artificial Intelligence*, Vol. 3, pp. 2228–2243 (2011).
25. L. Wang and L. Zhang, "Color image segmentation algorithm based on affinity propagation clustering," in *Foundations of Intelligent Systems, Advances in Intelligent and Soft Computing*, Vol. 122, pp. 731–739, Springer-Verlag, Berlin, Heidelberg (2011).

26. M. D. Levine and A. M. Nazif, "Dynamic measurement of computer generated image segmentations," *IEEE Trans. PAMI* **PAMI-7**(2), 155–164 (1985).
27. T. Kohonen, "The self-organisation maps," *Proc. IEEE* **78**(9), 502–512 (1990).

Kacem Chehdi received his PhD and the Habilitation à Diriger des Recherches degrees in signal processing and telecommunications from the University of Rennes 1, France, in 1986 and 1992, respectively. Since 1993, he has been a professor at the same institution. He is currently the head of the Signal and Multicomponent/Multimodal Image Processing Laboratory. His research activities concern adaptive processing (blind restoration and blind filtering) and pattern recognition. He is a member of SPIE.

Mariem Soltani is a PhD student at the University of Rennes 1, Lannion, France. She received her engineering and master's degrees in computer sciences and image processing from the National School of Computer Sciences, Tunisia, in 2009 and 2010, respectively.

Claude Cariou received his PhD degree in electronics from the University of Brest, France, in 1991. Since 1992, he has been assistant professor at the Ecole Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT)—University of Rennes 1, Lannion, France. His research interests include image analysis, pattern recognition, unsupervised classification, texture modeling and segmentation, image registration, and feature selection, especially with application to multi- and hyperspectral remote sensing imageries.