This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
Unsupervised methods for the classification of hyperspectral images with low spatial resolution

A. Villa, J. Chanussot, J.A. Benediktsson, C. Jutten, R. Dambreville

1 The work was in part carried out while A. Villa was a PhD student at the Gipsa-Lab and the University of Iceland.

Keywords: Unsupervised classification, Spatial resolution improvement, Hyperspectral data, Source separation, Spatial regularization

1. Introduction

Target and structure detection and image classification are two important applications of pattern recognition, used in many application domains, such as biomedical imaging and remote sensing [23,8,32,13]. In these fields, low image spatial resolution can highly affect the performance of the processing algorithms. Recent advances in sensor technology have made available images characterized by a very detailed spectral information on a wide spectral range, particularly suitable for these applications [24]. Hyperspectral images are composed of hundreds of bands with a very high spectral resolution, generally from the visible to the infra-red region. For every recorded pixel, the rich spectral information provides a complete spectral description and a better characterization of the observed surface, thus resulting in a very powerful tool for materials discrimination. Hyperspectral imaging is receiving continuously growing attention, especially in the remote sensing community, due to the advantageous characteristic of high spectral resolution data and to the already planned civilian space missions which will make available in the next future a huge quantity of hyperspectral data (amongst the others, PRISMA, planned by the Italian Space Agency ASI in 2014, EnMap planned by German Space Agency DLR in 2014, Hyper-J and HyspIRI, planned by the Japan Space Agency and NASA, respectively, in the next future, besides the already operating sensors like the widely used Hyperion and AVIRIS [38], both of NASA). However, a common drawback of hyperspectral sensors is the relatively low spatial resolution, which can vary from few to tens of meters, especially in the cases of high altitude sensors or instruments covering wide areas. There are many factors (such as imperfect imaging optics, atmospheric scattering, secondary illumination effects and sensor noise) that degrade the acquired image quality and make the development of new technology to improve the spatial resolution a very challenging task [11].

In the cases of structure detection and land cover classification, a low spatial resolution leads to the problem of mixed pixels, e.g., pixels containing mixture of different materials [28]. The usual assumption that every pixel of the image can be associated with a unique class label is no longer verified, and mixed pixels
Several techniques have been proposed recently to deal with the problem of mixed pixels and low spatial resolution of remote sensing images [12,17,2,14,31,20,29,36]. These techniques can be divided into three main groups. The first group includes techniques which use high spatial resolution images jointly with the low resolution images, in order to obtain a fused image with high spectral and spatial resolutions [12,17]. The resulting image can afterwards be used as input for classification. The main drawback of this approach, besides the need for an accurate coregistration of the two images, is the need for ancillary data, (e.g., a high spatial resolution image acquired over the same area), in general not easy to obtain.

The second group of techniques comprises super-resolution approaches independent from any high spatial resolution data. Tatem et al. propose an algorithm based on the Hopfield neural network [37] which does not need any secondary source of data to realize the super-resolution mapping. This method has a good efficiency but suffers from high computational cost. Gu et al. tackle the problem of high computational cost through a fast learning-based algorithm to integrate the spatial and spectral information of hyperspectral images, back propagation neural network and some ground truth information which is unassociated to the considered test data [16]. The output of such techniques is a series of abundance maps at a higher resolution. For each class considered, an abundance map is created.

Finally, several methods which classify images assuming the possibility of mixed pixels were proposed in recent years. Examples of such techniques are soft classification algorithms [29], which provide a set of images (one per class) expressing the degree of membership to the class, and linear Spectral Mixture Analysis (SMA) [20,35,39], which assumes every pixel to be the weighted sum of some constituent spectra, also called endmembers. Similar methods have been proposed for sub-pixel image labeling, based on the latest development of machine learning [5,7]. These techniques can partially overcome the weakness of full pixel classification methods when analyzing mixed pixels, and they are suitable to be used for the analysis of these scenarios. However, the final output is a classification map (or a series of maps) representing the membership degree (or the abundance) of each pixel with respect to a class. When trying to obtain a crisp output, the additional information provided by fuzzyness is lost. A scheme representing the behavior of different methods in the case of an image containing mixed pixels is shown in Fig. 1.

The work presented in this paper tries to tackle the problem of mixed pixels from a different viewpoint, in order to consider as input a Hyperspectral Image (HSI) with a given spatial resolution and obtain a thematic map where the distribution of the classes is depicted in a classification map with finer spatial resolution. The main novelty of the work is the proposal of a methodology able to provide thematic maps at a finer spatial resolution with respect to the analyzed image, in a totally unsupervised way. Methods previously proposed in the literature (such as spectral unmixing and soft classifiers) are able to retrieve the proportions of each pure class within a pixel, but they cannot obtain a classification map at a higher resolution. For example, VCA gives as output the abundance maps of the endmembers, but the spatial resolution of this map is the same as the original input image. If a resolution improvement is desired, further processing is needed.

The main principle is to sufficiently mine the data advantages of HSI by spectral unmixing and super-resolution mapping and to integrate the spectral and spatial information for resolution enhancement. One advantage of the proposed method is that no supplementary source associated with HSI is needed. The idea of the proposed method is to provide a unifying framework, able to address the problems highlighted here above, considering the rich information provided by hyperspectral data and exploiting the characteristics of both unsupervised classifiers and spectral unmixing to address the quantification of pure classes within mixed pixels, coupled with a spatial regularization which aims at correctly locate sub-pixels from a spatial viewpoint. According to the authors best knowledge, this is the first time that a similar unsupervised technique is proposed. Besides the novelty of the method, a careful investigation of the variables having an influence on the final classification accuracy is conducted, with particular attention to the well known problem of spectral variability. Preliminary results of this work were presented in [39], and encouraged to further develop the method.

The remainder of the paper is as follows. In Section 2, the general framework of the proposed method is presented, and spectral unmixing is analyzed in detail. Section 3 is focused on the different techniques used to obtain a map resolution improvement. Section 4 is devoted to a toy experiment with synthetic data, where the effectiveness of the method is tested on a scene where all the ground truth is known in detail. Experiments on two real hyperspectral data sets are discussed in Section 5, while the conclusions are finally drawn in Section 6.

2. Spectral unmixing

The general scheme of the proposed methodology is shown in Fig. 2. The first step is the determination of the classes within the image. Two approaches are proposed in this work, based respectively on source separation and unsupervised clustering. The reason for choosing two different methods is shown in Fig. 3,
where each pixel is represented by a vector \( \mathbf{x} \) denotes the spectral response of endmember \( z \), \( \Phi_z \) is a scalar value designating the fractional abundance of the endmember \( z \) at the pixel \( \mathbf{x} \), \( p \) is the total number of endmembers and \( \mathbf{w} \) is a noise vector.

A number of techniques have been recently proposed to retrieve endmembers by mean of source separation [33].

In [30], the Vertex Component Analysis (VCA) is proposed as an effective method for extracting the endmembers which are linearly mixed. VCA makes use of the concept of orthogonal projection. Algorithms based on this concept, start by selecting the pixel vector with maximum length in the scene as the first endmember. Then, they look for the pixel vector with the maximum absolute projection in the space orthogonal to the space linearly spanned by the initial pixel and labels that pixel as the second endmember. A third endmember is found by applying an orthogonal subspace projector to the original image, where the signature that has the maximum orthogonal projection in the space orthogonal to the space linearly spanned by the first two endmembers. This procedure is repeated until the desired number of endmembers \( p \) is found.

VCA, as opposed to the methodology previously described, exploits the fact that the endmembers are the vertices of a simplex and that the affine transformation of a simplex is also a simplex. As a result, VCA models the data using a positive cone, whose projection onto a properly chosen hyperplane is another simplex whose vertices are the final endmembers. After projecting the data onto the selected hyperplane, the VCA projects all image pixels to a random direction and uses the pixel with the largest projection as the first endmember. The other endmembers are identified in sequence by iteratively projecting the data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember is then selected as the pixel corresponding to the extreme projection, and the procedure is repeated until a set of \( p \) endmembers is found [30].

In our work, we have selected VCA due to its good performance and the very low computational burden.

2.2. Clustering based technique

Endmember extraction techniques are an easy way to retrieve endmembers, especially in the case of images comprising mixed pixels. However, these techniques are in general sensible to 'outliers', e.g., isolated pixels with anomalous values of reflectance, which are detected as extreme pixels and therefore endmembers. For such a reason, the second proposed technique is an extension of a traditional unsupervised classifier (the \( K \)-means classifier) as a method to detect classes composing the image, which is expected to be less sensitive to the issues related to the presence of outliers. Given a set of observations \( x_1, x_2, \ldots, x_n \), where each observation is a \( d \)-dimensional real vector, \( K \)-means clustering aims to partition the \( n \) observations into \( p \) sets so as to minimize the within-cluster sum of squares

\[
\min_S \sum_{i=1}^{n} \sum_{x \in S_i} ||x - \mu_i||^2
\]  

where \( \mu_i \) is the mean of points in the cluster \( S_i \).

The \( K \)-means classifier is first applied to the image. At the end of the classification process, the centroids of the classes found by the algorithm are retained as constituent spectra of the image.

Once the endmembers are extracted from the image, the abundance fractions of the elements within each pixel should be determined. Several algorithms have been developed to handle the linear mixing model according with the required physical constraint of abundance fractions, which are non-negativity (all the abundances must be greater than or equal to zero) and full

---

**Fig. 2.** Block diagram of the proposed approach. In a first step, thematic classes are identified through endmember extraction or unsupervised classification. Spectral unmixing is used to compute abundances of classes within each pixel. After splitting pixels in a number of sub-pixels and assigning them to a class according to the results of unmixing, a spatial regularization is performed to obtain the final map.
additivity (the sum of the endmember abundances within a pixel should be equal to one). Due to the efficiency from a computational point of view, a common choice is to use a Fully Constrained Least Squares (FCLS) algorithm, which satisfies both abundance constraints and is optimal in terms of least squares error [19].

When applied to real problems, the main problem is that the FCLS does not have a closed-form mathematical solution due to the non-negativity constraints; thus, a numerical solution is always required. To calculate the FCLS solution, the non-negativity constraint is considered first. The idea is to minimize the non-negative abundance values, which is mathematically expressed as
\[
\min(z - Sa)^T(z - Sa)
\]
where \( z \) is the observed value which should be represented by the product \( Sa \). By using Lagrange multipliers, a Lagrangian \( J \) is defined as
\[
J = \frac{1}{2}(z - Sa)^T(z - Sa) + \lambda^T(a - c)
\]
where \( a = c \), each member of the unknown constant \( L \times 1 \) vector \( c \) is non-negative to enforce the non-negativity constraint, and \( \lambda \) is the Lagrange multiplier denoted by an \( L \times 1 \) vector. The defined equation (3) allows the use of Lagrange multipliers because the non-negativity constraint has been substituted by equality constraints with the unknown vector \( c \). To calculate the estimate of \( a \), we take the partial derivative of \( J \) with respect to \( a \). Eq. (4) contains two unknown parameters, i.e., the abundance estimates and the Lagrange multipliers. Solving for these unknown parameters results in
\[
\hat{a} = (S^T S)^{-1}S^T (z - Sa), \quad \lambda = S^T(z - Sa)
\]
By iterating through (5) and (6), the numerical solution is provided for the non-negativity constraint. To begin this iterative method, we set all the Lagrange multipliers to zero and calculate the abundance. Note that this initial calculation is the unconstrained least squares solution for the abundance values. From this solution, we find those abundance values which are greater than zero and put them into the passive set \( P \). The remaining non-positive abundance values are placed in the active set \( R \). Eqs. (5) and (6) are iterated until all Lagrange multipliers in the passive set are zero and all Lagrange multipliers in the active set are either zero or negative. At this point, the Kuhn–Tucker conditions have been met, and an optimal solution for the abundance values has been found [18]. It should be noted that this solution only accounts for the non-negativity constraint. To handle the sum-to-one constraint, an easy modification of the aforementioned algorithm was developed to retain the optimality guaranteed under the Kuhn–Tucker conditions for numerical optimization on a finite computing machine. In the modification, the endmember matrix and pixel signatures are extended such that
\[
\tilde{S} = \begin{bmatrix} \delta S \\ 1^T \end{bmatrix}
\]
(7)
is the new endmember matrix and
\[
\tilde{a} = \begin{bmatrix} \delta z \\ 1 \end{bmatrix}
\]
(8)
is the new pixel signature, where \( \delta \) is a constant (typically, \( 1 \times 10^5 \)) [19]. The variable \( \delta \) controls how tightly the solution will sum to one so that the smaller values provide a better solution, but it may need a longer convergence time. The new endmember matrix and pixel signature are then used in (7) and (8) to obtain an abundance solution that subjects to both the non-negativity and sum-to-one constraints simultaneously.

The solution obtained by the FCLS algorithm is the optimal one. Using that algorithm, the abundance of each endmember in each pixel can be obtained. These abundances of the whole image and their spatial positions will be the inputs of super-resolution mapping step.

3. Improving spatial resolution

Spectral unmixing is useful to describe the scene at a sub-pixel level, but can only provide information about proportion of the endmembers within each pixel. Since the spatial location remains unknown, spectral unmixing does not perform any resolution enhancement. Here, we investigate two super-resolution mapping techniques, which take advantage of the information given...
spectral mixing analysis and use it to enhance the spatial resolution of thematic maps. The general scheme is shown in Fig. 4.

First, we set an abundance threshold to determine if a pixel can be considered as ‘pure’. Since a single spectral signature cannot represent extensivel y a class within the whole image, the abundance determination is negatively affected by spectral variability. Therefore, all the pixels with maximum abundance greater than this threshold are considered as entirely belonging to the corresponding class. The other pixels are considered as mixed. Then each pixel is split into a fixed number of sub-pixels, according to the desired resolution enhancement. Every sub-pixel is assigned to a class, in conformity with the fractional abundance computed in the first step. For each pixel, the number of sub-pixels n to assign to the class i is computed according to the equation

\[ n_i = \text{round} \left( \frac{\text{abd}_i}{1/N} \right) \]  

where \( \text{abd}_i \) is the fractional abundance of the class i within the considered pixel estimated with the FCLS, N is the total number of sub-pixels within each pixel, and \( \text{round}(x) \) returns the value of the closest integer to x.

As mentioned before, a well known problem of hyperspectral images source separation and spectral unmixing covering wide areas is that abundances determination is negatively related to the intra-class spectral-variability [15], and the assumption that a single endmember could extensively represent a class is generally far from reality. In order to investigate the influence of spectral variability on the final results, we have tested two possible approaches. When using an endmember extraction algorithm to retrieve the endmembers, the spectral signatures retrieved are used as ‘endmembers’ in the whole image, since such techniques retrieve the purest pixel of the images, which could be used in the whole data set to determine abundances within mixtures. In the case of endmembers that are extracted using clustering techniques, it is computed a preliminary ‘classification map’, where only the purest pixels are labeled (a pixel is considered as pure if its maximum abundance is higher than a chosen threshold). Then, the abundances of unlabeled pixels are re-computed by considering as ‘endmember candidates’ only a number of pixels in the spatial neighborhood of the considered mixed pixel. The endmember candidates are therefore chosen among the pixels labeled as ‘pure’ in the first step, in order to use local endmembers and to handle the problem of spectral variability. This is done because endmember retrieved by clustering techniques are more likely to be mixed or slightly mixed pixels, which cannot represent the correspondent class all in the whole image.

3.1. Simulated Annealing

The first approach to locate the sub-pixels is based on a Simulated Annealing mapping function. The algorithm is used to create random permutation of these sub-pixels, in order to minimize a chosen cost function. Relying on the spatial correlation tendency of landcovers, we assume that each endmember within a pixel should be spatially close to the same endmembers in the surrounding pixels. Therefore, the cost function C to be minimized is chosen as the perimeter of the areas belonging to the same class

\[ C = \sum_{i=1}^{I} \sum_{j=1}^{C_i} P_j \]  

where I is the number of the classes, \( C_i \) is the number of connected components of the class i, and \( P_j \) is the perimeter of the connected component j, computed according to the 8-connected border pixels model [22]. Simulated Annealing (SA) is a well established stochastic technique originally developed to model the natural process of crystallization [27]. This process is based on an analogy from thermodynamics where a system is slowly cooled in order to achieve its lowest energy state. More recently, SA has been proposed to solve global optimization problems [21], and it has been used in various fields.

The basic idea of the method is that, in order to avoid to be trapped in local minima, uphill movements, i.e., the points corresponding to worse objective function values could, sometimes, be accepted by the following iterative procedure. As with a greedy search, it accepts all changes that lead to improvements in the fitness of a solution. However, also changes which lead to worse solutions can be accepted. The probability of accepting a reversal is inversely proportional to the size of the reversal with the acceptance of smaller reversals, being more probable. This probability also decreases as the search continues, or as the system cools, allowing eventual convergence on a solution.

3.2. Pixel Swapping

The second algorithm investigated is the so-called Pixel Swapping algorithm, based on the concept of sub-pixels attractiveness introduced by Atkinson [3,4]. The original algorithm takes as input the sub-pixels hard land cover maps obtained after converting the fractional abundances of the classes and try to maximize the spatial correlation of same class sub-pixels while preserving the proportional composition within a “low-resolution” pixel. With respect to the original algorithm, two main differences are presented here:

- In Atkinson [3], the fractional abundances are considered as given information, while here we compute them in the first step of the proposed algorithm.
- The original algorithm was proposed for two-class problems. Here, we apply it to a multi-class classification problem.

The pixels swapping algorithm takes in consideration a weighted function representing the ‘attractiveness’ of a sub-pixel location.
For each sub-pixel position \( i \) within the pixel, the attractiveness \( O_i \) is represented by

\[
O_i = \sum_{j=1}^{g} \lambda_{ij} Z(X_j)
\]

(11)

where \( g \) is the number of neighboring pixels considered, \( Z(X_j) \) is the binary value of the class \( z \) in the \( j \)-th sub-pixel location \( X_j \) (1 if the sub-pixel belongs to the class \( z \), 0 otherwise), and \( \lambda_{ij} \) is a weight computed as

\[
\lambda_{ij} = \exp\left(-\frac{h_{ij}}{z}\right)
\]

(12)

where \( h_{ij} \) is the distance between the sub-pixel locations \( i \) and \( j \) and \( z \) is a range parameter of the exponential model. Several weighting functions were explored in [25] as possible alternatives to the exponential function. The authors suggested that a simple Nearest Neighbor model could provide comparable accuracy of more complex spatial models, such as the exponential weighting function, with a much simpler model. Therefore, (11) becomes the simple sum of the values in the nearest sub-pixels positions

\[
O_i = \sum_{j=1}^{g} Z(X_j)
\]

(13)

The method can be described as follows: within a pixel, for each class, the attractiveness value \( O \) is computed for each sub-pixel position \( i \) and class \( z \). If the least attractive value of a sub-pixel actually belonging to a class \( z \) is smaller than the most attractive value of a sub-pixel belonging to another class, the two sub-pixels are swapped. This procedure is repeated either for a previously fixed number of times or until a stopping criterion is reached.

### 4. Experiments on synthetic data

This section aims at giving an overall idea of the proposed method, when used for classifying synthetic data. Because of this reason, only the endmember extraction based method is considered, with Simulated Annealing sub-pixel mapping.

One of the main problems when dealing with spectral unmixing of real data is the difficulty to assess the results obtained, especially in the case of abundance fractions estimation. If reference spectra can be easily obtained from a laboratory especially in the case of abundance fractions estimation of real data is the difficulty to assess the results obtained, with Simulated Annealing sub-pixel mapping.

For each sub-pixel position \( i \) within the pixel, the attractiveness \( O_i \) is represented by

\[
O_i = \sum_{j=1}^{g} \lambda_{ij} Z(X_j)
\]

where \( g \) is the number of neighboring pixels considered, \( Z(X_j) \) is the binary value of the class \( z \) in the \( j \)-th sub-pixel location \( X_j \) (1 if the sub-pixel belongs to the class \( z \), 0 otherwise), and \( \lambda_{ij} \) is a weight computed as

\[
\lambda_{ij} = \exp\left(-\frac{h_{ij}}{z}\right)
\]

where \( h_{ij} \) is the distance between the sub-pixel locations \( i \) and \( j \) and \( z \) is a range parameter of the exponential model. Several weighting functions were explored in [25] as possible alternatives to the exponential function. The authors suggested that a simple Nearest Neighbor model could provide comparable accuracy of more complex spatial models, such as the exponential weighting function, with a much simpler model. Therefore, (11) becomes the simple sum of the values in the nearest sub-pixels positions

\[
O_i = \sum_{j=1}^{g} Z(X_j)
\]

The method can be described as follows: within a pixel, for each class, the attractiveness value \( O \) is computed for each sub-pixel position \( i \) and class \( z \). If the least attractive value of a sub-pixel actually belonging to a class \( z \) is smaller than the most attractive value of a sub-pixel belonging to another class, the two sub-pixels are swapped. This procedure is repeated either for a previously fixed number of times or until a stopping criterion is reached.

Despite most of these spectra are not closely related, some of them have a very high spectral correlation making more difficult the source separation process.

In this simulation, we have tried to perform an accurate simulation of data containing spatial characteristic close to reality. In order to create a realistic fractional abundance map, we have considered the ground truth of a widely used image in remote sensing application, the AVIRIS Indian Pine image.3 By considering the reference map of the AVIRIS data as ground truth of our simulated scene, we have substituted the original spectral values with those obtained from the USGS library spectra. Therefore, every spectrum will have a spatial distribution of a land cover class of the AVIRIS image (or, in some cases, more than one, since the AVIRIS image contain 16 different classes, while in our experiment we have only nine). No spectral variability is considered in this first experiment.

After creating the synthetic image of 144 × 144 pixels which will be used later to assess the results of our method, we perform a down-scaling by substituting each 3 × 3 window of pixels with its average value, obtaining a new image composed of 48 × 48 pixels, with the same number of bands but a lower spatial resolution (of course, being a simulated image, the spatial resolution is relative; however, we obtain an image covering the same area where mixture of pixels are incorporated). This image will be the input data of the proposed approach, to improve the spatial resolution of its classification map. The ground truths of the original and filtered images are shown in Fig. 6(a) and (b). In the image, a majority of pure pixels are present, but it also contains a number of areas with mixtures of materials (totally the low resolution image contains 1722 pure and 582 mixed pixels). Due to the lack of spectra variability this is not a particularly challenging scenario, but still it can show the possibility offered by the proposed method.

Results are shown in Fig. 6(d)–(f) and in Table 1. In order to have a quantitative assessment of the proposed method, we performed a comparison between the original high resolution image (Fig. 6a) and the image with enhanced resolution. The indicator that we have evaluated are the following: The correspondence of the retrieved spectra with the spectra used to build the image, the percentage of pixels which are not correctly retrieved after the spectral unmixing step, the percentage of pixels which are not correctly located after the Simulated Annealing step, and the percentage of mixed pixels (that are, pixels considered as pure in the high resolution image and mixed in the low resolution one) which are not correctly located after applying SA. More than 98% of the pixels are correctly located, which correspond to the 92.8% of the mixed pixels. All the pure pixels are correctly labeled.

From this simple experiment, the effectiveness of the proposed method can be evaluated. We can preliminary conclude that when the classes are correctly retrieved with spectral unmixing, the proposed spatial regularization technique provides very good results.

### 4.1. Spectral variability

In order to search for more challenging scenarios, the image was down-scaled of a factor 4, and the influence of noise has been considered. The filtered image is therefore composed of 1296 pixels, 818 pure and 478 mixed. Zero mean Gaussian noise is added to the original signal, in order to obtain the desired SNR. The use of Gaussian noise can be considered as a realistic

---

1. Villa et al. / Pattern Recognition 46 (2013) 1556–1568 1561
assumption [1,26,6] to simulate the spectral variability of the endmembers. Three different values of SNR have been tested: 50, 30 and 25 dB. When considering real hyperspectral images, the amount of noise is in general much lower than the one tested here; however, the Gaussian noise is useful to try to represent the spectral variability of real data, which is not considered in synthetic images. The presence of noise could affect the source separation step. If the endmembers are not correctly retrieved, the optimization step will start with wrong assumptions and inevitably lead to a bad result. As it can be seen from Table 1, the overall error increases with the noise, but still very good results are obtained. In the worst case, which is synthetic data with zoom factor of 4 and SNR of 25 dB, more than 91% of the image pixels are correctly labeled and positioned. Thus, the results obtained in the previous experiments are confirmed, stating the validity of the proposed spatial regularization and of the whole method proposed.

### Table 1

<table>
<thead>
<tr>
<th>Filtering window size</th>
<th>SNR</th>
<th>Inf</th>
<th>50 dB</th>
<th>30 dB</th>
<th>25 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean spectra correlation</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Pixel unc. labeled after SS</td>
<td>0.15%</td>
<td>0.83%</td>
<td>1.80%</td>
<td>2.51%</td>
<td></td>
</tr>
<tr>
<td>Pixel unc. positioned after SA</td>
<td>1.89%</td>
<td>6.40%</td>
<td>7.31%</td>
<td>8.82%</td>
<td></td>
</tr>
<tr>
<td>Pixel unc. positioned after PS</td>
<td>1.80%</td>
<td>5.62%</td>
<td>5.81%</td>
<td>6.10%</td>
<td></td>
</tr>
</tbody>
</table>

5. Experiments on real data

The experiments on real images were conducted by considering two different hyperspectral data. The first considered data set ROSIS data acquired over the University of Pavia, Italy, with 103 bands, ranging from 0.43 to 0.86 μm, with a 1.3 m spatial resolution. The very high value of the spatial resolution, which is not common in traditional hyperspectral satellite sensors, is due to fact that ROSIS is an airborne sensor. Here, we consider a
small segment (120 × 90 pixels) of the image, which contains several land cover classes. Fig. 7(a) shows a gray scale image of the 30th band of the scene. The main element of interest is the metal sheet structure in the center of the image.

The second image analyzed in our experiments is an AISA Eagle dataset. It contains 252 bands ranging from 395 to 975 nm in the visible and NIR spectral range. The original spatial resolution of the image was 2 m measured on ground, but in order to be treatable and still useful for the purposes of land cover interpretation it was downscaled to 6 m ground resolution while keeping the original spectral information as possible. The area is located in Hungary and contains arable lands near to the city of Heves. The area is mainly useful because of agricultural production. We considered a large subset of the image (400 × 500 pixels) containing six classes of interest.

The performances of the tested methods were evaluated in terms of overall accuracy (OA), that is the number of correctly classified test samples with respect to the total number of test samples,
Average Accuracy (AA), which represents the average of the classification accuracies for the individual classes, and the single classes accuracy. In the case of the ROSIS University experiment, where the classification of a single class is evaluated, the two accuracy indicators are equivalent.

5.1. ROSIS data set

The experiments carried out on the ROSIS data set are intended to evaluate the usefulness of the proposed method as a tool for structure detection. Two different tests were performed. The first one was on the original data, where all the pixels are considered as pure, in order to see the behavior of the proposed algorithms in such a situation. In the second experiment, the spatial resolution of the image was artificially degraded of a factor 3, so that the obtained images have a spatial resolution of 3.9 m, which is a realistic assumption in the case of airborne/satellite hyperspectral sensors. The method considered for downsampling is the same as in the previous experiments. We introduce in this way a number of mixed pixels, useful to evaluate the performance of our method in such a situation. In order to have a comparison with a traditional unsupervised classification method, we have also classified both images with a K-means classifier. The number of classes to select was set to 5, after applying the Virtual Dimensionality method (setting the probability of false alarm to 0.001).

Besides the number of classes, the only parameter which needs to be set in the proposed method is the threshold to determine whether a pixel can be considered as ‘pure’ after the first step. Instead of choosing an absolute value, we considered the difference between the two biggest abundances within a pixel, and set this value to 0.4. The decision to consider a relative value as threshold was taken by considering the characteristics of hyperspectral data, which are in general subject to high spectral variability. When performing spectral unmixing, endmembers which do not belong to a pixel could result in a small, but larger than zero abundance, mainly because of spectral variability or noise influence. With the proposed method, if a pixel contains two classes with abundances 0.65 and 0.35, it will be considered as mixed. However, if several classes are included in the pixel, the largest abundance being 0.65 and the others smaller than 0.2, the pixel will be considered as pure, since we assume that low abundances are related to spectral variability and noise.

The performance of the methods was evaluated on the classification of the metal sheet structure present in the middle of the image. In order to have a quantitative comparison of the results, in the case of low resolution data, the classification map obtained with the traditional unsupervised classifier was evaluated by comparison with the low resolution ground truth available, while the proposed methods is evaluated by comparison with the high resolution ground truth data. However, we would like to highlight that all the methods take as input the low resolution data. A simple Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>ROSIS original Before PP</th>
<th>ROSIS original After PP</th>
<th>ROSIS low resolution Before PP</th>
<th>ROSIS low resolution After PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM</td>
<td>50.86%</td>
<td>50.71%</td>
<td>93.75%</td>
<td>96.46%</td>
</tr>
<tr>
<td>KM–SU–SA</td>
<td>95.89%</td>
<td>96.91%</td>
<td>97.10%</td>
<td>98.35%</td>
</tr>
<tr>
<td>KM–SU–PS</td>
<td>95.24%</td>
<td>96.29%</td>
<td>96.91%</td>
<td>98.02%</td>
</tr>
<tr>
<td>VCA–SU–SA</td>
<td>96.95%</td>
<td>99.91%</td>
<td>97.12%</td>
<td>98.78%</td>
</tr>
<tr>
<td>VCA–SU–PS</td>
<td>96.55%</td>
<td>99.76%</td>
<td>96.85%</td>
<td>98.43%</td>
</tr>
</tbody>
</table>

Fig. 8. (a) Original ground truth. (b) K-means classification map, before post-processing. (c) Proposed method K-means + spectral unmixing classification map, before post-processing. (d) K-means classification map, after post-processing. (e) Proposed method K-means + spectral unmixing classification map, after post-processing. (f) Proposed method VCA + spectral unmixing classification map, after post-processing.
post-processing was applied to the classification map, in order to eliminate sparse pixels. For each pixel, a $3 \times 3$ window including its surrounding was used, and the value set to the most represented class within the window.

Both from Table 2 and Fig. 7 can be noticed the improvement provided by the proposed methods. Quite surprisingly, the $K$-means classifier provides better results in the case of low resolution data (also if the spatial accuracy of the method is clearly lower). The reason for this improvement is mainly due to two facts: (1) pixels labeled as "structure" in the low resolution data are composed by the average value of $9$ pixels of the original image, this mitigating the problem of spectral variability and (2) the number of pixels labeled as "structure" is much less than in the original case, since all the samples which were averaged with pixels belonging to other classes or unknown, were considered as mixed and therefore discarded from the ground truth.

It is highly remarkable that the proposed method obtains comparable results in the two cases, retrieving the metal structure as it is represented in the high resolution reference data. The qualitative improvement can be easily seen in Fig. 7.

5.2. AISA data set

The second experiment was carried out on the AISA data set, after reducing the spatial resolution of a factor 5. In spite of the high spatial resolution degradation, most of the pixels of the data set are to be considered as pure, since the image is mainly composed by large agricultural fields (high resolution ground truth represented in Fig. 8(a)).

After the unsupervised classification, every cluster was assigned to the label of the class that was better represented, taking care that each cluster was assigned to only one class. The overall classification accuracy was then computed along with the accuracies of the single classes and the average class accuracies. As in the previous experiment, the $K$-means output map was compared with the low resolution ground truth obtained after filtering (not considering pixels which become mixed), while the proposed method was compared with the high resolution ground truth. We want to stress that this type of comparison is highly unfavorable to our method, which is expected to correctly classify pixels which are mixed in the input image, and to correctly locate the obtained sub-pixels in order to have the same spatial distribution of the original image, while for the $K$-means classifier these pixels are not considered in the ground truth. The post-processing considered in the first experiment was applied also in this case to eliminate isolated pixels in the classification maps.

The quantitative results are shown in Table 3. Both proposed methods show better performances in terms of overall classification accuracy. The use of spectral unmixing with global end-members results in a high percentage of mixed pixel, as it can be noticed from the improvement obtained with the classification post-processing. Instead, the unsupervised classifier with local endmembers shows slightly better results in terms of average class accuracy after the post-processing step. The results of the experiment suggest that once the additional information about sub-pixel class abundances is retrieved by mean of spectral unmixing, the classification errors due to spectral variability can be easily corrected with a simple majority voting post-processing.

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Acc. (%)</th>
<th>Average Acc. (%)</th>
<th>Class 1 (%)</th>
<th>Class 2 (%)</th>
<th>Class 3 (%)</th>
<th>Class 4 (%)</th>
<th>Class 5 (%)</th>
<th>Class 6 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AISA data set—before PP</td>
<td>51.61%</td>
<td>61.37</td>
<td>93.75</td>
<td>56.41</td>
<td>95.83</td>
<td>59.15</td>
<td>6.67</td>
<td>56.44</td>
</tr>
<tr>
<td>KM–SU–SA</td>
<td>75.72</td>
<td>64.20</td>
<td>59.72</td>
<td>87.34</td>
<td>99.19</td>
<td>46.90</td>
<td>7.08</td>
<td>85.03</td>
</tr>
<tr>
<td>KM–SU–PS</td>
<td>75.65</td>
<td>64.15</td>
<td>59.21</td>
<td>87.32</td>
<td>99.11</td>
<td>46.91</td>
<td>7.00</td>
<td>84.97</td>
</tr>
<tr>
<td>VCA–SU–SA</td>
<td>59.69</td>
<td>56.83</td>
<td>58.26</td>
<td>67.87</td>
<td>36.69</td>
<td>73.50</td>
<td>51.19</td>
<td>58.96</td>
</tr>
<tr>
<td>VCA–SU–PS</td>
<td>59.54</td>
<td>56.41</td>
<td>58.11</td>
<td>67.88</td>
<td>30.60</td>
<td>73.21</td>
<td>51.13</td>
<td>58.87</td>
</tr>
<tr>
<td>AISA data set—after PP</td>
<td>52.75</td>
<td>65.60</td>
<td>100</td>
<td>55.75</td>
<td>100</td>
<td>71.83</td>
<td>6.67</td>
<td>59.09</td>
</tr>
<tr>
<td>KM–SU–SA</td>
<td>76.24</td>
<td>64.37</td>
<td>60.21</td>
<td>88.07</td>
<td>99.38</td>
<td>45.84</td>
<td>7.16</td>
<td>85.58</td>
</tr>
<tr>
<td>KM–SU–PS</td>
<td>76.11</td>
<td>64.21</td>
<td>60.22</td>
<td>88.01</td>
<td>99.38</td>
<td>45.64</td>
<td>7.18</td>
<td>85.50</td>
</tr>
<tr>
<td>VCA–SU–SA</td>
<td>70.57</td>
<td>65.73</td>
<td>75.63</td>
<td>77.41</td>
<td>28.18</td>
<td>94.75</td>
<td>70.05</td>
<td>59.77</td>
</tr>
<tr>
<td>VCA–SU–PS</td>
<td>70.51</td>
<td>65.70</td>
<td>75.61</td>
<td>77.40</td>
<td>28.12</td>
<td>94.75</td>
<td>70.06</td>
<td>59.60</td>
</tr>
</tbody>
</table>

Fig. 9. Variation of the overall classification accuracy versus the threshold parameter to determine the purity of a pixel (with PP): (a) AISA data set. (b) ROSIS low resolution data set. Blue bar: Pixel Swapping. Red bar: Simulated Annealing. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)
5.3. Discussion

The results obtained in the previous experiments prove that the proposed method has a very high potentiality for the unsupervised classification of hyperspectral images with low spatial resolution. In the case of highly mixed scene, as in the first experiment, endmember extraction based methods provide better results. On the contrary, when the number of pure pixels is much higher than the number of mixed ones (see for example the AISA data set experiment), the characteristics of clustering based methods have proven to be more suitable. The two approaches tested for sub-pixel locations have shown to be equivalent in terms of classification accuracy, with slightly better results provided by Simulated Annealing. In the proposed method, the only parameter having an influence on the overall classification accuracy obtained is the threshold to determine if a pixel can be considered as ‘pure’. How the classification accuracy changes by changing the value of the parameter can be seen in Fig. 9. It can be noticed that the proposed method shows similar classification accuracies for the three tested values, demonstrating that the choice of the parameters not crucial for the classification. As could be expected, a high value of the comparative threshold to determine if a pixel can be considered as ‘pure’ provides slightly higher accuracies, since only the most reliable pixels are labeled for the preliminary classification. By setting a low value of the threshold parameter, the preliminary classification map will tend to be like a common hard classification map obtained with a traditional classifier.

Regarding the computational burden, the two techniques considered to determine land cover classes, are equivalent in terms of processing time. The main difference between the methods that were tested in our experiments resides in the super-resolution algorithms used to locate sub-pixels. Fig. 10 shows the value of the perimeter of connected areas (which is the cost function considered in our experiments for Simulated Annealing) versus the number of iterations. It can be noticed that the Pixel Swapping method reaches the stability condition after few iterations (usually in the order of tens iterations), while Simulated Annealing, due to the random permutations performed, needs a much higher number of iterations to reach the same results. The computational burden required by Simulated Annealing is much higher in the case of AISA data set, due to the larger number of pixels and the higher resolution enhancement factor.

6. Conclusions

Unsupervised classification of hyperspectral images in the presence of mixed pixels was addressed. Two methods for structure detection and improvement of the spatial resolution of classification maps were proposed. The method exploits the advantages of source separation, unsupervised classifiers and spectral unmixing algorithms, in order to determine the fractional abundances of the classes at a sub-pixel scale. A spatial regularization by Simulated Annealing is finally performed to spatially locate the land cover classes within each pixel. Experiments were carried out on a synthetic and two real data sets. The experimental results show that the proposed method clearly outperforms classical unsupervised classification techniques when areas with mixtures of materials are located in the scene, providing excellent results both from a visually and quantitative point of view. Further research will be devoted to the investigation of advanced methods to better discriminate pure and mixed pixels, and of the possibility of alternative techniques of spatial regularization.

Acknowledgments

This work was supported in part by the European Community’s Marie Curie Research Training Networks Programme under Grant MRTN-CT-2006-035927. The authors would like to thank P. Gamba (University of Pavia) for sharing the Pavia ROSIS data set.

References

Alberto Villa received the B.S. and M.S. degrees in electronic engineering from the Universityof Pavia, Pavia, Italy, in 2005 and 2008, respectively. In 2011, he obtained the Ph.D. degree (a jointdegree) from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, and the University of Iceland, Reykjavik, Iceland. He was a Visiting Researcher at the HYPHERCOMP Lab, University of Extremadura, from September 2010 to February 2011. Since July 2011, he has been working as a Research Engineer for Aresys srl, a spin-off company of Politecnico di Milano dealing with SAR imaging and ground based radar. His research interests are in the areas of SAR antenna model, spectral unmixing, machine learning, hyperspectral imaging, signal and image processing.

Jocelyn Chanussot received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995 and the Ph.D. degree from the University of Savoie, Annecy, France, in 1998. In 1999, he was with the Geography Imagery Perception Laboratory, Délegation Générale pour l’Armement (French National Defense Department). Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005, was an Associate Professor from 2005 to 2007, and is currently a Professor of signal and image processing. He is currently conducting his research with the Grenoble Image Speech Signals and Automatics Laboratory, (GIPSA-Lab, affiliated with CNRS and Grenoble INP). His research interests include image analysis, multicomponent image processing, non-linear filtering and data fusion in remote sensing. He was an Associate Editor for Pattern Recognition (2006–2008).

Dr. Chanussot was an Associate Editor for the IEEE Geoscience and Remote Sensing Letters (2005–2007). He has been an Associate Editor for the IEEE Transactions on Image Processing (2009). Since 2009, he has been an Associate Editor for the IEEE Transactions on Geoscience & Remote Sensing 46 (May (5)) (2008) 1359–1374.

Jón Atli Benediktsson received the Cand. Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively. He is currently the Pro Rector for Academic Affairs and a Professor of Electrical and Computer Engineering with the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing and signal processing, and he has published extensively in those fields.

Dr. Benediktsson is the 2011 President of the IEEE Geoscience and Remote Sensing Society (GRSS) but he has been on the GRSS AdCom since 2000. He was Editor of the IEEE Transactions on Geoscience and Remote Sensing from 2003 to 2008. He was the recipient of the Stevan J. Krist of Award from Purdue University in 1991 as
Christian Jutten received the Ph.D. and Docteur des Sciences degrees from the Grenoble Institute of Technology, Grenoble, France, in 1981 and 1987. After being an Associate Professor with the Electrical Engineering Department, Grenoble Institute of Technology (1982–1989), and a Visiting Professor at the Swiss Federal Polytechnic Institute, Lausanne, Switzerland, in 1989, he became a Full Professor with the University Joseph Fourier of Grenoble, more precisely in the sciences and technologies department. For 30 yr, his research interests are learning in neural networks, blind source separation, and independent component analysis, including theoretical aspects (separability and source separation in non-linear mixtures) and applications (biomedical, seismic, speech, astrophysics, and chemical sensor array). He is the author or coauthor of four books, 18 invited papers, and more than 60 papers in international journals and 160 communications in international conferences. He has been a Scientific Advisor for signal and image processing at the French Ministry of Research (1996–1998) and for the French National Research Center (CNRS, 2003–2006). Dr. Jutten is a member of the technical committee “Blind Signal Processing” of the IEEE Circuits and Systems Society and of the technical committee “Machine Learning for Signal Processing” of the IEEE Signal Processing Society. He has been the Associate Editor of the IEEE Transactions on Circuits and Systems (1994–1995). He was Co-organizer of the 1st International Conference on Blind Signal Separation and Independent Component Analysis (Aussois, France, January 1999). For his contributions in source separation and independent component analysis, he received the EURASIP Best Paper Award in 1992, the Medal Blondel in 1997 from SEE (French Electrical Engineering Society) and, in 2008, has been elevated as a senior member of the Institut Universitaire de France.

Romain Dambreville is actually working on his Ph.D. at the Grenoble Institute of Technology.