Spectral Unmixing-Based Clustering of High-Spatial Resolution Hyperspectral Imagery

Eleftheria A. Mylona, Olga A. Sykioti, Konstantinos D. Koutroumbas, and Athanasios A. Rontogiannis, *Member, IEEE*

***Abstract*—This paper introduces a novel unsupervised spectral unmixing-based clustering method for high-spatial resolution hyperspectral images (*HSIs*). In contrast to most clustering methods reported so far, which are applied on the spectral signature representations of the image pixels, the idea in the proposed method is to apply clustering on the abundance representations of the pixels. Specifically, the proposed method comprises two main processing stages namely: an unmixing stage (consisting of the endmember extraction and abundance estimation (AE) substages) and a clustering stage. In the former stage, suitable endmembers are selected first as the most representative pure pixels. Then, the spectral signature of each pixel is expressed as a linear combination of the endmembers' spectral signatures and the pixel itself is represented by the relative abundance vector, which is estimated via an efficient AE algorithm. The resulting abundance vectors associated with the *HSI* pixels are next fed to the clustering stage. Eventually, the pixels are grouped into clusters, in terms of their associated abundance vectors and not their spectral signatures. Experiments are performed on a synthetic *HSI* dataset as well as on three airborne *HSI* datasets of high-spatial resolution containing vegetation and urban areas. The experimental results corroborate the effectiveness of the proposed method and demonstrate that it outperforms state-of-the-art clustering techniques in terms of overall accuracy, average accuracy, and kappa coefficient.**

***Index Terms*—Abundance estimation (AE), clustering, endmember extraction (EE), hyperspectral imagery (HSI), spectral unmixing (SU).**

I. Introduction

HYPERSPECTRAL imaging has enabled applications and detailed mapping possibilities in a wide variety of Earth studies. In particular, airborne hyperspectral images (*HSIs*) offer high-spatial resolution with detailed spectral accuracy. This versatility enhances the identification, modeling, and detailed classification of various natural and man-made materials. *HSIs* are collected via hyperspectral sensors and are represented as data cubes consisting of numerous contiguous spectral bands of narrow bandwidths. A significant characteristic of *HSIs*, which makes their processing more challenging, is the presence of *mixed pixels*, which depict surface regions consisting of two or more distinct materials. The data for each mixed pixel correspond to the total reflectance of all the materials present within the pixel in numerous spectral bands from the surface depicted by the pixel, which form the *spectral signature* of the pixel. The key objectives in *HSI* processing are: 1) the detection of the constituent components of mixed *HSI* pixels as well as the proportions in which they appear, which will allow the production of abundance maps per material and 2) the identification of spectrally homogeneous regions. The first objective is tackled via spectral unmixing (SU) and the second via the use of clustering algorithms.

In this study, we focus on the problem of identifying spectrally homogeneous regions, via clustering (unsupervised) techniques, which, in contrast to their supervised counterparts, they do not require any externally labeled set of pixels. Most clustering techniques proposed in this field are applied on the spectral signature representations of the pixels. In contrast, the key idea of the proposed methodology is to apply clustering on the abundance vector representations of the *HSI* pixels, since the latter representation is likely to lead to more well-separated clusters. To this end, SU is applied first on the spectral representations of the pixels, in order to extract the corresponding abundance vectors, and then, clustering is applied on the abundance vector pixels representations.

SU [1]–[6] of *HSIs* has been widely applied to environmental studies. It consists of two main substages, namely 1) endmember extraction and 2) abundance estimation (AE). EE [7]–[11] is a challenging process since the aim is to mine the purest pixels (*endmembers*) of each spectrally distinct material of a *HSI*. The latter almost always consists of mixed pixels, which are also affected by noise spectra. Ideally, each endmember ought to have the maximum possible abundance of a single physical material present in the *HSI* under study and minimum (close to zero) abundance for the rest of the physical materials. Moreover, the determination of the number of endmembers is critical since an underestimated number may result in poor representation of the mixed *HSI* pixels under study, whereas an overestimated number may comprise a lot of mixed signatures. Popular endmember extraction algorithms (*EEAs*) include *VCA* [12], *N-FINDR* variants [13], and *MVSA* [14]. Other related algorithms are discussed in [16]–[18].

The aim of *AE* is the decomposition of the spectral signatures of mixed pixels into a selection of spectral signatures

Manuscript received September 14, 2016; revised November 21, 2016 and January 24, 2017; accepted March 13, 2017. Manuscript received September 14, 2016. This work was supported by the PHySIS Project under Contract 640174 within the H2020 Framework Program of the European Commission. *(Corresponding author: Eleftheria A. Mylona.)*

The authors are with the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens (NOA), Athens 15236, Greece (e-mail: emylona@noa.gr; sykioti@noa.gr; koutroum@noa.gr; tronto@noa.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2017.2687703

corresponding to the reflectance of pure physical materials (*endmembers*). The latter is usually extracted by the image itself via *EE* (however, in some cases they are selected from specific spectral libraries). *AE* results in a set of corresponding fractions (*abundances*), which indicate the proportion of each endmember present in a given pixel. Clearly, the ultimate success of *AE* depends heavily on the appropriate selection of endmembers. Since only a small number of the available materials' spectra are expected to be present in a *HSI* pixel (especially in high- resolution *HSIs*), the abundance vectors are expected to be *sparse*.

Clustering [19], [20] partitions a set of pixels from the input image into groups. Some of the most known clustering approaches are the *k-means* [21], the Fuzzy *C-Means* (*FCM*) [22], the Possibilistic *C-Means* (*PCM*) [23] and their variants, e.g., [24], [25]. The aforementioned algorithms are suitable for recovering compact clusters and they use specific vectors, (called *representatives*) to represent the clusters that underlie in the current dataset. In contrast to these algorithms, that provide a single data clustering, in Hierarchical Agglomerative Clustering (*HAC*) [26], [27], the data are organized into an effective hierarchy of nested clusterings. *HAC* requires a metric in order to calculate the dissimilarity between pairs of pixels and a linkage so as to measure the dissimilarity between clusters.

A. Related Work

It should be mentioned that the literature on clustering techniques applied on *HSIs* is limited. In [28], a graph data structure is generated to represent the tree crowns weighted with the Euclidean distance. A minimum spanning tree is generated using Kruskal's algorithm and edges above a length threshold are removed to generate independent clusters. In [29], an unsupervised hierarchical cluster analysis to phytoplankton pigment data is applied with the aim of discriminating different phytoplankton assemblages in open ocean environments. Several types of optical data vectors are used as input to *HAC* including objects consisting of reflectance values of hyperspectral data. Also, in [30], a new clustering algorithm, named Adaptive Possibilistic C-Means *(APCM)*, is applied on *HSIs*.

In [31], a clustering procedure is proposed, which consists of three processes: 1) EE, 2) unmixing and 3) hardening process via the winner-takes-all approach, in order to produce reconstructed pixels spectra. In [32], the proposed work utilizes the Gauss Mixture Vector Quantization algorithm to learn the mixture analysis and explores the cluster analysis with correlation distance. In [33], *SU* is combined with *k-means* cluster analysis for accurate geological mapping. The data are first classified into two categories: hydrothermal alteration areas and unaltered rocks. SU is applied to hydrothermal alteration areas and *k-means* clustering to unaltered rocks as two separate approaches. In [34], the proposed work generates classification maps based on *k-means* clustering and Gradient Flow. SU is conducted using the Max-D algorithm to automatically find endmembers. It should be highlighted that, in all previous methods, the unmixing and clustering processes are utilized as two separate steps, in the sense that their results are extracted independently from each other and are combined next.

In this paper, a novel unsupervised *SU-based clustering* method (*SUBC*) for *HSIs* is proposed. *SUBC* consists of two processing stages namely: 1) SU, which consists of an *EEA*, followed by a (sparse) AE algorithm and 2) a clustering algorithm. The first process identifies suitable endmembers based on the *VCA* algorithm [12]. Then, *AE* is applied on each image pixel, in order to provide its abundance representation, using the sparsity-promoting *BiICE* algorithm [35]. Finally, the recently proposed *APCM* clustering algorithm [30] uses the abundance representations of the pixels, in order to group them into clusters. It should be noted that the abundance pixel representations adopted in the proposed methodology ensures (in general) a common sparsity pattern for pixels in the same cluster. To the best of our knowledge, this is the first attempt of utilizing the abundance representation of pixels generated by *SU* as input to a clustering algorithm with the aim to enhance classification in *HSIs*.

The proposed *SUBC* method is evaluated on a synthetic *HSI* dataset as well as on three airborne *HSI* datasets of high-spatial resolution (the agricultural area of Salinas Valley, CA, USA, the land cover at Washington DC Mall, USA, and the urban area of the Pavia center, Italy) and its performance is compared in terms of overall accuracy (*OA*), average accuracy (*AA*) and kappa coefficient with that of state-of-the-art clustering techniques.

The paper is organized as follows. Section II introduces the proposed *SUBC* method. Section III demonstrates the results obtained by the proposed method as well as comparisons with state-of-the-art clustering algorithms. Conclusion and future research directions are summarized in Section IV.

II. Proposed *SUBC* Method

In this section, we first present the motivation and contribution of this study and then we describe in detail the proposed unmixing-based clustering algorithm.

A. Motivation and Contribution

In general, classification algorithms [36], [37] (both supervised and unsupervised) developed so far are applied directly on the *L*-dimensional spectral band vectors of the pixels. However, such (usually high dimensional) representations may contain a lot of redundant information, which may cause pixels depicting different areas to be not well separated from each other in the *L*-dimensional spectral domain. Clearly, this renders the work of the classification algorithms more difficult. Apart from the above issue, most classification schemes used for *HSI* processing do not focus on exploiting the available fine spectral resolution, that is, they do not consider at all information *within* the pixel. A further consequence of this is that such schemes do not exploit the fact that each *HSI* pixel contains only a few of the materials existing in the whole *HSI* (equivalently, the spectral signature of each pixel is expected to result from the linear combination of only a few endmember spectral signatures, which implies that the corresponding abundance vectors will be *sparse*).

The approach that we adopt in this paper in order to leverage the above issues is to employ sparsity-promoting SU techniques in order to represent each pixel by its abundance vector (with

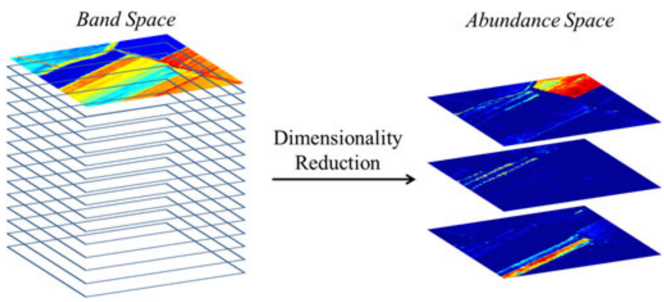


Fig. 1. Conceptual illustration of the dimensionality reduction achieved, moving from the original band space (usually consisting of hundreds of spectral bands) to the "less correlated" low-dimensional abundance space.

respect to a set of endmembers) and not by its spectral signature. The rationale behind this choice is twofold. First, the dimension of the abundance vector space (which equals to the number of the endmembers depicted in the *HSI* under study) is usually much lower than the dimension of the spectral signature space (number of spectral bands) (see Fig. 1). Since the corresponding original feature space (the space where each band defines an axis) is high dimensional, the Hughes phenomenon [38] ("curse" of dimensionality) appears. In light of this, the original high-dimensional space of the *HSI* is transformed to the dimensionally reduced space of abundance vectors [39].

Second, assuming that the endmembers are pure pixels, the (sparse) abundance vectors are expected to form clusters, which are likely to lie in different subspaces in the abundance space. It is, thus, anticipated that different classes will form more easily distinguishable clusters in the abundance vectors space. Generally speaking, adoption of the abundance representation is expected to ease the work of the classification methods. However, we have to keep in mind that the abundance retrieval requires a very good estimation of the endmembers that have a physical meaning in order to work properly, which, in practice, is not straightforward.

In the SU stage of the *SUBC* an *EEA* is first employed, which identifies appropriate endmembers of the image. Next, a sparse *AE* algorithm is used that is based on the endmembers extracted by the *EEA*, in order to produce the abundance fractions for each pixel, which in turn form the abundance vector of the pixel. These vectors of all pixels are fed to the second stage of the *SUBC* method, where a clustering algorithm groups pixels based on their abundance representations.

An additional feature concerning the mapping to the abundance space that should be highlighted is that the number of clusters and the number of endmembers are (in general) different. A cluster formed according to the abundances usually corresponds to a region where a single (or a few) endmembers have high proportion, whereas all other endmembers have low proportions. However, it can also correspond to the mixture of several endmembers of varied proportions. The block diagram of *SUBC* is depicted in Fig. 2.

B. Spectral Unmixing

1) Endmember Extraction: Aiming at detecting suitable endmembers, we utilize the *VCA* algorithm [12], which takes as input the spectral signatures of the pixels, as can be seen in Fig. 2. Each pixel can be viewed as a vector in an L-dimensional Euclidean space, where each spectral band is assigned to one axis of the space. Based on the aforesaid data points, the *VCA* algorithm returns a prespecified number of endmembers via iteratively projecting data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until the number of endmembers is exhausted [12]. Then, *SUBC* continues in estimating the abundance fractions of each endmember via *AE*.

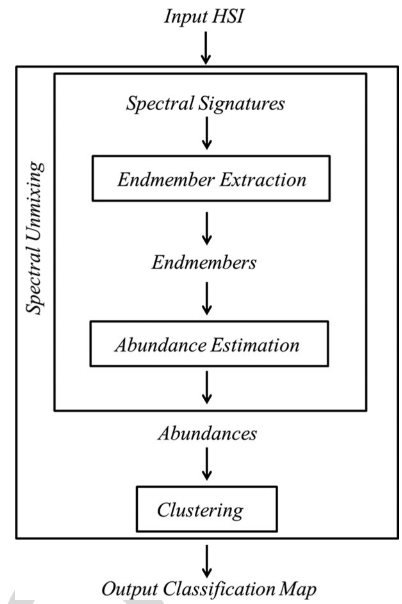


Fig. 2. Block diagram of *SUBC*.

2) Abundance Estimation: The selection of appropriate endmembers is crucial so as to correctly estimate the abundance fractions. Usually, the spectral signature of the pixel, denoted by $\boldsymbol{y}$, is assumed to follow the Linear Mixing Model [40] according to which it can be expressed as a linear combination of its endmembers' spectra as follows:

$$y = \Phi x + n \tag{1}$$

where $\Phi = [\varphi_1, \varphi_2, ..., \varphi_p] \in \Re_+^{L\times p},\ L \gg p$, is the mixing matrix comprising the endmembers' spectra (L-dimensional vectors $\phi_{i,}\ i = 1,\ 2, ..., p$), $\boldsymbol{x}$ is a $p \times 1$ vector consisting of the corresponding abundance fractions, named *abundance vector*, and $\boldsymbol{n}$ is an $L \times 1$ additive noise vector, which is assumed to be a zero-mean Gaussian distributed random vector with independent and identically distributed elements.

Due to the physical constraints of the unmixing problem, the abundance fractions for each pixel should satisfy the following two constraints:

$$x_i \geq 0,\ i = 1,\ 2,\ ...,\ N, \quad \sum_{i=1}^{N} x_i = 1 \tag{2}$$

that is, the abundances should be nonnegative and they must sum to 1. Furthermore, the *abundance* vector is expected to be sparse, i.e., only a few of its elements will be nonzero, since the

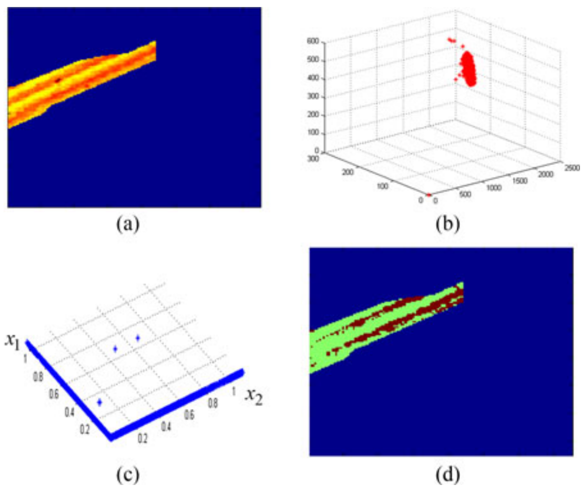


Fig. 3. (a) Class of the *HSI* containing two subclasses; (b) representation of pixels in the original space; (c) representation of abundance vectors $x = x_1, x_2$; and (d) clustering result emerged from *SUBC*.

area depicted by a single pixel is likely to embed only a small fraction of the different materials encountered in the whole *HSI*.

In this study, the abundance vector for each pixel is estimated via a variational Bayes algorithm called *BiICE* [35] (see Fig. 2) that imposes sparsity on the abundance vector and is based on an appropriately defined hierarchical Bayesian model. In algorithmic form, the abundance vector can be obtained as follows:

$$x = BiICE(\Phi, y). \tag{3}$$

BiICE is computationally efficient, provides sparse solutions without requiring the fine-tuning of any parameters, and converges fast to accurate values even for highly correlated data. The determined abundance vectors $\boldsymbol{x}$ are further used for the representation of their associated pixels at the *clustering* process.

In order to unravel the advantages of using the abundance representation of the pixels instead of the traditional band representation, we consider the following simplified case. For illustration purposes, we form an *RGB* image selecting three appropriate bands from a small area of one of the *HSIs* considered in Section III-B. The considered area [see Fig. 3(a)] is a class consisting of two subclasses. The representation of the pixels in the original space is depicted in red color in Fig. 3(b). Assuming two endmembers (one from each subclass), Fig. 3(c) depicts the abundance vectors stemmed from *BiICE* in blue color. Note that, due to the imposed sparsity, almost all pixels are concentrated around the two axes. Finally, Fig. 3(d) depicts the classification map produced by *SUBC*.

It should be highlighted that the pixels in the original space formulate one compact cloud with a few outliers and, thus, it is difficult to be naturally divided into two separate groups. On the contrary, the abundance vectors formulate two compact clouds tangent to the axes, which are highly distinguished. This is the advantage that characterizes the representation of the pixels using their abundance vectors and eases *SUBC* to correctly identify the two subclasses, via its second stage. It should also be reminded here that the abundance vectors are characterized by sparsity (i.e., the existence of zeros in vectors $\boldsymbol{x}$), which promotes data distinctions.

C. Clustering

The *clustering* stage, which is applied on the abundance representations of the *HSI* pixels under study, employs the *APCM* algorithm [30] (see Fig. 2). Let $X = \{x_i \in \Re^p, i = 1, \ldots, N\}$ be a set of N p-dimensional data vectors to be clustered and $\Theta = \{\theta_j \in \Re^p, j = 1, \ldots, m\}$ be a set of m vectors (called *representatives*) that will be used for the representation of the clusters formed by the points in X. Let $U = [u_{ij}], i = 1, \ldots, N, j = 1, \ldots, m$ be an $N \times m$ matrix whose (i, j) entry stands for the so-called *degree of compatibility* of $\boldsymbol{x}_i$ with the jth cluster denoted by C_j and represented by the vector θ_j. The *APCM* algorithm emerges from the optimization of the cost function of the original *PCM* described as follows:

$$J_{PCM}(\Theta, U) = \sum_{j=1}^{m} [\sum_{i=1}^{N} u_{ij} ||\mathbf{x}_i - \theta_j||^2 + \gamma_j \sum_{i=1}^{N} (u_{ij} \ln u_{ij} - u_{ij})]. \tag{4}$$

In contrast to the classical *PCM,* where $\gamma_j' s$ remain constant during the execution of the algorithm, in *APCM* $\gamma_j' s$ are adapted at each iteration through the adaptation of the corresponding $\eta_j' s$. This is achieved by setting $\gamma_j = \frac{\hat{\eta}}{\alpha} \eta_j$ and adapting η_j (which is a measure of the mean absolute deviation of the current form of cluster C_j) at each iteration of the algorithm. Note that $\eta_j' s$ and α are constant quantities (for more details see [30]).

The output of the algorithm is a classification map consisting of clusters formed based on the abundances produced in *SU*. The clusters that are formed usually correspond to regions where a few abundances have high values of fractions, whereas the remaining ones exhibit low values (that is, they are aggregated around certain subspaces in the abundance space).

III. Experimental Results and Discussion

SUBC has been experimentally evaluated in four case studies: a synthetic and three real airborne *HSI* datasets of high-spatial resolution. The synthetic *HSI* dataset has been generated with various values of additive noise in order to test the sensitivity of the proposed method under different noise levels. The first airborne *HSI* dataset represents a challenging area of various plant species on an agricultural area, where discrimination between the species is impeded by numerous factors such as the similar spectral signatures of the pixels as well as the absence of reference spectra. The second airborne *HSI* dataset represents a land cover of mixed vegetation and urban materials whose spectral signatures patterns vary. The third airborne *HSI* dataset represents a mainly urban area, where the spectral signatures of the materials present are not characterized by specific patterns.

A. Synthetic HSI Dataset

The experimental evaluation of *SUBC* has been conducted on a 100×100 synthetic *HSI* dataset consisting of five different

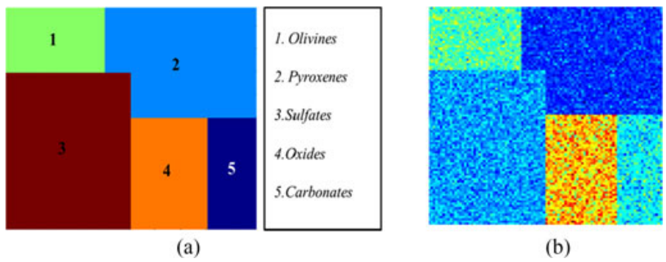


Fig. 4. (a) Reference map of synthetic *HSI* dataset and (b) 100th band added with noise at 20 dB.

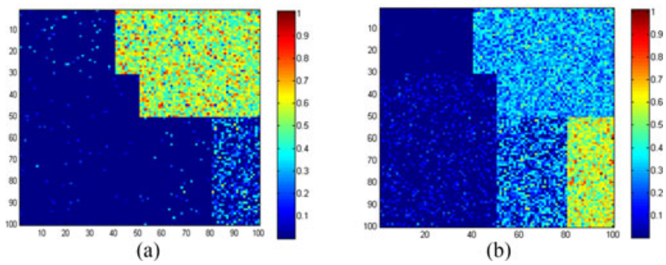


Fig. 5. Estimated abundance maps for two endmembers (a) pyroxenes and (b) carbonates extracted from synthetic *HSI* via *BiICE*. Abundance values range from 0 (blue) to 1 (red).

regions artificially generated. The spectral signatures have been obtained by the U.S. Geological Survey Spectral Library [41]. The data cube contains areas with mineral signatures of five general mineral classes: 1) olivines; 2) pyroxenes; 3) sulfates; 4) oxides; and 5) carbonates. The *HSI* under study comprises 109 spectral bands. For the generation of the synthetic hyperspectral data cube, seven endmembers have been randomly selected and for each mineral class seven pure pixels have been assigned. It should be highlighted that for each mineral class more than one endmembers have been randomly assigned. Each one of the five regions consists of a linear combination of different randomly selected different endmembers contaminated by additive Gaussian zero mean noise.

Fig. 4(a) depicts the reference map, while Fig. 4(b) shows the 100th band of the synthetic *HSI* contaminated by 20-dB additive noise. It should be noted that noise is added in all bands of the synthetic *HSI* dataset and experiments have been conducted with different *SNRs* in the range of 20–40 dB. Fig. 5 illustrates abundance maps obtained from *BiICE* for two endmembers 1) pyroxenes and 2) carbonates extracted from the synthetic *HSI* under study. In Fig. 6, *SUBC* is compared with state-of-the-art clustering algorithms namely *k-means*, complete-link *HAC*, *FCM*, and *APCM*. It should be highlighted that all these algorithms are applied on the spectral signatures of the pixels, whereas the clustering procedure in *SUBC* is applied on the abundance representations of the pixels (due to the philosophy of the method). As shown in Fig. 6, classes 1, 3, and 4 are correctly identified by all tested algorithms, while the superiority of the proposed *SUBC* algorithm is clearly demonstrated in the identification of classes 2 and 5.

Table I contains the results obtained by *k-means*, *HAC*, *FCM*, *APCM*, and *SUBC* in terms of *OA* and *kappa* coefficient based on the obtained confusion matrix for 20-dB *SNR* [33]. Table II demonstrates the results in terms of *AA* (fraction of true positives and true negatives) for each class. We observe that *SUBC* outperforms all existing clustering techniques and offers an almost 100% *OA* and *AA*. It should be noted here that similar results have also beenobtained for all other values of *SNR* tested in the range 20–40 dB.

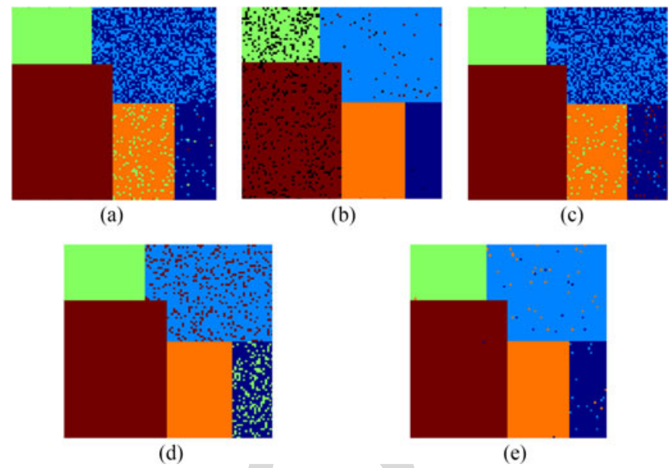


Fig. 6. Clustering results emerged from: (a) *k-means*; (b) *HAC*; (c) *FCM*; (d) *APCM*; and (e) *SUBC* on the synthetic HSI under study.

TABLE I
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SYNTHETIC *HSI* DATASET IN TERMS OF *OA* AND *Kappa Coefficient*

	OA(%)	kappa
k-means	86.95	0.76
HAC	93.51	0.87
FCM	90.01	0.89
APCM	97.73	0.90
SUBC	99.28	092

TABLE II
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SYNTHETIC *HSI* DATASET IN TERMS OF *AA* FOR EACH CLASS

Class	k-means	HAC	FCM	APCM	SUBC
1	99.76	92.51	99.70	99.72	99.78
2	77.52	93.21	77.20	97.31	98.66
3	99.24	82.24	99.30	99.25	99.26
4	75.03	99.69	85.87	99.53	99.50
5	83.20	99.90	87.98	92.84	99.20

B. Airborne HSI Datasets

SUBC has been also experimentally evaluated on the *HSI* airborne dataset of the Salinas Valley, CA, USA [42], which constitutes an arduous clustering scenario. Salinas *HSI* has been collected by the Airborne Visible Infra-Red Imaging Spectrometer (*AVIRIS*) sensor over an agricultural area of Salinas Valley, California. The *AVIRIS* sensor, developed by NASA's Jet Propulsion Laboratory [43], generates calibrated radiance

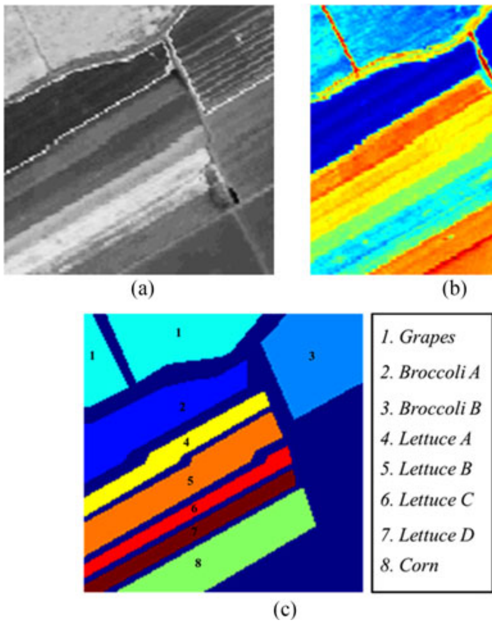


Fig. 7. (a) First PCA band; (b) 117th band of Salinas Valley; and (c) masked reference map [42].

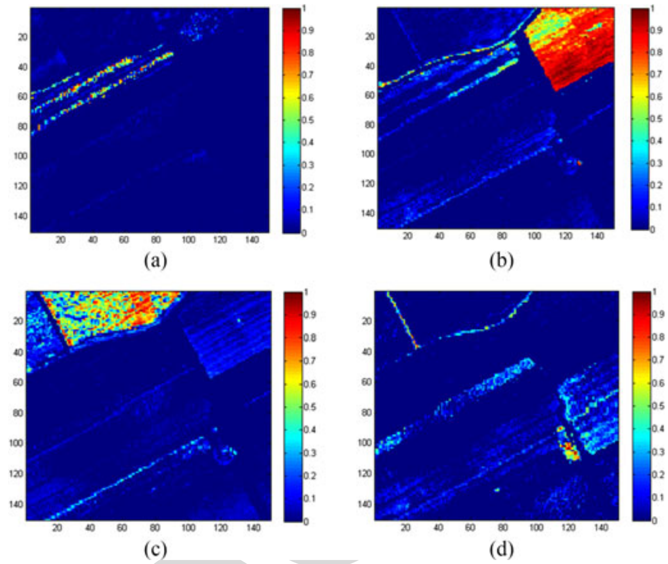


Fig. 8. Estimated abundance maps for four endmembers extracted from Salinas Valley *HSI* via *BiICE*. Abundance values range from 0 (blue) to 1 (red).

images in 224 contiguous spectral bands with wavelengths from 400 to 2500 nm. Moreover, it is characterized by high-spatial resolution of 3.7-m pixels. The number of bands is reduced to 204 by removing 20 water absorption bands. Salinas Valley *HSI* consists of vegetables and vineyard fields. Its masked reference classification map comprises eight classes: corn, two types of broccoli, four types of lettuce and grapes [42]. Fig. 7 depicts: (a) first *PCA* band; (b) 117th band; and (c) masked reference map of a 150×150 subimage of the Salinas Valley *HSI*.

Ideally, one would have a digital spectral library of reference spectra of the mapped plant species. However, such a publicly available library does not exist for the specific plant species. In addition, it is not known how many spectra would be required to represent the changing spectral signatures, as a function of the growing season. This unavoidably leads to the selection of the endmembers from the image itself. Doing so, Fig. 8(a)–(d) depict estimated abundance maps stemmed from *BiICE* for four endmembers extracted from Salinas Valley *HSI*. Fig. 8(a) and (b) correspond to two types of broccoli, Fig. 8(c) to one type of grapes and Fig. 8(d) to a (most probably) construction.

Aiming at a quantitative evaluation, *SUBC* is compared against *k-means*, *HAC*, *FCM*, and *APCM* in terms of *OA* and *AA* computed by the obtained confusion matrix as can be seen in Tables III and IV, respectively. We see from Tables III and IV that *SUBC* achieves *OA*, *kappa*, and *AA* values which are higher than that of the other state-of-the-art clustering algorithms. Fig. 9 illustrates clustering results emerged from: (a) *k-means;* (b) *HAC;* (c) *FCM*; (d) *APCM*; and (c) *SUBC* on the Salinas *HSI* dataset. It should be mentioned that the results obtained by *APCM* and *SUBC* demonstrate the correct identification of all classes and subclasses as can be seen by examining the first *PCA* band in Fig 7(a).

TABLE III
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SALINAS *HSI* DATASET IN TERMS OF *OA* AND *Kappa Coefficient*

	OA(%)	kappa
k-means	72.67	0.70
HAC	87.07	0.75
FCM	82.46	0.70
APCM	91.34	0.78
SUBC	93.04	0.80

TABLE IV
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SALINAS *HSI* DATASET IN TERMS OF *AA* FOR EACH CLASS

Class	k-means	HAC	FCM	APCM	SUBC
Grapes	73.67	94.27	74.92	87.92	94.77
Broccoli A	74.43	73.82	92.83	92.79	93.49
Broccoli B	73.56	73.93	90.12	90.82	91.52
Lettuce A	72.43	89.38	72.81	92.27	93.37
Lettuce B	73.21	91.59	70.62	91.39	92.36
Lettuce C	70.23	92.72	91.29	92.12	92.79
Lettuce D	71.54	93.91	92.46	93.24	93.52
Corn	72.29	86.94	74.63	90.17	92.50

SUBC has also been quantitatively evaluated on the *HSI* airborne dataset of the Pavia Center [42]. The image has been acquired by the reflective optics system imaging spectrometer sensor over an urban area of the city center. The flight was operated by the German Aerospace Agency under the HySens project managed by the German Aerospace Center (*DLR*). The original data consist of 115 spectral bands (with the spectral range from 0.43 to 0.86 μm) and has a high-spatial resolution of 1.3 m. However, noisy bands were previously removed leading to a total of 102 bands. Four thematic classes are present in the

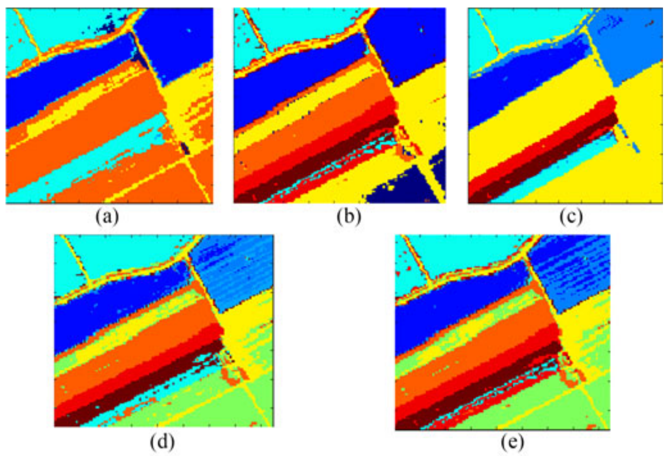


Fig. 9. Clustering results emerged from: (a) *k-means*; (b) *HAC*; (c) *FCM*; (d) *APCM*; and (e) *SUBC* on the Salinas *HSI*.

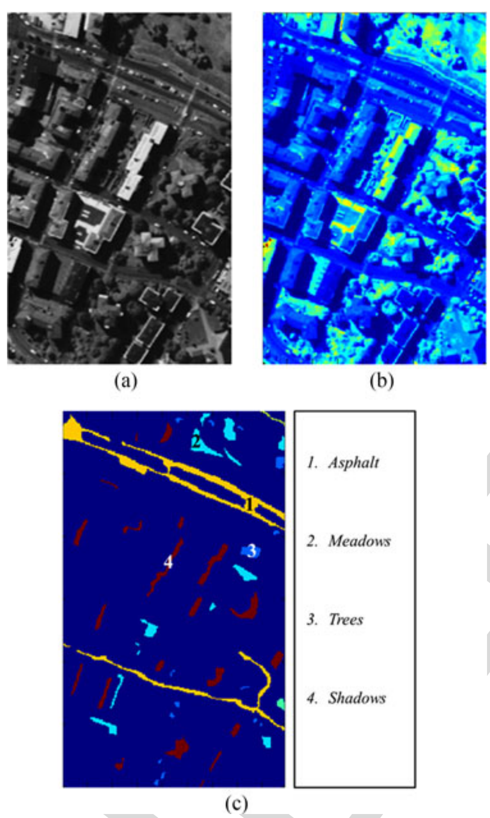


Fig. 10. (a) First PCA band; (b) 80th band of Pavia center; and (c) masked reference map [42] (1-yellow, 2-light blue, 3-dark blue, and 4-brown).

scene: 1) asphalt; 2) meadows; 3) trees; and 4) shadows, according to the reference classification map provided by [42]. Fig. 10 depicts: (a) first *PCA* band; (b) 80th band; and 3) masked reference map of a 300 × 177 subimage of the Pavia center *HSI* [42]. Fig. 11(a) and (b) depicts estimated abundance maps stemmed from *BiICE* for two endmembers: (a) shadow and (b) manmade material.

In the scope of a quantitative evaluation, *SUBC* is compared against *k-means*, *HAC*, *FCM*, and *APCM* in terms of the *OA* and *kappa* coefficient computed by the obtained confusion matrix as can be seen in Table V and in terms of the *AA* as can be seen in Table VI, while the clustering results of all algorithms are shown in Fig. 12. Again, *SUBC* provides the best clustering performance as witnessed by its *OA, kappa*, and *AA* values, which are the highest among all its competitors.

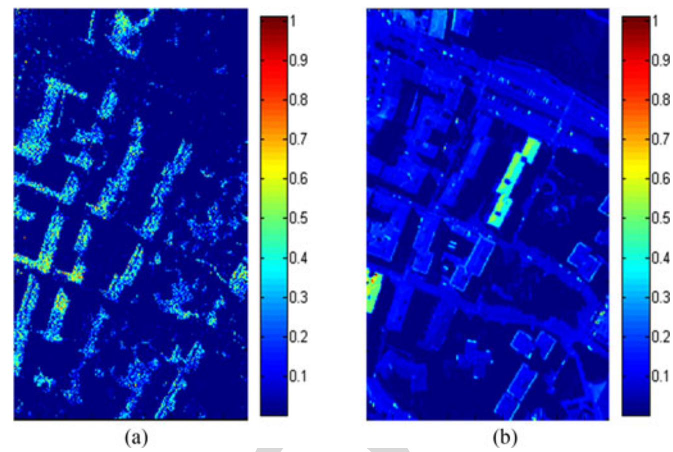


Fig. 11. Estimated abundance maps for two endmembers (a) shadow, (b) manmade material extracted from Pavia center *HSI* via *BiICE*. Abundance values range from 0 (blue) to 1 (red).

TABLE V
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON PAVIA *HSI* DATASET IN TERMS OF *OA* AND *Kappa Coefficient*

	OA(%)	Kappa
k-means	93.26	0.80
HAC	37.03	0.71
FCM	92.46	0.78
APCM	93.38	0.79
SUBC	96.30	0.83

TABLE VI
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON PAVIA *HSI* DATASET IN TERMS OF *AA* FOR EACH CLASS

Class	k-means	HAC	FCM	APCM	SUBC
Asphalt	94.28	25.93	94.90	95.01	97.31
Meadows	90.62	16.72	92.61	91.68	96.71
Trees	92.25	21.09	86.51	90.47	94.37
Shadows	95.89	84.39	95.82	96.36	96.81

Finally, *SUBC* has been qualitatively evaluated on the *HSI* airborne dataset of the Washington DC mall [44]. The image has been acquired by the airborne mounted Hyperspectral Digital Imagery Collection Experiment sensor. The sensor system used in this case measured pixel response in 210 bands in the 0.4–2.4 μm region of the visible and infrared spectrum. Bands in the 0.9–1.4 μm region, where the atmosphere is opaque, have been omitted from the dataset leaving 191 bands. Moreover, the dataset exhibits high-spatial resolution (2.8 m). Five thematic land cover *classes* are present in the scene: 1) roof; 2) grass; 3) trees; 4) water; and 5) asphalt road, according to the Q3

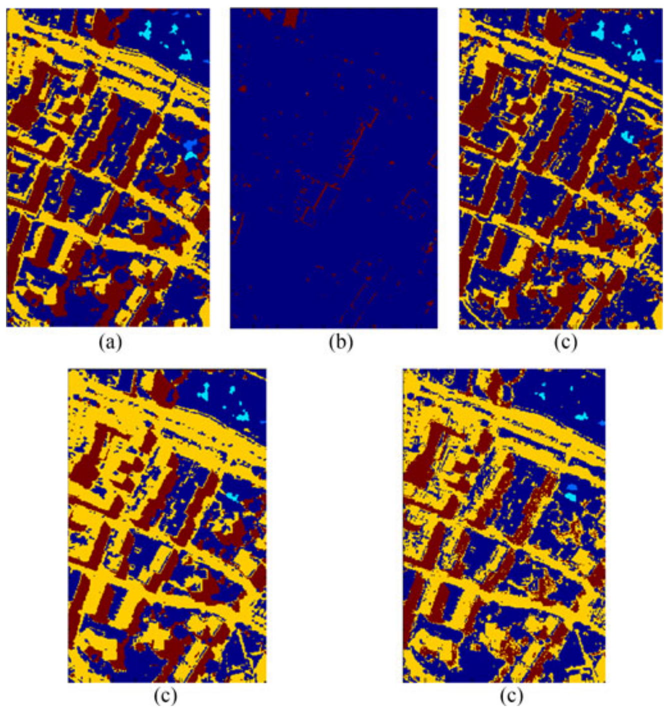


Fig. 12. Clustering results emerged from: (a) *k-means*, (b) *HAC*, (c) *FCM*, (d) *APCM* and (e) *SUBC* on the Pavia center *HSI*.

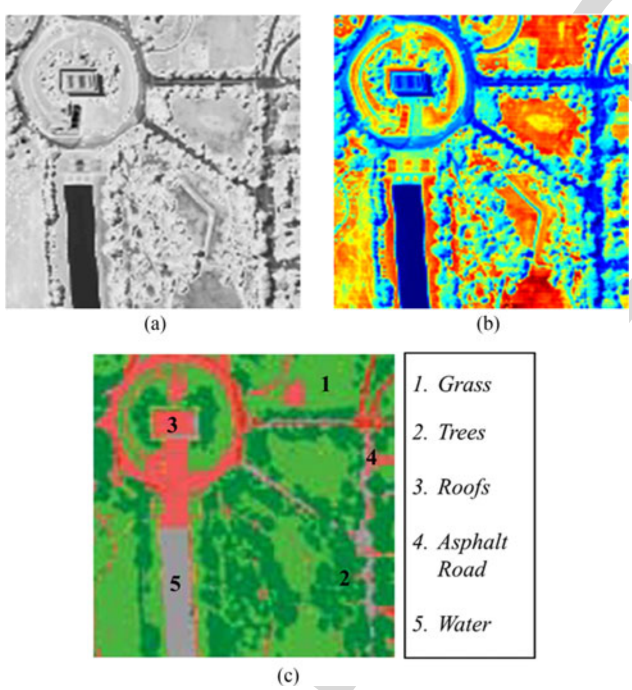


Fig. 13. (a) First PCA band; (b) 100th band of Washington DC; and (c) reference map [44].

classification map provided by [49] and used here as a reference map.

Fig. 13 depicts: (a) first *PCA* band; (b) 100th band; and (c) reference map of a 100×100 subimage of the Washington DC mall *HSI* [44]. It should be noticed that the reference map is provided only for qualitative visualization assessment and it is not accurate for a thorough quantitative assessment. Fig. 14(a)

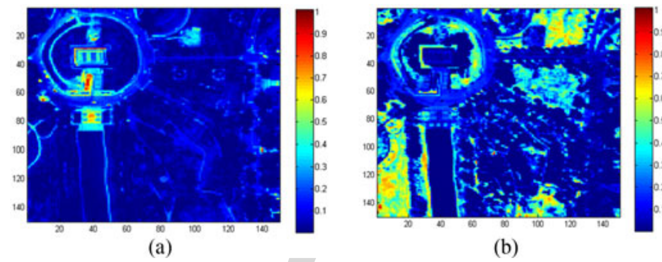


Fig. 14. Estimated abundance maps for two endmembers: (a) manmade material and(b) (most probably) soil/grass of class 1 extracted from Washington DC HSI via BiICE. Abundance values range from 0 (blue) to 1 (red).

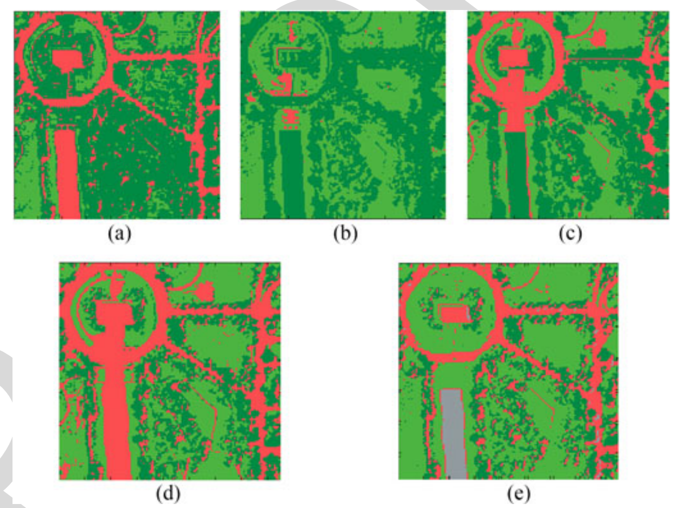


Fig. 15. Clustering results emerged from: (a) k-means; (b) HAC; (c) FCM; (d) APCM; and (e) SUBC on the Washington DC HSI.

and (b) depicts estimated abundance maps stemmed from *BiICE* for two endmembers: 1) manmade material and 2) (most probably) soil/grass of class 1. Fig. 15 illustrates clustering results emerged from: (a) *k-means*; (b) *HAC*; (c) *FCM*; (d) *APCM*; and (e) *SUBC* on the Washington DC *HSI* dataset. It should be highlighted that, apart from *SUBC*, all other algorithms falsely classify water and asphalt road pixels to one class. On the other hand, *SUBC* correctly distinguishes pixels that belong to the water class from all other pixels that belong to the remaining classes.

As it has been highlighted throughout the paper, the key idea of the proposed method is to perform unmixing at its first stage, in order to take the abundance representations of the pixels and then, at the second stage, to perform clustering based on the pixels abundance vector representations. Clearly, one could choose any unmixing method in the first stage and any clustering method in the second stage of the algorithm. In order to justify the choice of *BiICE* in the first stage, we compare it against two *AE* algorithms: 1) a quadratic programming (*QP*) technique [45], which does not exploit sparsity and 2) the sparse unmixing by variable splitting and augmented Lagrangian (*SUnSAL*) algorithm [46], which, as *BiICE*, imposes sparsity. That is, we substitute *BiICE* with *QP* and *SUnSAL* at the first stage of the proposed method. Leaned on Table VII, which depicts the *OA*

TABLE VII
COMPARATIVE RESULTS OF *SU* ALGORITHMS ON *HSI* DATASETS IN TERMS OF *OA*

	Synthetic	Salinas	Pavia
QP	83.67	74.32	71.20
SUnSAL	97.50	82.71	87.52
BiICE	99.28	93.04	96.30

of the three cases, the *QP* algorithm attains the worst performance (since it does not take into account that by the nature of the problem, the abundance vectors exhibit sparsity), whereas *SUnSAL* exhibits significantly improved performance compared to *QP*, yet inferior, compared to *BiICE,* especially for real data. Moreover, *SUnSAL* comes at the additional expense of manually fine-tuning nontrivial parameters, such as a sparsity promoting parameter λ.

The choice of *APCM* in the second stage of the algorithm is justified mainly by the fact that it is able to estimate automatically the underlying number of clusters in the dataset. Moreover, focusing on the first four lines of Tables I, III, and IV, the *OA* of *APCM* is significantly higher from all other state of the art clustering methods (note that all these algorithms are applied on the same dataset, i.e., the spectral signature representations of the *HSI* pixels).

IV. CONCLUSION AND FUTURE DIRECTIONS

The key challenge of the proposed method (*SUBC*) is the identification of spatially homogeneous regions comprising different materials. The method consists of two main stages (unmixing and clustering) and generates three significant (by)products, namely: 1) endmembers; 2) abundance vectors (abundance maps); and 3) clusters (classification maps). The key feature of *SUBC* is the utilization of the abundance representations of the *HSI* pixels (as they result from the unmixing stage) in the clustering stage. The advantage of using the abundance representation instead of the basic spectral representation of the pixels is that the former, in contrast to the latter, provides subpixel level information, which in turn favors more detailed classification maps. Moreover, the abundance representation is likely to give rise to more well-discriminated clusters that live on subspaces of the abundance space, due to the fact that only a few materials are expected to contribute to the formation of a *HSI* pixel (sparsity issue). As a consequence, subspace clustering algorithms could also be considered as an alternative in the final stage of the algorithm, since the abundance representations are likely to lead to clusters that live to subspaces of the abundance space. *SUBC* is unsupervised and does not require class information knowledge of the dataset under study. Moreover, it is image independent, it alleviates the "curse of dimensionality" issue and enhances localization and accuracy since it operates in the subpixel level of information. However, it is noted again that the correct identification of the endmembers number and their correspondence to physical objects/materials is undoubtedly the most critical step for successful SU and, as a consequence, for the clustering processes.

Experimental results show that *SUBC* compares favorably to other related methods. This gives us confidence to claim that the performance of the proposed method remains consistent with high-spatial resolution airborne data. It is capable of identifying compact regions and spectral regions that lack training data.

In terms of future directions, the full potential of this algorithm will be investigated with additional hyperspectral acquisitions of higher mixture complexity. In addition, this study could be reinforced and expanded in the case of existing and future satellite hypespectral data imagery of lower spatial resolutions where increased complexity issues for the tasks of 1) endmember identification; 2) resolving shadowing effects; and 3) facing oblique viewing and illumination angles arise. Moreover, subspace clustering algorithms could be utilized, since as we discussed earlier, they suit nicely in the nature of the problem in the abundance space.

ACKNOWLEDGMENT

The AVIRIS data were provided courtesy of NASA/JPL-Caltech in Pasadena, CA, USA, via the EXELIS Visual Information Solutions Inc., 2012.

REFERENCES

[1] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised multitemporal spectral unmixing for detecting multiple changes in Hyperspectral Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2733–2748, May 2016.

[2] C. Shi and L. Wang, "Incorporating spatial information in spectral unmixing: A review," *Remote Sens. Environ.*, vol. 149, pp. 70–87, 2014.

[3] T. Meyer and G. S. Okin, "Evaluation of spectral unmixing techniques using MODIS in a structurally complex savanna environment for retrieval of green vegetation, nonphotosynthetic vegetation and soil fractional cover," *Remote Sens. Environ.*, vol. 161, pp. 122–130, 2015.

[4] Y. Yuan, M. Fu, and X. Lu, "Substance dependence constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 2975–2986, Jun. 2015.

[5] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[6] Y. Yuan, Y. Feng, and X. Lu, "Projection-based subspace learning for hyperspectral unmixing," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2632–2643, Jun. 2015.

[7] B. Zhang, X. Sun, L. Gao, and L. Yang, "Endmember extraction of hyperspectral remote sensing images based on the discrete particle swarm optimization algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4173–4176, Nov. 2011.

[8] P. E. Dennison and D. A. Roberts, "Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE," *Remote Sens. Environ.*, vol. 87, pp. 123–135, 2003.

[9] C. Song, "Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?" *Remote Sens. Environ.*, vol. 95, pp. 248–263, 2005.

[10] R. L. Powell, D. A. Roberts, P. E. Dennison, and L. L. Hess, "Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil," *Remote Sens. Environ.*, vol. 106, pp. 253–267, 2007.

[11] T. C. Eckmann, D. A. Roberts, and C. J. Still, "Using multiple endmember spectral mixture analysis to retrieve subpixel fire properties from MODIS," *Remote Sens. Environ.*, vol. 112, pp. 3773–3783, 2008.

[12] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[13] T. H. Chan, W. K. Ma, A. Ambikapathi, and C. Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.

[14] J. Li *et al.*, "Minimum volume simplex analysis: a fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.

[15] A. Plaza and C.-I. Chang, "Impact of initialization on design of endmember extraction algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3397–3407, Nov. 2006.

[16] A. Zare, P. Gader, O. Bchir, and H. Frigui, "Piecewise convex multiple-model endmember detection and spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2853–2862, May 2013.

[17] A. Zare, P. Gader, and G. Casella, "Sampling piece-wise convex unmixing and endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1655–1665, Mar. 2013.

[18] B. Somers, M. Zortea, A. Plaza, and G. P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 396–408, Apr. 2012.

[19] M. R. Anderberg, *Cluster Analysis for Applications*. New York, NY, USA: Academic, 1973.

[20] B. S. Everitt, S. Landau, and M. Leese, *Cluster Analysis*. London, U.K.: Wiley, 2001.

[21] R. O. Duda and P. Hart, *Pattern Classification and Scene Analysis*. Hoboken, NJ, USA: Wiley, 1973.

[22] J. C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.

[23] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.

[24] S. D. Xenaki, K. D. Koutroumbas, and A. A. Rontogiannis, "Sparsity-aware possibilistic clustering algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 6, pp. 1611–1626, Dec. 2016.

[25] R. Inokuchi and S. Miyamoto, "Sparse possibilistic clustering with L1 regularization," in *Proc. IEEE Int. Conf. Granular Comput.*, Fremont, CA, USA, 2007, p. 442.

[26] S. Theodoridis and K. Koutroumbas, *Pattern Recognition Applications*, 4th ed. New York, NY, USA: Academic, 2008.

[27] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.

[28] P. Bunting, R. M. Lucas, K. Jones, and A. R. Bean, "Characterisation and mapping of forest communities by clustering individual tree crowns," *Remote Sens. Environ.*, vol. 114, pp. 2536–2547, 2010.

[29] E. Torrecilla, D. Stramski, R. A. Reynolds, E. Millán-Núnez, and J. Piera, "Cluster Analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean," *Remote Sens. Environ.*, vol. 115, pp. 2578–2593, 2011.

[30] S. D. Xenaki, K. D. Koutroumbas, and A. A. Rontogiannis, "A novel adaptive possibilistic clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 4, pp. 791–810, Aug. 2016.

[31] H. Gholizadeh, M. J. V. Zoej, and B. Mojaradi, "A novel hyperspectral image clustering method based on spectral unmixing," in *Proc. Aerosp. Conf.*, 2012, pp. 1–5.

[32] M. Parente and A. Zymnis, "Statistical clustering and mineral spectral unmixing in AVIRIS hyperspectral image of Cuprite, NV," CS2009 Report, 2005.

[33] N. Ishidoshiro *et al.*, "Geological mapping by combining spectral unmixing and cluster analysis for hyperspectral data," in *Proc. Int. Archives Photogrammetry, Remote Sens. Spatial Inform. Sci.*, 2016, pp. 431–435.

[34] J. D. Kwong, D. W. Messinger, and W. D. Middleton, "Hyperspectral
Q4 clustering and unmixing for studying the ecology of state formation and complex societies," in *Proc. SPIE Opt. Eng. Appl., Int. Soc. Opt. Photon.*, 2009.

[35] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585–599, Feb. 2012.

[36] J. Li, M. Khodadadzadeh, A. Plaza, X. Jia, and J. M. Bioucas-Dias, "A discontinuity preserving relaxation scheme for spectral-spatial hyperspectral image classification," *IEEE J. Select. Topics Appl. Earth Observ.Remote Sens.*, vol. 9, no. 2, pp. 625–639, Feb. 2016.

[37] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.

[38] W. Ma, C. Gong, Y. Hu, P. Meng, and F. Xu, "The Hughes phenomenon in hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai Lake," in *Proc. Int. Symp. Photoelectron. Detection Imag. 2013: Imag. Spectrometer Technol. Appl.*, Beijing, China, 2013.

[39] A. Plaza, P. Martínez, J. Plaza, and R. Pérez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.

[40] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process.* Q5
Mag., vol. 19, no. 1, pp. 44–57, 2002.

[41] R. N. Clark *et al.*, USGS digital spectral library splib06a: U.S.
Geological Survey, Digital Data Series 231, [Online]. Available: Q6
http://library.usgs.gov/spectral.lib06.

[42] [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=
Hyperspectral_Remote_Sensing_Scenes Q7

[43] R. O. Green *et al.*, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, 1998.

[44] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[45] T. F. Coleman and Y. Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM J. Optimization*, vol. 6, pp. 1040–1058, 1996.

[46] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

Eleftheria A. Mylona received the B.Sc. degree in applied mathematical and physical sciences from the National Technical University of Athens, Athens, Greece, the M.Sc. degree in acoustics from the University of Edinburgh, Scotland, U.K., and the Ph.D. degree in image analysis from the National Kapodistrian University of Athens, Athens, in 2006, 2007, and 2014, respectively.

From 2014, she has been a Postdoctoral Researcher in the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens. She has coauthored 18 research articles on image analysis and pattern recognition and has been actively involved in EU and Greek research and development projects. Her research interests include pattern recognition, image analysis, and machine learning.

Dr. Mylona received a scholarship co-financed by the European Union and Greek national funds through the research funding program Heracleitus II.

Olga A. Sykioti received the B.S. degree in geology from the National and Kapodistrian University of Athens, Athens, Greece, in 1988. She received the M.S. degrees in remote sensing (1990), in hydrology, geostatistics, and geochemistry (1991), and the Ph.D. degree in geosciences and remote sensing (1994) from Pierre & Marie Curie University, Paris, France.

From 1995 to 2005, she was a Satellite Image Processing Postdoctoral Researcher and later as a Research Assistant in the National Observatory of Athens. During 2006–2010, she was an Assistant Researcher and since 2011 she has been an Associate Researcher at the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing of NOA. She has authored more than 50 peer-reviewed scientific publications in journals and international conference proceedings. She has participated in 24 international, European, European Space Agency and national funded research projects. Her research interests include hyperspectral and microwave remote sensing applications in geosciences with emphasis in geology/mineral detection and mapping, landscape dynamics, monitoring of natural ecosystems, ground deformation and tectonics. In 2009, she was awarded with a Fulbright Scholar fellowship.

Konstantinos D. Koutroumbas received the Diploma degree from the University of Patras, Patras, Greece (1989), the M.Sc. degree in advanced methods in computer science from the Queen Mary College of the University of London London, U.K. (1990), and the Ph.D. degree from the University of Athens, Athens, Greece (1995).

Since 2001, he has been in the Institute of Astronomy, Astrophysics, Space Applications and the Remote Sensing of the National Observatory of Athens, Greece, where he is currently a Senior Researcher. His research interests include mainly pattern recognition, time series estimation and their application to remote sensing and the estimation of characteristic quantities of the upper atmosphere. He has coauthored the books *Pattern Recognition* (first, second, third, fourth editions) and *Introduction to Pattern Recognition: A MATLAB Approach*. He has over 3000 citations in his work.

Athanasios A. Rontogiannis (M'97) received the five-year Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1991, the M.A.Sc. degree in electrical and computer engineering from the University of Victoria, Victoria, B.C., Canada, in 1993, and the Ph.D. degree in communications and signal processing from the University of Athens, in 1997.

From 1998 to 2003, he was in the University of Ioannina, Ionnina, Greece. In 2003, he joined the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens, where since 2011 he has been a Senior Researcher. His research interests are in the general areas of statistical signal processing and wireless communications with emphasis on adaptive estimation, hyperspectral image processing, Bayesian compressive sensing, channel estimation/equalization and cooperative communications. He has served at the editorial boards of the *EURASIP Journal on Advances in Signal Processing*, Springer (since 2008) and the *EURASIP Signal Processing Journal*, Elsevier (since 2011).

Dr. Rontogiannis is a Member of the IEEE Signal Processing and Communication Societies and the Technical Chamber of Greece.

QUERIES

Q1. Author: When accessing and uploading your corrections at the Author Gateway, please note we cannot accept new source files as corrections for your paper. Do not send new Latex, Word, or PDF files, as we cannot simply "overwrite" your paper. Please submit your corrections as an annotated PDF or as clearly written list of corrections, with location in paper. You can also upload revised graphics to the Gateway.–

Q2. Author: Unless invited or otherwise informed, a mandatory Excessive Paper Length charge of $200.00 per page (beginning with page 7 and beyond) is required for papers in excess of six (6) printed pages for JSTARS. If you have any questions regarding overlength page charges, need an invoice, or have any other billing questions, please contact reprints@ieee.org as they handle these billing requests.

Q3. Author: Ref. [49] citation is present in the sentence "Five thematic land cover *classes* are present in the scene ... " but we have only total 46 references in reference list. Please check and suggest.

Q4. Author: Please provide page range in Refs. [34] and [38].

Q5. Author: please provide month in Ref. [40].

Q6. Author: Please provide year in Ref. [41].

Q7. Author: Please provide bibliographic details in Ref. [42].

Q8. Author: Please provide the subject of study in Diploma and the Ph.D. degree in the biography of the author "Konstantinos D. Koutroumbas."

Spectral Unmixing-Based Clustering of High-Spatial Resolution Hyperspectral Imagery

Eleftheria A. Mylona, Olga A. Sykioti, Konstantinos D. Koutroumbas, and Athanasios A. Rontogiannis, *Member, IEEE*

***Abstract*—This paper introduces a novel unsupervised spectral unmixing-based clustering method for high-spatial resolution hyperspectral images (*HSIs*). In contrast to most clustering methods reported so far, which are applied on the spectral signature representations of the image pixels, the idea in the proposed method is to apply clustering on the abundance representations of the pixels. Specifically, the proposed method comprises two main processing stages namely: an unmixing stage (consisting of the endmember extraction and abundance estimation (AE) substages) and a clustering stage. In the former stage, suitable endmembers are selected first as the most representative pure pixels. Then, the spectral signature of each pixel is expressed as a linear combination of the endmembers' spectral signatures and the pixel itself is represented by the relative abundance vector, which is estimated via an efficient AE algorithm. The resulting abundance vectors associated with the *HSI* pixels are next fed to the clustering stage. Eventually, the pixels are grouped into clusters, in terms of their associated abundance vectors and not their spectral signatures. Experiments are performed on a synthetic *HSI* dataset as well as on three airborne *HSI* datasets of high-spatial resolution containing vegetation and urban areas. The experimental results corroborate the effectiveness of the proposed method and demonstrate that it outperforms state-of-the-art clustering techniques in terms of overall accuracy, average accuracy, and kappa coefficient.**

***Index Terms*—Abundance estimation (AE), clustering, endmember extraction (EE), hyperspectral imagery (HSI), spectral unmixing (SU).**

I. Introduction

HYPERSPECTRAL imaging has enabled applications and detailed mapping possibilities in a wide variety of Earth studies. In particular, airborne hyperspectral images (*HSIs*) offer high-spatial resolution with detailed spectral accuracy. This versatility enhances the identification, modeling, and detailed classification of various natural and man-made materials. *HSIs* are collected via hyperspectral sensors and are represented as data cubes consisting of numerous contiguous spectral bands of narrow bandwidths. A significant characteristic of *HSIs*, which makes their processing more challenging, is the presence of *mixed pixels*, which depict surface regions consisting of two or more distinct materials. The data for each mixed pixel correspond to the total reflectance of all the materials present within the pixel in numerous spectral bands from the surface depicted by the pixel, which form the *spectral signature* of the pixel. The key objectives in *HSI* processing are: 1) the detection of the constituent components of mixed *HSI* pixels as well as the proportions in which they appear, which will allow the production of abundance maps per material and 2) the identification of spectrally homogeneous regions. The first objective is tackled via spectral unmixing (SU) and the second via the use of clustering algorithms.

In this study, we focus on the problem of identifying spectrally homogeneous regions, via clustering (unsupervised) techniques, which, in contrast to their supervised counterparts, they do not require any externally labeled set of pixels. Most clustering techniques proposed in this field are applied on the spectral signature representations of the pixels. In contrast, the key idea of the proposed methodology is to apply clustering on the abundance vector representations of the *HSI* pixels, since the latter representation is likely to lead to more well-separated clusters. To this end, SU is applied first on the spectral representations of the pixels, in order to extract the corresponding abundance vectors, and then, clustering is applied on the abundance vector pixels representations.

SU [1]–[6] of *HSIs* has been widely applied to environmental studies. It consists of two main substages, namely 1) endmember extraction and 2) abundance estimation (AE). EE [7]–[11] is a challenging process since the aim is to mine the purest pixels (*endmembers*) of each spectrally distinct material of a *HSI*. The latter almost always consists of mixed pixels, which are also affected by noise spectra. Ideally, each endmember ought to have the maximum possible abundance of a single physical material present in the *HSI* under study and minimum (close to zero) abundance for the rest of the physical materials. Moreover, the determination of the number of endmembers is critical since an underestimated number may result in poor representation of the mixed *HSI* pixels under study, whereas an overestimated number may comprise a lot of mixed signatures. Popular endmember extraction algorithms (*EEAs*) include *VCA* [12], *N-FINDR* variants [13], and *MVSA* [14]. Other related algorithms are discussed in [16]–[18].

The aim of *AE* is the decomposition of the spectral signatures of mixed pixels into a selection of spectral signatures

Manuscript received September 14, 2016; revised November 21, 2016 and January 24, 2017; accepted March 13, 2017. Manuscript received September 14, 2016. This work was supported by the PHySIS Project under Contract 640174 within the H2020 Framework Program of the European Commission. *(Corresponding author: Eleftheria A. Mylona.)*

The authors are with the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens (NOA), Athens 15236, Greece (e-mail: emylona@noa.gr; sykioti@noa.gr; koutroum@noa.gr; tronto@noa.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2017.2687703

corresponding to the reflectance of pure physical materials (*endmembers*). The latter is usually extracted by the image itself via *EE* (however, in some cases they are selected from specific spectral libraries). *AE* results in a set of corresponding fractions (*abundances*), which indicate the proportion of each endmember present in a given pixel. Clearly, the ultimate success of *AE* depends heavily on the appropriate selection of endmembers. Since only a small number of the available materials' spectra are expected to be present in a *HSI* pixel (especially in high- resolution *HSIs*), the abundance vectors are expected to be *sparse*.

Clustering [19], [20] partitions a set of pixels from the input image into groups. Some of the most known clustering approaches are the *k-means* [21], the Fuzzy *C-Means* (*FCM*) [22], the Possibilistic *C-Means* (*PCM*) [23] and their variants, e.g., [24], [25]. The aforementioned algorithms are suitable for recovering compact clusters and they use specific vectors, (called *representatives*) to represent the clusters that underlie in the current dataset. In contrast to these algorithms, that provide a single data clustering, in Hierarchical Agglomerative Clustering (*HAC*) [26], [27], the data are organized into an effective hierarchy of nested clusterings. *HAC* requires a metric in order to calculate the dissimilarity between pairs of pixels and a linkage so as to measure the dissimilarity between clusters.

A. Related Work

It should be mentioned that the literature on clustering techniques applied on *HSIs* is limited. In [28], a graph data structure is generated to represent the tree crowns weighted with the Euclidean distance. A minimum spanning tree is generated using Kruskal's algorithm and edges above a length threshold are removed to generate independent clusters. In [29], an unsupervised hierarchical cluster analysis to phytoplankton pigment data is applied with the aim of discriminating different phytoplankton assemblages in open ocean environments. Several types of optical data vectors are used as input to *HAC* including objects consisting of reflectance values of hyperspectral data. Also, in [30], a new clustering algorithm, named Adaptive Possibilistic C-Means *(APCM)*, is applied on *HSIs*.

In [31], a clustering procedure is proposed, which consists of three processes: 1) EE, 2) unmixing and 3) hardening process via the winner-takes-all approach, in order to produce reconstructed pixels spectra. In [32], the proposed work utilizes the Gauss Mixture Vector Quantization algorithm to learn the mixture analysis and explores the cluster analysis with correlation distance. In [33], *SU* is combined with *k-means* cluster analysis for accurate geological mapping. The data are first classified into two categories: hydrothermal alteration areas and unaltered rocks. SU is applied to hydrothermal alteration areas and *k-means* clustering to unaltered rocks as two separate approaches. In [34], the proposed work generates classification maps based on *k-means* clustering and Gradient Flow. SU is conducted using the Max-D algorithm to automatically find endmembers. It should be highlighted that, in all previous methods, the unmixing and clustering processes are utilized as two separate steps, in the sense that their results are extracted independently from each other and are combined next.

In this paper, a novel unsupervised *SU-based clustering* method (*SUBC*) for *HSIs* is proposed. *SUBC* consists of two processing stages namely: 1) SU, which consists of an *EEA*, followed by a (sparse) AE algorithm and 2) a clustering algorithm. The first process identifies suitable endmembers based on the *VCA* algorithm [12]. Then, *AE* is applied on each image pixel, in order to provide its abundance representation, using the sparsity-promoting *BiICE* algorithm [35]. Finally, the recently proposed *APCM* clustering algorithm [30] uses the abundance representations of the pixels, in order to group them into clusters. It should be noted that the abundance pixel representations adopted in the proposed methodology ensures (in general) a common sparsity pattern for pixels in the same cluster. To the best of our knowledge, this is the first attempt of utilizing the abundance representation of pixels generated by *SU* as input to a clustering algorithm with the aim to enhance classification in *HSIs*.

The proposed *SUBC* method is evaluated on a synthetic *HSI* dataset as well as on three airborne *HSI* datasets of high-spatial resolution (the agricultural area of Salinas Valley, CA, USA, the land cover at Washington DC Mall, USA, and the urban area of the Pavia center, Italy) and its performance is compared in terms of overall accuracy (*OA*), average accuracy (*AA*) and kappa coefficient with that of state-of-the-art clustering techniques.

The paper is organized as follows. Section II introduces the proposed *SUBC* method. Section III demonstrates the results obtained by the proposed method as well as comparisons with state-of-the-art clustering algorithms. Conclusion and future research directions are summarized in Section IV.

II. Proposed *SUBC* Method

In this section, we first present the motivation and contribution of this study and then we describe in detail the proposed unmixing-based clustering algorithm.

A. Motivation and Contribution

In general, classification algorithms [36], [37] (both supervised and unsupervised) developed so far are applied directly on the L-dimensional spectral band vectors of the pixels. However, such (usually high dimensional) representations may contain a lot of redundant information, which may cause pixels depicting different areas to be not well separated from each other in the L-dimensional spectral domain. Clearly, this renders the work of the classification algorithms more difficult. Apart from the above issue, most classification schemes used for *HSI* processing do not focus on exploiting the available fine spectral resolution, that is, they do not consider at all information *within* the pixel. A further consequence of this is that such schemes do not exploit the fact that each *HSI* pixel contains only a few of the materials existing in the whole *HSI* (equivalently, the spectral signature of each pixel is expected to result from the linear combination of only a few endmember spectral signatures, which implies that the corresponding abundance vectors will be *sparse*).

The approach that we adopt in this paper in order to leverage the above issues is to employ sparsity-promoting SU techniques in order to represent each pixel by its abundance vector (with

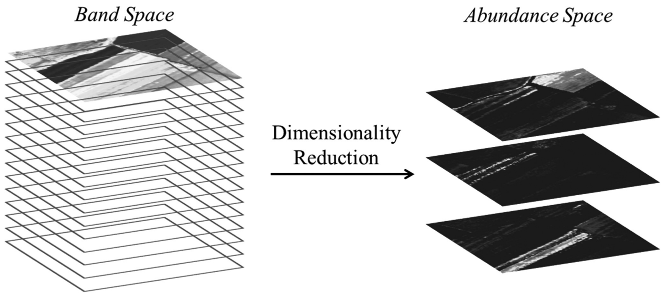


Fig. 1. Conceptual illustration of the dimensionality reduction achieved, moving from the original band space (usually consisting of hundreds of spectral bands) to the "less correlated" low-dimensional abundance space.

respect to a set of endmembers) and not by its spectral signature. The rationale behind this choice is twofold. First, the dimension of the abundance vector space (which equals to the number of the endmembers depicted in the *HSI* under study) is usually much lower than the dimension of the spectral signature space (number of spectral bands) (see Fig. 1). Since the corresponding original feature space (the space where each band defines an axis) is high dimensional, the Hughes phenomenon [38] ("curse" of dimensionality) appears. In light of this, the original high-dimensional space of the *HSI* is transformed to the dimensionally reduced space of abundance vectors [39].

Second, assuming that the endmembers are pure pixels, the (sparse) abundance vectors are expected to form clusters, which are likely to lie in different subspaces in the abundance space. It is, thus, anticipated that different classes will form more easily distinguishable clusters in the abundance vectors space. Generally speaking, adoption of the abundance representation is expected to ease the work of the classification methods. However, we have to keep in mind that the abundance retrieval requires a very good estimation of the endmembers that have a physical meaning in order to work properly, which, in practice, is not straightforward.

In the SU stage of the *SUBC* an *EEA* is first employed, which identifies appropriate endmembers of the image. Next, a sparse *AE* algorithm is used that is based on the endmembers extracted by the *EEA*, in order to produce the abundance fractions for each pixel, which in turn form the abundance vector of the pixel. These vectors of all pixels are fed to the second stage of the *SUBC* method, where a clustering algorithm groups pixels based on their abundance representations.

An additional feature concerning the mapping to the abundance space that should be highlighted is that the number of clusters and the number of endmembers are (in general) different. A cluster formed according to the abundances usually corresponds to a region where a single (or a few) endmembers have high proportion, whereas all other endmembers have low proportions. However, it can also correspond to the mixture of several endmembers of varied proportions. The block diagram of *SUBC* is depicted in Fig. 2.

B. Spectral Unmixing

1) Endmember Extraction: Aiming at detecting suitable endmembers, we utilize the *VCA* algorithm [12], which takes

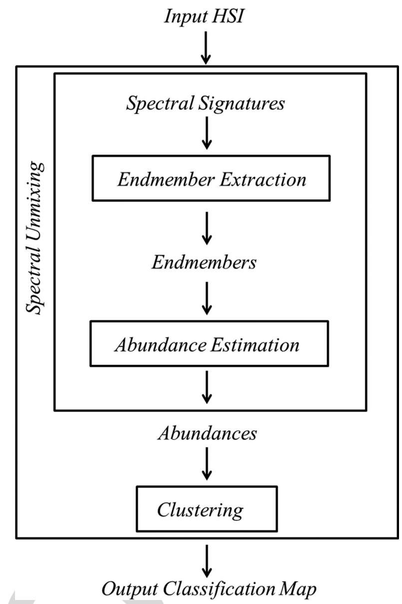


Fig. 2. Block diagram of *SUBC*.

as input the spectral signatures of the pixels, as can be seen in Fig. 2. Each pixel can be viewed as a vector in an L-dimensional Euclidean space, where each spectral band is assigned to one axis of the space. Based on the aforesaid data points, the *VCA* algorithm returns a prespecified number of endmembers via iteratively projecting data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until the number of endmembers is exhausted [12]. Then, *SUBC* continues in estimating the abundance fractions of each endmember via *AE*.

2) Abundance Estimation: The selection of appropriate endmembers is crucial so as to correctly estimate the abundance fractions. Usually, the spectral signature of the pixel, denoted by $\boldsymbol{y}$, is assumed to follow the Linear Mixing Model [40] according to which it can be expressed as a linear combination of its endmembers' spectra as follows:

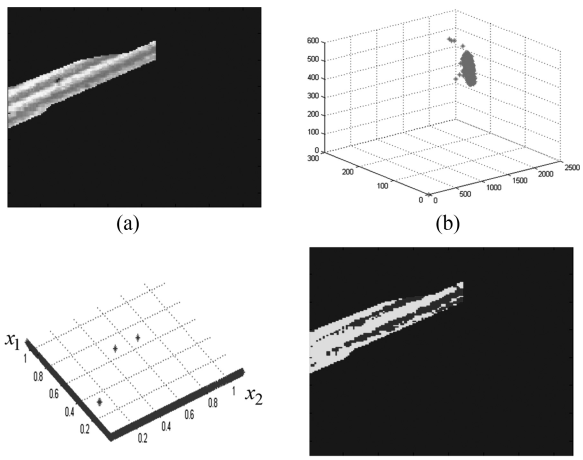
$$y = \Phi x + n \tag{1}$$

where $\Phi = [\varphi_1, \varphi_2, ..., \varphi_p] \in \Re_+^{L\times p},\ L \gg p$, is the mixing matrix comprising the endmembers' spectra (L-dimensional vectors $\phi_{i,}\ i = 1,\ 2, ..., p$), $\boldsymbol{x}$ is a $p \times 1$ vector consisting of the corresponding abundance fractions, named *abundance vector*, and $\boldsymbol{n}$ is an $L \times 1$ additive noise vector, which is assumed to be a zero-mean Gaussian distributed random vector with independent and identically distributed elements.

Due to the physical constraints of the unmixing problem, the abundance fractions for each pixel should satisfy the following two constraints:

$$x_i \geq 0,\ i = 1,\ 2,\ ...,\ N, \quad \sum_{i=1}^{N} x_i = 1 \tag{2}$$

that is, the abundances should be nonnegative and they must sum to 1. Furthermore, the *abundance* vector is expected to be sparse, i.e., only a few of its elements will be nonzero, since the



(a) (b) (c) (d)

Fig. 3. (a) Class of the *HSI* containing two subclasses; (b) representation of pixels in the original space; (c) representation of abundance vectors $x = x_1, x_2$; and (d) clustering result emerged from *SUBC*.

area depicted by a single pixel is likely to embed only a small fraction of the different materials encountered in the whole *HSI*.

In this study, the abundance vector for each pixel is estimated via a variational Bayes algorithm called *BiICE* [35] (see Fig. 2) that imposes sparsity on the abundance vector and is based on an appropriately defined hierarchical Bayesian model. In algorithmic form, the abundance vector can be obtained as follows:

$$x = BiICE(\Phi, y). \tag{3}$$

BiICE is computationally efficient, provides sparse solutions without requiring the fine-tuning of any parameters, and converges fast to accurate values even for highly correlated data. The determined abundance vectors ***x*** are further used for the representation of their associated pixels at the *clustering* process.

In order to unravel the advantages of using the abundance representation of the pixels instead of the traditional band representation, we consider the following simplified case. For illustration purposes, we form an *RGB* image selecting three appropriate bands from a small area of one of the *HSIs* considered in Section III-B. The considered area [see Fig. 3(a)] is a class consisting of two subclasses. The representation of the pixels in the original space is depicted in red color in Fig. 3(b). Assuming two endmembers (one from each subclass), Fig. 3(c) depicts the abundance vectors stemmed from *BiICE* in blue color. Note that, due to the imposed sparsity, almost all pixels are concentrated around the two axes. Finally, Fig. 3(d) depicts the classification map produced by *SUBC*.

It should be highlighted that the pixels in the original space formulate one compact cloud with a few outliers and, thus, it is difficult to be naturally divided into two separate groups. On the contrary, the abundance vectors formulate two compact clouds tangent to the axes, which are highly distinguished. This is the advantage that characterizes the representation of the pixels using their abundance vectors and eases *SUBC* to correctly identify the two subclasses, via its second stage. It should also be reminded here that the abundance vectors are characterized by sparsity (i.e., the existence of zeros in vectors ***x***), which promotes data distinctions.

C. Clustering

The *clustering* stage, which is applied on the abundance representations of the *HSI* pixels under study, employs the *APCM* algorithm [30] (see Fig. 2). Let $X = \{x_i \in \Re^p, i = 1, ..., N\}$ be a set of *N p*-dimensional data vectors to be clustered and $\Theta = \{\theta_j \in \Re^p, j = 1, ..., m\}$ be a set of *m* vectors (called *representatives*) that will be used for the representation of the clusters formed by the points in *X*. Let $U = [u_{ij}], i = 1, ..., N, j = 1, ..., m$ be an $N \times m$ matrix whose (i, j) entry stands for the so-called *degree of compatibility* of $\mathbf{x}_i$ with the *j*th cluster denoted by C_j and represented by the vector θ_j. The *APCM* algorithm emerges from the optimization of the cost function of the original *PCM* described as follows:

$$J_{PCM}(\Theta, U) = \sum_{j=1}^{m} [\sum_{i=1}^{N} u_{ij} ||\mathbf{x}_i - \theta_j||^2 + \gamma_j \sum_{i=1}^{N} (u_{ij} \ln u_{ij} - u_{ij})]. \tag{4}$$

In contrast to the classical *PCM,* where $\gamma_j{}'s$ remain constant during the execution of the algorithm, in *APCM* $\gamma_j{}'s$ are adapted at each iteration through the adaptation of the corresponding $\eta_j{}'s$. This is achieved by setting $\gamma_j = \frac{\hat{\eta}}{\alpha}\eta_j$ and adapting η_j (which is a measure of the mean absolute deviation of the current form of cluster C_j) at each iteration of the algorithm. Note that $\eta_j{}'s$ and α are constant quantities (for more details see [30]).

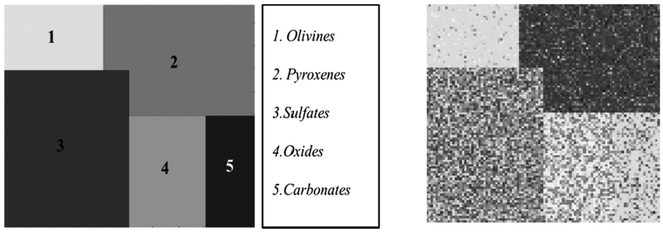
The output of the algorithm is a classification map consisting of clusters formed based on the abundances produced in *SU*. The clusters that are formed usually correspond to regions where a few abundances have high values of fractions, whereas the remaining ones exhibit low values (that is, they are aggregated around certain subspaces in the abundance space).

III. Experimental Results and Discussion

SUBC has been experimentally evaluated in four case studies: a synthetic and three real airborne *HSI* datasets of high-spatial resolution. The synthetic *HSI* dataset has been generated with various values of additive noise in order to test the sensitivity of the proposed method under different noise levels. The first airborne *HSI* dataset represents a challenging area of various plant species on an agricultural area, where discrimination between the species is impeded by numerous factors such as the similar spectral signatures of the pixels as well as the absence of reference spectra. The second airborne *HSI* dataset represents a land cover of mixed vegetation and urban materials whose spectral signatures patterns vary. The third airborne *HSI* dataset represents a mainly urban area, where the spectral signatures of the materials present are not characterized by specific patterns.

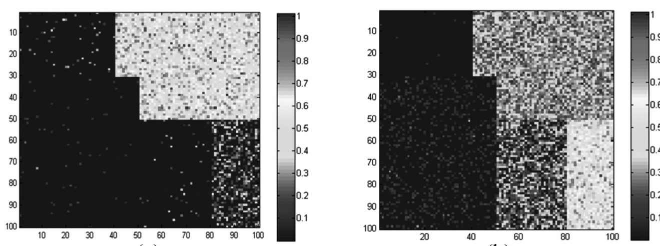
A. Synthetic HSI Dataset

The experimental evaluation of *SUBC* has been conducted on a 100×100 synthetic *HSI* dataset consisting of five different



(a) (b)

Fig. 4. (a) Reference map of synthetic *HSI* dataset and (b) 100th band added with noise at 20 dB.



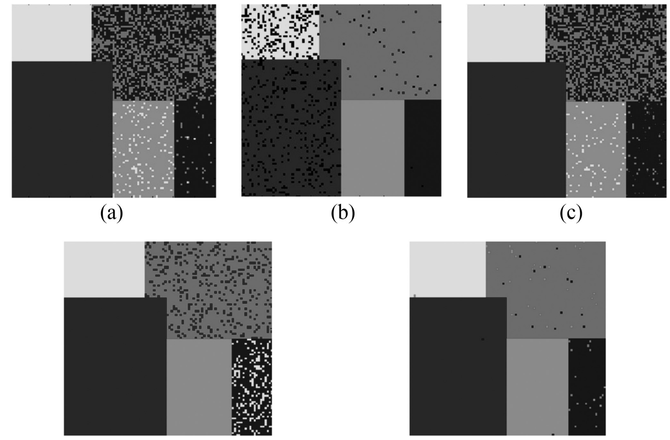
(a) (b)

Fig. 5. Estimated abundance maps for two endmembers (a) pyroxenes and (b) carbonates extracted from synthetic *HSI* via *BiICE*. Abundance values range from 0 (blue) to 1 (red).

regions artificially generated. The spectral signatures have been obtained by the U.S. Geological Survey Spectral Library [41]. The data cube contains areas with mineral signatures of five general mineral classes: 1) olivines; 2) pyroxenes; 3) sulfates; 4) oxides; and 5) carbonates. The *HSI* under study comprises 109 spectral bands. For the generation of the synthetic hyperspectral data cube, seven endmembers have been randomly selected and for each mineral class seven pure pixels have been assigned. It should be highlighted that for each mineral class more than one endmembers have been randomly assigned. Each one of the five regions consists of a linear combination of different randomly selected different endmembers contaminated by additive Gaussian zero mean noise.

Fig. 4(a) depicts the reference map, while Fig. 4(b) shows the 100th band of the synthetic *HSI* contaminated by 20-dB additive noise. It should be noted that noise is added in all bands of the synthetic *HSI* dataset and experiments have been conducted with different *SNRs* in the range of 20–40 dB. Fig. 5 illustrates abundance maps obtained from *BiICE* for two endmembers 1) pyroxenes and 2) carbonates extracted from the synthetic *HSI* under study. In Fig. 6, *SUBC* is compared with state-of-the-art clustering algorithms namely *k-means*, complete-link *HAC*, *FCM*, and *APCM*. It should be highlighted that all these algorithms are applied on the spectral signatures of the pixels, whereas the clustering procedure in *SUBC* is applied on the abundance representations of the pixels (due to the philosophy of the method). As shown in Fig. 6, classes 1, 3, and 4 are correctly identified by all tested algorithms, while the superiority of the proposed *SUBC* algorithm is clearly demonstrated in the identification of classes 2 and 5.

Table I contains the results obtained by *k-means*, *HAC*, *FCM*, *APCM*, and *SUBC* in terms of *OA* and *kappa* coefficient based on the obtained confusion matrix for 20-dB *SNR* [33]. Table II demonstrates the results in terms of *AA* (fraction of true positives and true negatives) for each class. We observe that *SUBC* outperforms all existing clustering techniques and offers an almost 100% *OA* and *AA*. It should be noted here that similar results have also beenobtained for all other values of *SNR* tested in the range 20–40 dB.



(a) (b) (c)

(d) (e)

Fig. 6. Clustering results emerged from: (a) *k-means*; (b) *HAC*; (c) *FCM*; (d) *APCM*; and (e) *SUBC* on the synthetic HSI under study.

TABLE I
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SYNTHETIC *HSI* DATASET IN TERMS OF *OA* AND *Kappa Coefficient*

	OA(%)	kappa
k-means	86.95	0.76
HAC	93.51	0.87
FCM	90.01	0.89
APCM	97.73	0.90
SUBC	99.28	092

TABLE II
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SYNTHETIC *HSI* DATASET IN TERMS OF *AA* FOR EACH CLASS

Class	k-means	HAC	FCM	APCM	SUBC
1	99.76	92.51	99.70	99.72	99.78
2	77.52	93.21	77.20	97.31	98.66
3	99.24	82.24	99.30	99.25	99.26
4	75.03	99.69	85.87	99.53	99.50
5	83.20	99.90	87.98	92.84	99.20

B. Airborne HSI Datasets

SUBC has been also experimentally evaluated on the *HSI* airborne dataset of the Salinas Valley, CA, USA [42], which constitutes an arduous clustering scenario. Salinas *HSI* has been collected by the Airborne Visible Infra-Red Imaging Spectrometer (*AVIRIS*) sensor over an agricultural area of Salinas Valley, California. The *AVIRIS* sensor, developed by NASA's Jet Propulsion Laboratory [43], generates calibrated radiance

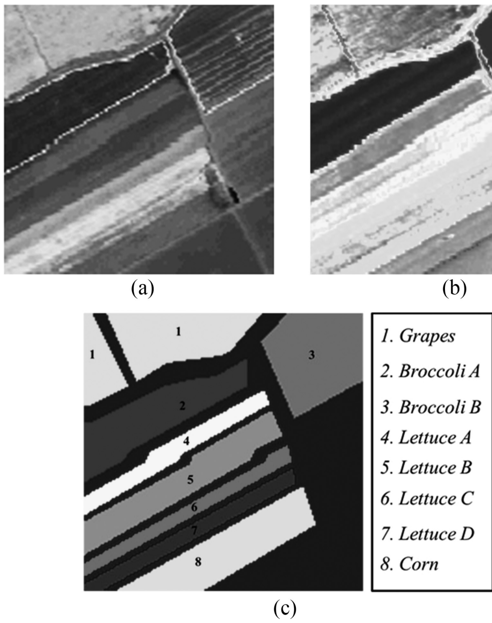


Fig. 7. (a) First PCA band; (b) 117th band of Salinas Valley; and (c) masked reference map [42].

images in 224 contiguous spectral bands with wavelengths from 400 to 2500 nm. Moreover, it is characterized by high-spatial resolution of 3.7-m pixels. The number of bands is reduced to 204 by removing 20 water absorption bands. Salinas Valley *HSI* consists of vegetables and vineyard fields. Its masked reference classification map comprises eight classes: corn, two types of broccoli, four types of lettuce and grapes [42]. Fig. 7 depicts: (a) first *PCA* band; (b) 117th band; and (c) masked reference map of a 150×150 subimage of the Salinas Valley *HSI*.

Ideally, one would have a digital spectral library of reference spectra of the mapped plant species. However, such a publicly available library does not exist for the specific plant species. In addition, it is not known how many spectra would be required to represent the changing spectral signatures, as a function of the growing season. This unavoidably leads to the selection of the endmembers from the image itself. Doing so, Fig. 8(a)–(d) depict estimated abundance maps stemmed from *BiICE* for four endmembers extracted from Salinas Valley *HSI*. Fig. 8(a) and (b) correspond to two types of broccoli, Fig. 8(c) to one type of grapes and Fig. 8(d) to a (most probably) construction.

Aiming at a quantitative evaluation, *SUBC* is compared against *k-means*, *HAC*, *FCM*, and *APCM* in terms of *OA* and *AA* computed by the obtained confusion matrix as can be seen in Tables III and IV, respectively. We see from Tables III and IV that *SUBC* achieves *OA*, *kappa*, and *AA* values which are higher than that of the other state-of-the-art clustering algorithms. Fig. 9 illustrates clustering results emerged from: (a) *k-means;* (b) *HAC;* (c) *FCM*; (d) *APCM*; and (c) *SUBC* on the Salinas *HSI* dataset. It should be mentioned that the results obtained by *APCM* and *SUBC* demonstrate the correct identification of all classes and subclasses as can be seen by examining the first *PCA* band in Fig 7(a).

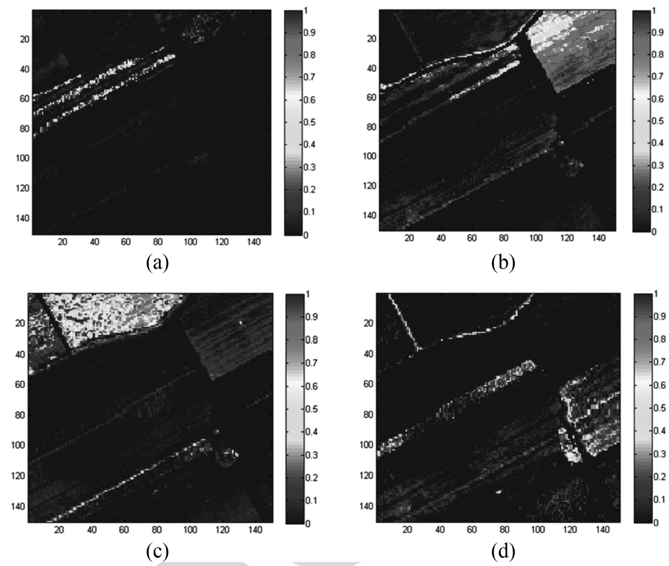


Fig. 8. Estimated abundance maps for four endmembers extracted from Salinas Valley *HSI* via *BiICE*. Abundance values range from 0 (blue) to 1 (red).

TABLE III
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SALINAS *HSI* DATASET IN TERMS OF *OA* AND *Kappa Coefficient*

	OA(%)	kappa
k-means	72.67	0.70
HAC	87.07	0.75
FCM	82.46	0.70
APCM	91.34	0.78
SUBC	93.04	0.80

TABLE IV
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON SALINAS *HSI* DATASET IN TERMS OF *AA* FOR EACH CLASS

Class	k-means	HAC	FCM	APCM	SUBC
Grapes	73.67	94.27	74.92	87.92	94.77
Broccoli A	74.43	73.82	92.83	92.79	93.49
Broccoli B	73.56	73.93	90.12	90.82	91.52
Lettuce A	72.43	89.38	72.81	92.27	93.37
Lettuce B	73.21	91.59	70.62	91.39	92.36
Lettuce C	70.23	92.72	91.29	92.12	92.79
Lettuce D	71.54	93.91	92.46	93.24	93.52
Corn	72.29	86.94	74.63	90.17	92.50

SUBC has also been quantitatively evaluated on the *HSI* airborne dataset of the Pavia Center [42]. The image has been acquired by the reflective optics system imaging spectrometer sensor over an urban area of the city center. The flight was operated by the German Aerospace Agency under the HySens project managed by the German Aerospace Center (*DLR*). The original data consist of 115 spectral bands (with the spectral range from 0.43 to 0.86 μm) and has a high-spatial resolution of 1.3 m. However, noisy bands were previously removed leading to a total of 102 bands. Four thematic classes are present in the

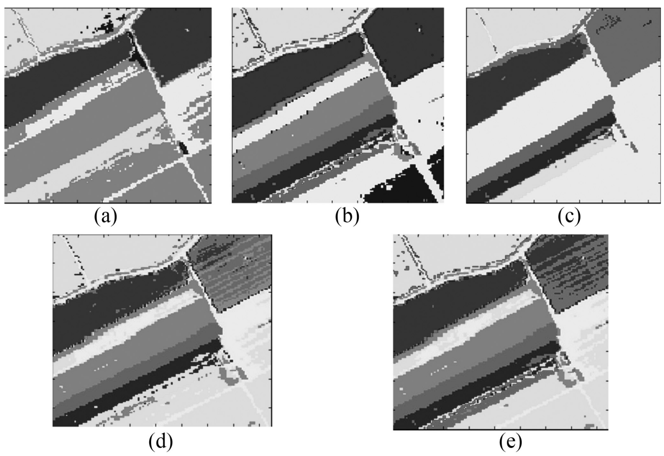


Fig. 9. Clustering results emerged from: (a) *k-means*; (b) *HAC*; (c) *FCM*; (d) *APCM*; and (e) *SUBC* on the Salinas *HSI*.

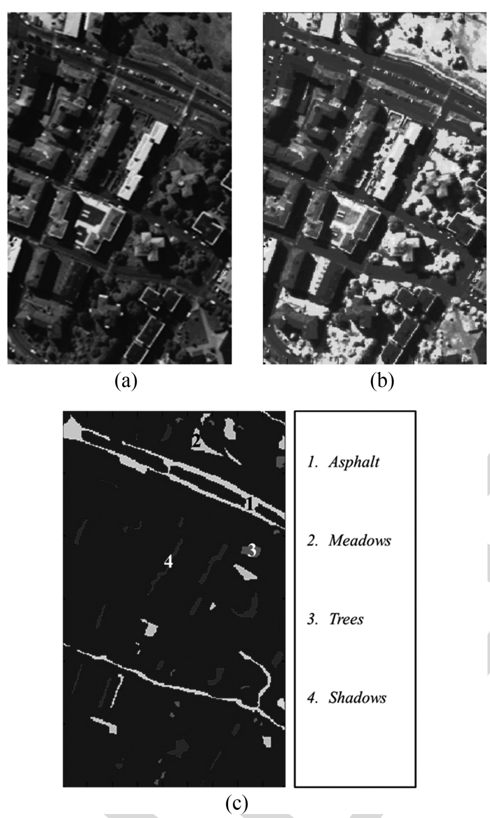


Fig. 10. (a) First PCA band; (b) 80th band of Pavia center; and (c) masked reference map [42] (1-yellow, 2-light blue, 3-dark blue, and 4-brown).

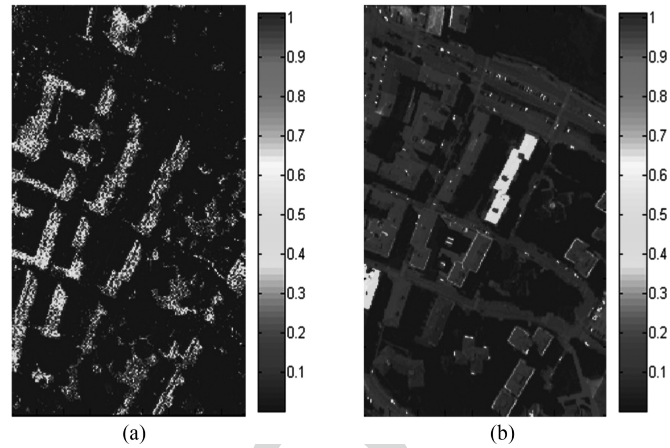


Fig. 11. Estimated abundance maps for two endmembers (a) shadow, (b) manmade material extracted from Pavia center *HSI* via *BiICE*. Abundance values range from 0 (blue) to 1 (red).

TABLE V
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON PAVIA *HSI* DATASET IN TERMS OF *OA* AND *Kappa Coefficient*

	OA(%)	Kappa
k-means	93.26	0.80
HAC	37.03	0.71
FCM	92.46	0.78
APCM	93.38	0.79
SUBC	96.30	0.83

TABLE VI
COMPARATIVE RESULTS OF CLUSTERING ALGORITHMS ON PAVIA *HSI* DATASET IN TERMS OF *AA* FOR EACH CLASS

Class	k-means	HAC	FCM	APCM	SUBC
Asphalt	94.28	25.93	94.90	95.01	97.31
Meadows	90.62	16.72	92.61	91.68	96.71
Trees	92.25	21.09	86.51	90.47	94.37
Shadows	95.89	84.39	95.82	96.36	96.81

scene: 1) asphalt; 2) meadows; 3) trees; and 4) shadows, according to the reference classification map provided by [42]. Fig. 10 depicts: (a) first *PCA* band; (b) 80th band; and 3) masked reference map of a 300 × 177 subimage of the Pavia center *HSI* [42]. Fig. 11(a) and (b) depicts estimated abundance maps stemmed from *BiICE* for two endmembers: (a) shadow and (b) manmade material.

In the scope of a quantitative evaluation, *SUBC* is compared against *k-means*, *HAC*, *FCM*, and *APCM* in terms of the *OA* and *kappa* coefficient computed by the obtained confusion matrix as can be seen in Table V and in terms of the *AA* as can be seen in Table VI, while the clustering results of all algorithms are shown in Fig. 12. Again, *SUBC* provides the best clustering performance as witnessed by its *OA, kappa*, and *AA* values, which are the highest among all its competitors.

Finally, *SUBC* has been qualitatively evaluated on the *HSI* airborne dataset of the Washington DC mall [44]. The image has been acquired by the airborne mounted Hyperspectral Digital Imagery Collection Experiment sensor. The sensor system used in this case measured pixel response in 210 bands in the 0.4–2.4 μm region of the visible and infrared spectrum. Bands in the 0.9–1.4 μm region, where the atmosphere is opaque, have been omitted from the dataset leaving 191 bands. Moreover, the dataset exhibits high-spatial resolution (2.8 m). Five thematic land cover *classes* are present in the scene: 1) roof; 2) Q3 grass; 3) trees; 4) water; and 5) asphalt road, according to the

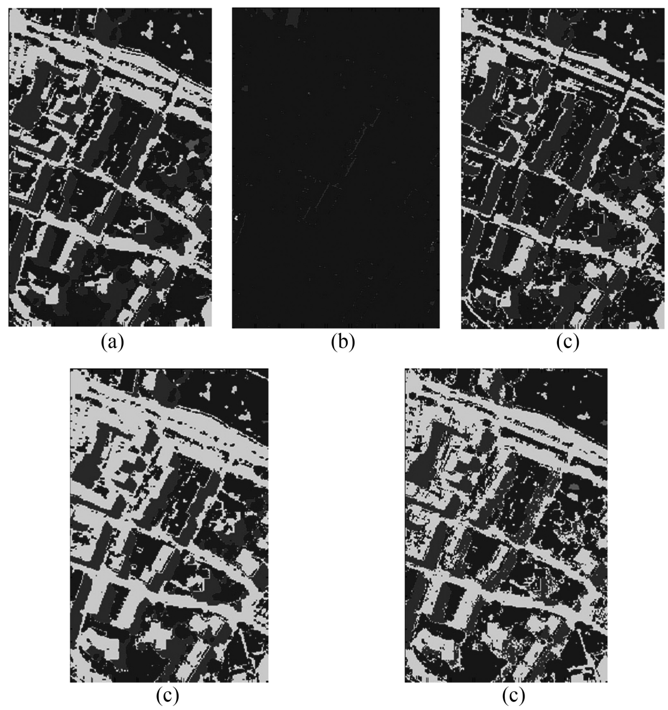


Fig. 12. Clustering results emerged from: (a) *k-means*, (b) *HAC*, (c) *FCM*, (d) *APCM* and (e) *SUBC* on the Pavia center *HSI*.

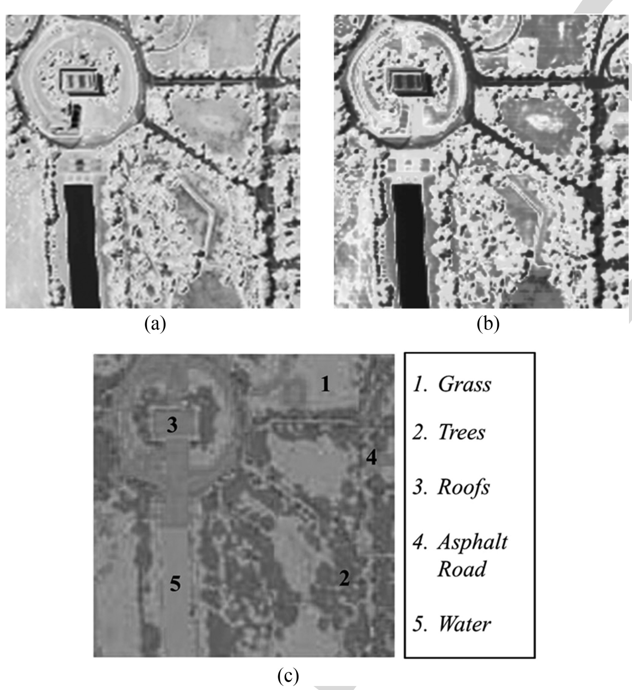


Fig. 13. (a) First PCA band; (b) 100th band of Washington DC; and (c) reference map [44].

classification map provided by [49] and used here as a reference map.

Fig. 13 depicts: (a) first *PCA* band; (b) 100th band; and (c) reference map of a 100×100 subimage of the Washington DC mall *HSI* [44]. It should be noticed that the reference map is provided only for qualitative visualization assessment and it is not accurate for a thorough quantitative assessment. Fig. 14(a)

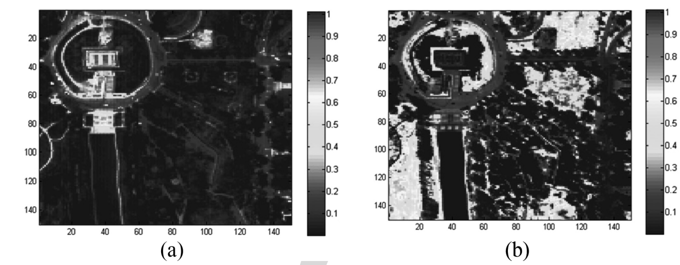


Fig. 14. Estimated abundance maps for two endmembers: (a) manmade material and(b) (most probably) soil/grass of class 1 extracted from Washington DC HSI via BiICE. Abundance values range from 0 (blue) to 1 (red).

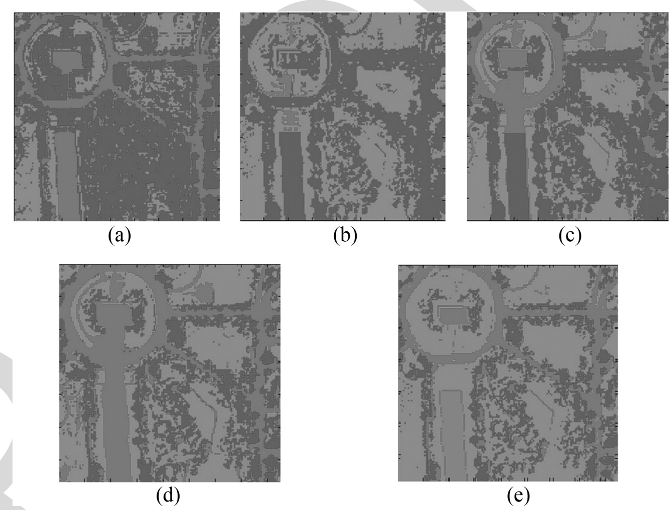


Fig. 15. Clustering results emerged from: (a) k-means; (b) HAC; (c) FCM; (d) APCM; and (e) SUBC on the Washington DC HSI.

and (b) depicts estimated abundance maps stemmed from *BiICE* for two endmembers: 1) manmade material and 2) (most probably) soil/grass of class 1. Fig. 15 illustrates clustering results emerged from: (a) *k-means*; (b) *HAC*; (c) *FCM*; (d) *APCM*; and (e) *SUBC* on the Washington DC *HSI* dataset. It should be highlighted that, apart from *SUBC*, all other algorithms falsely classify water and asphalt road pixels to one class. On the other hand, *SUBC* correctly distinguishes pixels that belong to the water class from all other pixels that belong to the remaining classes.

As it has been highlighted throughout the paper, the key idea of the proposed method is to perform unmixing at its first stage, in order to take the abundance representations of the pixels and then, at the second stage, to perform clustering based on the pixels abundance vector representations. Clearly, one could choose any unmixing method in the first stage and any clustering method in the second stage of the algorithm. In order to justify the choice of *BiICE* in the first stage, we compare it against two *AE* algorithms: 1) a quadratic programming (*QP*) technique [45], which does not exploit sparsity and 2) the sparse unmixing by variable splitting and augmented Lagrangian (*SUnSAL*) algorithm [46], which, as *BiICE*, imposes sparsity. That is, we substitute *BiICE* with *QP* and *SUnSAL* at the first stage of the proposed method. Leaned on Table VII, which depicts the *OA*

TABLE VII
COMPARATIVE RESULTS OF *SU* ALGORITHMS ON *HSI* DATASETS IN TERMS OF *OA*

	Synthetic	Salinas	Pavia
QP	83.67	74.32	71.20
SUnSAL	97.50	82.71	87.52
BiICE	99.28	93.04	96.30

of the three cases, the *QP* algorithm attains the worst performance (since it does not take into account that by the nature of the problem, the abundance vectors exhibit sparsity), whereas *SUnSAL* exhibits significantly improved performance compared to *QP*, yet inferior, compared to *BiICE,* especially for real data. Moreover, *SUnSAL* comes at the additional expense of manually fine-tuning nontrivial parameters, such as a sparsity promoting parameter λ.

The choice of *APCM* in the second stage of the algorithm is justified mainly by the fact that it is able to estimate automatically the underlying number of clusters in the dataset. Moreover, focusing on the first four lines of Tables I, III, and IV, the *OA* of *APCM* is significantly higher from all other state of the art clustering methods (note that all these algorithms are applied on the same dataset, i.e., the spectral signature representations of the *HSI* pixels).

IV. CONCLUSION AND FUTURE DIRECTIONS

The key challenge of the proposed method (*SUBC*) is the identification of spatially homogeneous regions comprising different materials. The method consists of two main stages (unmixing and clustering) and generates three significant (by)products, namely: 1) endmembers; 2) abundance vectors (abundance maps); and 3) clusters (classification maps). The key feature of *SUBC* is the utilization of the abundance representations of the *HSI* pixels (as they result from the unmixing stage) in the clustering stage. The advantage of using the abundance representation instead of the basic spectral representation of the pixels is that the former, in contrast to the latter, provides subpixel level information, which in turn favors more detailed classification maps. Moreover, the abundance representation is likely to give rise to more well-discriminated clusters that live on subspaces of the abundance space, due to the fact that only a few materials are expected to contribute to the formation of a *HSI* pixel (sparsity issue). As a consequence, subspace clustering algorithms could also be considered as an alternative in the final stage of the algorithm, since the abundance representations are likely to lead to clusters that live to subspaces of the abundance space. *SUBC* is unsupervised and does not require class information knowledge of the dataset under study. Moreover, it is image independent, it alleviates the "curse of dimensionality" issue and enhances localization and accuracy since it operates in the subpixel level of information. However, it is noted again that the correct identification of the endmembers number and their correspondence to physical objects/materials is undoubtedly the most critical step for successful SU and, as a consequence, for the clustering processes.

Experimental results show that *SUBC* compares favorably to other related methods. This gives us confidence to claim that the performance of the proposed method remains consistent with high-spatial resolution airborne data. It is capable of identifying compact regions and spectral regions that lack training data.

In terms of future directions, the full potential of this algorithm will be investigated with additional hyperspectral acquisitions of higher mixture complexity. In addition, this study could be reinforced and expanded in the case of existing and future satellite hypespectral data imagery of lower spatial resolutions where increased complexity issues for the tasks of 1) endmember identification; 2) resolving shadowing effects; and 3) facing oblique viewing and illumination angles arise. Moreover, subspace clustering algorithms could be utilized, since as we discussed earlier, they suit nicely in the nature of the problem in the abundance space.

ACKNOWLEDGMENT

The AVIRIS data were provided courtesy of NASA/JPL-Caltech in Pasadena, CA, USA, via the EXELIS Visual Information Solutions Inc., 2012.

REFERENCES

[1] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised multitemporal spectral unmixing for detecting multiple changes in Hyperspectral Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2733–2748, May 2016.

[2] C. Shi and L. Wang, "Incorporating spatial information in spectral unmixing: A review," *Remote Sens. Environ.*, vol. 149, pp. 70–87, 2014.

[3] T. Meyer and G. S. Okin, "Evaluation of spectral unmixing techniques using MODIS in a structurally complex savanna environment for retrieval of green vegetation, nonphotosynthetic vegetation and soil fractional cover," *Remote Sens. Environ.*, vol. 161, pp. 122–130, 2015.

[4] Y. Yuan, M. Fu, and X. Lu, "Substance dependence constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 2975–2986, Jun. 2015.

[5] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[6] Y. Yuan, Y. Feng, and X. Lu, "Projection-based subspace learning for hyperspectral unmixing," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2632–2643, Jun. 2015.

[7] B. Zhang, X. Sun, L. Gao, and L. Yang, "Endmember extraction of hyperspectral remote sensing images based on the discrete particle swarm optimization algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4173–4176, Nov. 2011.

[8] P. E. Dennison and D. A. Roberts, "Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE," *Remote Sens. Environ.*, vol. 87, pp. 123–135, 2003.

[9] C. Song, "Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?" *Remote Sens. Environ.*, vol. 95, pp. 248–263, 2005.

[10] R. L. Powell, D. A. Roberts, P. E. Dennison, and L. L. Hess, "Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil," *Remote Sens. Environ.*, vol. 106, pp. 253–267, 2007.

[11] T. C. Eckmann, D. A. Roberts, and C. J. Still, "Using multiple endmember spectral mixture analysis to retrieve subpixel fire properties from MODIS," *Remote Sens. Environ.*, vol. 112, pp. 3773–3783, 2008.

[12] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[13] T. H. Chan, W. K. Ma, A. Ambikapathi, and C. Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.
[14] J. Li *et al.*, "Minimum volume simplex analysis: a fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.
[15] A. Plaza and C.-I. Chang, "Impact of initialization on design of endmember extraction algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3397–3407, Nov. 2006.
[16] A. Zare, P. Gader, O. Bchir, and H. Frigui, "Piecewise convex multiple-model endmember detection and spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2853–2862, May 2013.
[17] A. Zare, P. Gader, and G. Casella, "Sampling piece-wise convex unmixing and endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1655–1665, Mar. 2013.
[18] B. Somers, M. Zortea, A. Plaza, and G. P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 396–408, Apr. 2012.
[19] M. R. Anderberg, *Cluster Analysis for Applications*. New York, NY, USA: Academic, 1973.
[20] B. S. Everitt, S. Landau, and M. Leese, *Cluster Analysis*. London, U.K.: Wiley, 2001.
[21] R. O. Duda and P. Hart, *Pattern Classification and Scene Analysis*. Hoboken, NJ, USA: Wiley, 1973.
[22] J. C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.
[23] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.
[24] S. D. Xenaki, K. D. Koutroumbas, and A. A. Rontogiannis, "Sparsity-aware possibilistic clustering algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 6, pp. 1611–1626, Dec. 2016.
[25] R. Inokuchi and S. Miyamoto, "Sparse possibilistic clustering with L1 regularization," in *Proc. IEEE Int. Conf. Granular Comput.*, Fremont, CA, USA, 2007, p. 442.
[26] S. Theodoridis and K. Koutroumbas, *Pattern Recognition Applications*, 4th ed. New York, NY, USA: Academic, 2008.
[27] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.
[28] P. Bunting, R. M. Lucas, K. Jones, and A. R. Bean, "Characterisation and mapping of forest communities by clustering individual tree crowns," *Remote Sens. Environ.*, vol. 114, pp. 2536–2547, 2010.
[29] E. Torrecilla, D. Stramski, R. A. Reynolds, E. Millán-Núnez, and J. Piera, "Cluster Analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean," *Remote Sens. Environ.*, vol. 115, pp. 2578–2593, 2011.
[30] S. D. Xenaki, K. D. Koutroumbas, and A. A. Rontogiannis, "A novel adaptive possibilistic clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 4, pp. 791–810, Aug. 2016.
[31] H. Gholizadeh, M. J. V. Zoej, and B. Mojaradi, "A novel hyperspectral image clustering method based on spectral unmixing," in *Proc. Aerosp. Conf.*, 2012, pp. 1–5.
[32] M. Parente and A. Zymnis, "Statistical clustering and mineral spectral unmixing in AVIRIS hyperspectral image of Cuprite, NV," CS2009 Report, 2005.
[33] N. Ishidoshiro *et al.*, "Geological mapping by combining spectral unmixing and cluster analysis for hyperspectral data," in *Proc. Int. Archives Photogrammetry, Remote Sens. Spatial Inform. Sci.*, 2016, pp. 431–435.
[34] J. D. Kwong, D. W. Messinger, and W. D. Middleton, "Hyperspectral
Q4 clustering and unmixing for studying the ecology of state formation and complex societies," in *Proc. SPIE Opt. Eng. Appl., Int. Soc. Opt. Photon.*, 2009.
[35] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585–599, Feb. 2012.
[36] J. Li, M. Khodadadzadeh, A. Plaza, X. Jia, and J. M. Bioucas-Dias, "A discontinuity preserving relaxation scheme for spectral-spatial hyperspectral image classification," *IEEE J. Select. Topics Appl. Earth Observ.Remote Sens.*, vol. 9, no. 2, pp. 625–639, Feb. 2016.
[37] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.
[38] W. Ma, C. Gong, Y. Hu, P. Meng, and F. Xu, "The Hughes phenomenon in hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai Lake," in *Proc. Int. Symp. Photoelectron. Detection Imag. 2013: Imag. Spectrometer Technol. Appl.*, Beijing, China, 2013.
[39] A. Plaza, P. Martínez, J. Plaza, and R. Pérez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.
[40] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process.* Q5
Mag., vol. 19, no. 1, pp. 44–57, 2002.
[41] R. N. Clark *et al.*, USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231, [Online]. Available: Q6
http://library.usgs.gov/spectral.lib06.
[42] [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes Q7
[43] R. O. Green *et al.*, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, 1998.
[44] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
[45] T. F. Coleman and Y. Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM J. Optimization*, vol. 6, pp. 1040–1058, 1996.
[46] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

Eleftheria A. Mylona received the B.Sc. degree in applied mathematical and physical sciences from the National Technical University of Athens, Athens, Greece, the M.Sc. degree in acoustics from the University of Edinburgh, Scotland, U.K., and the Ph.D. degree in image analysis from the National Kapodistrian University of Athens, Athens, in 2006, 2007, and 2014, respectively.

From 2014, she has been a Postdoctoral Researcher in the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens. She has coauthored 18 research articles on image analysis and pattern recognition and has been actively involved in EU and Greek research and development projects. Her research interests include pattern recognition, image analysis, and machine learning.

Dr. Mylona received a scholarship co-financed by the European Union and Greek national funds through the research funding program Heracleitus II.

Olga A. Sykioti received the B.S. degree in geology from the National and Kapodistrian University of Athens, Athens, Greece, in 1988. She received the M.S. degrees in remote sensing (1990), in hydrology, geostatistics, and geochemistry (1991), and the Ph.D. degree in geosciences and remote sensing (1994) from Pierre & Marie Curie University, Paris, France.

From 1995 to 2005, she was a Satellite Image Processing Postdoctoral Researcher and later as a Research Assistant in the National Observatory of Athens. During 2006–2010, she was an Assistant Researcher and since 2011 she has been an Associate Researcher at the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing of NOA. She has authored more than 50 peer-reviewed scientific publications in journals and international conference proceedings. She has participated in 24 international, European, European Space Agency and national funded research projects. Her research interests include hyperspectral and microwave remote sensing applications in geosciences with emphasis in geology/mineral detection and mapping, landscape dynamics, monitoring of natural ecosystems, ground deformation and tectonics. In 2009, she was awarded with a Fulbright Scholar fellowship.

Konstantinos D. Koutroumbas received the Diploma degree from the University of Patras, Patras, Greece (1989), the M.Sc. degree in advanced methods in computer science from the Queen Mary College of the University of London London, U.K. (1990), and the Ph.D. degree from the University of Athens, Athens, Greece (1995).

Since 2001, he has been in the Institute of Astronomy, Astrophysics, Space Applications and the Remote Sensing of the National Observatory of Athens, Greece, where he is currently a Senior Researcher. His research interests include mainly pattern recognition, time series estimation and their application to remote sensing and the estimation of characteristic quantities of the upper atmosphere. He has coauthored the books *Pattern Recognition* (first, second, third, fourth editions) and *Introduction to Pattern Recognition: A MATLAB Approach*. He has over 3000 citations in his work.

Athanasios A. Rontogiannis (M'97) received the five-year Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1991, the M.A.Sc. degree in electrical and computer engineering from the University of Victoria, Victoria, B.C., Canada, in 1993, and the Ph.D. degree in communications and signal processing from the University of Athens, in 1997.

From 1998 to 2003, he was in the University of Ioannina, Ionnina, Greece. In 2003, he joined the Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens, where since 2011 he has been a Senior Researcher. His research interests are in the general areas of statistical signal processing and wireless communications with emphasis on adaptive estimation, hyperspectral image processing, Bayesian compressive sensing, channel estimation/equalization and cooperative communications. He has served at the editorial boards of the *EURASIP Journal on Advances in Signal Processing*, Springer (since 2008) and the *EURASIP Signal Processing Journal*, Elsevier (since 2011).

Dr. Rontogiannis is a Member of the IEEE Signal Processing and Communication Societies and the Technical Chamber of Greece.

QUERIES

Q1. Author: When accessing and uploading your corrections at the Author Gateway, please note we cannot accept new source files as corrections for your paper. Do not send new Latex, Word, or PDF files, as we cannot simply “overwrite” your paper. Please submit your corrections as an annotated PDF or as clearly written list of corrections, with location in paper. You can also upload revised graphics to the Gateway.–

Q2. Author: Unless invited or otherwise informed, a mandatory Excessive Paper Length charge of $200.00 per page (beginning with page 7 and beyond) is required for papers in excess of six (6) printed pages for JSTARS. If you have any questions regarding overlength page charges, need an invoice, or have any other billing questions, please contact reprints@ieee.org as they handle these billing requests.

Q3. Author: Ref. [49] citation is present in the sentence “Five thematic land cover *classes* are present in the scene . . . ” but we have only total 46 references in reference list. Please check and suggest.

Q4. Author: Please provide page range in Refs. [34] and [38].

Q5. Author: please provide month in Ref. [40].

Q6. Author: Please provide year in Ref. [41].

Q7. Author: Please provide bibliographic details in Ref. [42].

Q8. Author: Please provide the subject of study in Diploma and the Ph.D. degree in the biography of the author “Konstantinos D. Koutroumbas.”