

Classification of Hyperspectral Images by Using Extended Morphological Attribute Profiles and Independent Component Analysis

Mauro Dalla Mura, *Student Member, IEEE*, Alberto Villa, *Student Member, IEEE*, Jon Atli Benediktsson, *Fellow, IEEE*, Jocelyn Chanussot, *Senior Member, IEEE*, and Lorenzo Bruzzone, *Fellow, IEEE*

Abstract—In this letter, a technique based on independent component analysis (ICA) and extended morphological attribute profiles (EAPs) is presented for the classification of hyperspectral images. The ICA maps the data into a subspace in which the components are as independent as possible. APs, which are extracted by using several attributes, are applied to each image associated with an extracted independent component, leading to a set of extended EAPs. Two approaches are presented for including the computed profiles in the analysis. The features extracted by the morphological processing are then classified with an SVM. The experiments carried out on two hyperspectral images proved the effectiveness of the proposed technique.

Index Terms—Attribute filters, decision fusion, extended attribute profile (EAP), independent component analysis (ICA), mathematical morphology, remote sensing.

I. INTRODUCTION

THE EXPLOITATION of spatial information is very important for the classification of high-resolution hyperspectral images, particularly when considering urban areas, and it is advisable to consider geometrical features in the analysis in order to derive spatially accurate maps [1].

Manuscript received July 19, 2010; revised October 14, 2010; accepted October 25, 2010. This work was supported in part by the European community's Marie Curie Research Training Networks Program, Hyperspectral Imaging Network (HYPER-I-NET), under Contract MRTN-CT-2006-035927 and in part by the Research Fund of the University of Iceland and the University of Trento.

M. Dalla Mura is with the Department of Information Engineering and Computer Science, University of Trento, 38100 Trento, Italy, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: dallamura@disi.unitn.it).

A. Villa is with the Grenoble Images Parole Signals Automatics Laboratory, Signal and Image Department, Grenoble Institute of Technology, 38402 Saint Martin d'Hères, France, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: alberto.villa@hyperinet.eu).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is).

J. Chanussot is with the Grenoble Images Parole Signals Automatics Laboratory, Signal and Image Department, Grenoble Institute of Technology, 38402 Saint Martin d'Hères, France (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

L. Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38100 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

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/LGRS.2010.2091253

The spatial characteristics of the objects in an image can be modeled with several approaches. A widely used technique for extracting spatial features is based on mathematical morphology. Among all the operators belonging to this framework, morphological connected operators [2] proved to be suitable for extracting spatial information while preserving the geometrical characteristics of the structures in the image (i.e., without distorting the borders). In [3], morphological profiles (MPs), a sequence of multiscale connected operators, were applied to high-resolution hyperspectral images by reducing the high dimensionality of the data by principal component analysis (PCA) and computing the profiles on the first principal components extracted, leading to the definition of extended MPs (EMPs). Due to the limitations of PCA, when extracting the information sources from the high-dimensional data, it was proposed to perform independent component analysis (ICA) before the computation of the MPs [4].

The characterization of spatial information obtained by the application of an MP is particularly suitable for representing the multiscale variability of the structures in the image, but it is not sufficient to model other geometrical features. To avoid this limitation, the use of morphological attribute filters instead of the conventional operators based on the geodesic reconstruction was proposed [5]. The application of attribute filters in a multilevel way leads to the definition of attribute profiles (APs) [5], which permit to model other geometrical characteristics rather than the size of the objects. Moreover, APs showed interesting characteristics when extended to hyperspectral images [6]. In greater details, analogous to [3], the APs were applied to the first principal components extracted from a hyperspectral image, generating an extended AP (EAP).

In this letter, we extend the work presented in [7] by presenting a technique based on EAPs and ICA for the classification of hyperspectral images. Moreover, we investigate two approaches for combining the information extracted by EAPs computed with different attributes.

This letter is organized as follows. In Section II, the ICA is discussed, while in Section III, the concepts of morphological attribute filters and EAPs are presented. The proposed approaches for dealing with multiple EAPs are presented in Section IV. The experimental results are illustrated in Section V. Finally, conclusions are drawn in Section VI.

II. ICA

Hyperspectral sensors record images with hundreds of bands and a very high spectral resolution. The very detailed spectral description provided by these kinds of images increases the capability to distinguish between land-cover classes, thus achieving accurate classification maps. However, the analysis of this huge amount of data presents some methodological issues which need to be addressed. In particular, the high dimensionality of the data is a critical problem, due to the appearance of the Hughes phenomenon: After a certain threshold, if the number of features increases, the generalization capability of the classifier decreases when a fixed number of training samples are used. The threshold mainly depends on the number of samples used to train the classifier. Because of these reasons, feature reduction is often applied as a preprocessing step before the classification of hyperspectral data, in order to avoid the curse of dimensionality, and to reduce the computation time. Although it is not optimal for classification, the PCA is often used for such a task, due to its simplicity and ease of use. The principle of PCA is to project the data into an orthogonal space, so that the eigenvectors correspondent to the greatest eigenvalues retain the maximum variance of the data. Because the PCA is based on the analysis of covariance matrix and second-order statistics, it can neglect some important information, particularly when few components are retained. In this letter, we propose to use ICA for feature reduction, as an effective alternative to PCA. ICA consists of finding a linear decomposition of the observed data into statistically independent components (ICs). Given an observation model $\mathbf{x} = \mathbf{A}\mathbf{s}$, where \mathbf{x} is the vector of the observed signals, \mathbf{A} is a matrix of scalars corresponding to the mixing coefficient, and \mathbf{s} is the vector of the source signals, the ICA finds a separating matrix \mathbf{W} such that $\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}$, where \mathbf{y} is a vector of ICs.

Independence is a much stronger assumption than decorrelation, which can be obtained with PCA or factor analysis. In ICA, the concept of independence can be summarized as follows: Each component should not provide any information about higher (than second) order statistics of the other components. However, there are several methods for estimating ICA. In this letter, we have used the algorithm JADE (Joint Approximate Diagonalization of Eigenmatrices), due to good results shown when used for feature reduction in hyperspectral remote sensing data [8]. Due to space constraints, we refer the reader interested in more details about the general framework of ICA to [9].

III. EAPs

EAPs are based on the concept of the AP, which is a generalization of the widely used MPs [6]. Analogous to the definition of the EMPs, EAPs are generated by concatenating many APs. Each AP is computed on one of the c features (FRs) extracted by a feature reduction transformation (e.g., PCA) from a multivariate image (e.g., the hyperspectral image) [5]

$$EAP = \{AP(FR_1), AP(FR_2), \dots, AP(FR_c)\}. \quad (1)$$

The AP is an extension of the MP, obtained by processing a scalar grayscale image f , according to a criterion T , with n morphological attribute thickening (ϕ^T) and n attribute thinning (γ^T) operators, instead of the conventional morphological filters by reconstruction

$$AP(f) = \{\phi_n^T(f), \phi_{n-1}^T, \dots, \phi_1^T(f), f, \gamma_1^T(f), \dots, \gamma_{n-1}^T(f), \gamma_n^T(f)\}. \quad (2)$$

Attribute filters are connected operators which operate on the connected components (i.e., regions of isointensity spatially connected pixels) that compose an image, according to a given criterion [2]. The criterion associated to the transformation is evaluated on each connected component of the image. In general, the criterion compares the value of an arbitrary attribute $attr$ (e.g., area, volume, standard deviation, etc.) measured on the component C against a given reference value λ (which is the filter parameter), e.g., $T(C) = attr(C) > \lambda$. If the criterion is verified, then the regions are kept unaffected; otherwise, they are set to the gray level of a darker or brighter surrounding region, according to if the transformation performed is extensive (i.e., thickening) or antiextensive (i.e., thinning), respectively. When the criterion considered in the analysis is increasing (i.e., if it is verified for a connected component, then it will also be verified by all the regions brighter or darker, according to the transformation, including the component), the attribute thinning and thickening operators are actually opening and closing transformations. Nonincreasing criteria do not have a unique definition when considering grayscale images. In fact, different effects can be obtained by the operators with a nonincreasing criterion according to the filtering rule selected [10].

Attribute filters can be efficiently computed by taking advantage of the representation of the input image as a rooted hierarchical tree of the connected components of the image. The tree is obtained by the Max-tree algorithm [10]. The approach based on this data representation is particularly useful when computing an AP, since the image is converted to the tree only once (this is the most demanding stage of the filtering) and processed several times with the different criteria. Examples of EAPs computed with different attributes are reported in (Fig. 1).

IV. APPROACHES TO DEAL WITH MULTIPLE EAPs

The choice of the most suitable attribute and range of thresholding values (λ 's) for extracting the information on the geospatial objects is certainly a complex task, particularly when *a priori* information on the scene is not available. A possible approach attempting to overcome this issue relies on the computation of EAPs with different kinds of attributes. However, this leads to the problem of properly exploiting, in the analysis, the different information extracted by the computed EAPs.

A simple strategy is the *stacked vector approach* (SVA), which combines the EAPs by concatenating them in a single vector of features [also called extended multi-AP (EMAP) [6]; see Fig. 2(a)]. However, even if complementary information can be extracted by considering different attributes, a great redundancy is present in the features extracted. Thus, it is advisable that a classification algorithm with excellent penalization capability is used for classifying the features in order to handle

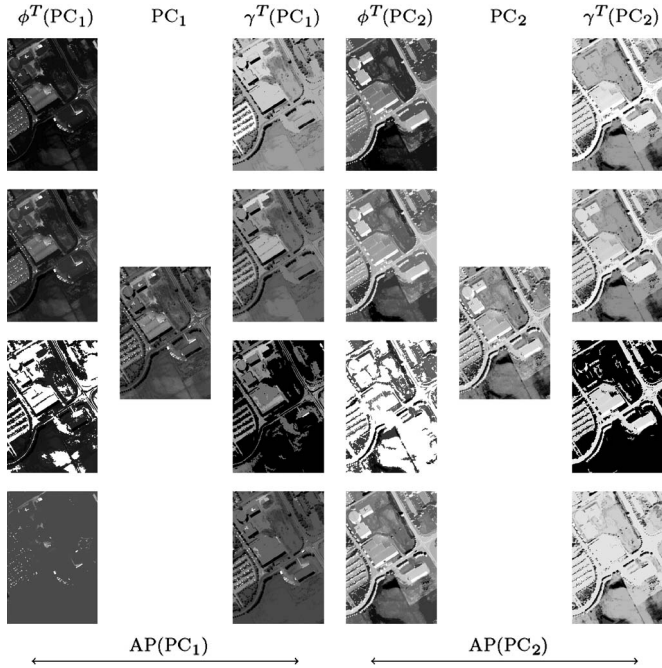


Fig. 1. Examples of EAPs computed on the first two PCs of a portion of the image Fig. 3(a). Each row shows an EAP built by different attributes. Attributes starting from the first row are area, length of the diagonal of the bounding box, moment of inertia, and standard deviation. Each EAP is composed by the concatenation of two APs computed on PC_1 and PC_2 . Each AP is composed of three levels: A thickening image ϕ^T , the original PC, and a thinning image γ^T . All the thickening and thinning transformations were computed with the following attribute values λ 's. Area: 5000; length of the diagonal: 100; moment of inertia: 0.5; and standard deviation: 50.

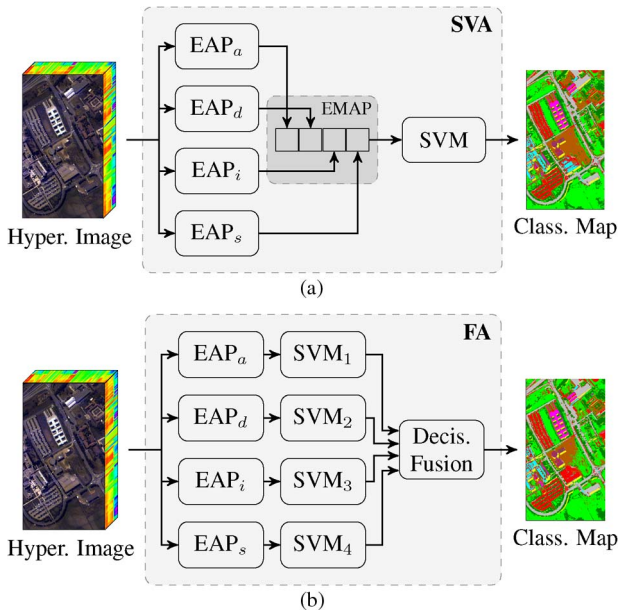


Fig. 2. Proposed approaches for dealing with multiple EAPs. (a) SVA and (b) FA.

the increased dimensionality which can lead to the Hughes phenomenon.

Another approach is the *fusion approach* (FA) that is based on the separate classification of each EAP and on the fusion of the results obtained by the independent classifiers in order to generate the final decision map [see Fig. 2(b)]. In comparison

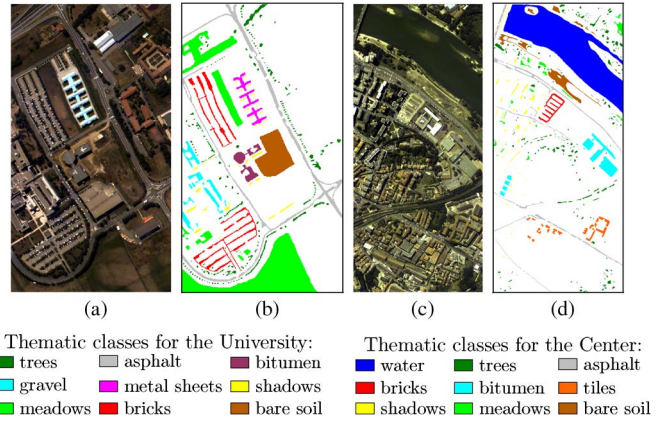


Fig. 3. ROSIS Pavia data sets: True color representation and test set for (a), (b) University and (c), (d) Center.

to the SVA, the FA keeps low the dimensionality of the data and increases the robustness of the results, particularly if the different EAPs generate complementary errors.

In this letter, an SVM classifier is considered with the one-against-one multiclass strategy. The fusion rule considered when combining the results of the single classifiers relies on the sum of the votes of the classifiers applied to the four EAPs, assigning each pixel to a class, according to the majority voting scheme. Obviously, other decision criteria can be applied.

V. EXPERIMENTAL ANALYSIS

The experimental analysis was carried out on two hyperspectral images acquired over the city of Pavia (Italy) by the ROSIS-03 (Reflective Optics Systems Imaging Spectrometer) hyperspectral sensor. The two images have a geometrical resolution of 1.3 m. The first one shows the university campus (610×340 pixels), while the second one was acquired on the city center (1096×489 pixels). In the following, we will refer to the two data sets as “University” and “Center,” respectively. The original data are composed of 115 spectral bands, ranging from 0.43 to $0.86 \mu\text{m}$ with a band of 4 nm . However, noisy bands were previously discarded, leading to 103 and 102 channels for the two images, respectively. Nine thematic land-cover classes were identified in the university campus: Trees, Asphalt, Bitumen, Gravel, Metal sheets, Shadows, Self-blocking Bricks, Meadows, and Bare soil. For this data set, a total of 3921 and 42 776 pixels were available as training and test sets, respectively. In the center area, the thematic classes found were Water, Tree, Meadow, Self-blocking Bricks, Soil, Asphalt, Bitumen, Tile, and Shadow. The training and test sets for this data set were composed of 5536 and 103 539 samples, respectively. The true color representation of the images and the test sets taken as reference are shown in Fig. 3.

In the analysis carried out, all the samples of the training set were used for the University data set, while for the Center data sets, only 50 samples (randomly chosen from the full training set for each class) were considered. All the experiments conducted on the latter data set were run ten times with a set of different training samples each time.

From both the two hyperspectral images, four components extracted by PCA and ICA were considered. The first four PCs

TABLE I
 ROSIS UNIVERSITY DATA SET. CLASSIFICATION ACCURACIES OBTAINED BY CLASSIFYING THE HYPERSPECTRAL IMAGE (SPECT.), THE FOUR COMPONENTS EXTRACTED (4 COMP.), EACH SINGLE EAP, AND THE DATA WITH THE PROPOSED APPROACHES: SVA AND FA

	Spect.	4 Comp.	EAP _a	EAP _d	EAP _i	EAP _s	SVA	FA
Feat.	103	4	36	36	36	36	144	(144)
Principal Component Analysis								
OA (%)	77.89	72.92	90.00	85.42	69.80	86.56	77.81	89.21
κ (%)	72.34	66.25	87.06	81.24	63.22	82.82	71.08	86.06
AA (%)	85.78	81.55	92.04	89.55	82.48	91.15	86.84	92.04
Independent Component Analysis								
OA (%)	77.89	74.64	91.26	87.94	93.57	87.69	94.47	91.69
κ (%)	72.34	68.22	88.55	84.31	91.63	84.14	92.80	89.13
AA (%)	85.78	77.18	92.36	91.72	95.73	90.92	96.58	94.11

TABLE II
 ROSIS CENTER DATA SET. CLASSIFICATION ACCURACIES OBTAINED BY CLASSIFYING THE HYPERSPECTRAL IMAGE (SPECT.), THE FOUR COMPONENTS EXTRACTED (4 COMP.), EACH SINGLE EAP, AND THE DATA WITH THE PROPOSED APPROACHES: SVA AND FA. THE RESULTS REPORTED ARE THE AVERAGE OF THE ACCURACIES OBTAINED IN 10 TRIALS WITH 50 TRAINING SAMPLES PER CLASS RANDOMLY CHOSEN FOR EACH TRIAL

	Spect.	4 Comp.	EAP _a	EAP _d	EAP _i	EAP _s	SVA	FA
Feat.	102	4	36	36	36	36	144	(144)
Principal Component Analysis								
OA (%)	96.25	96.24	98.40	97.83	97.81	98.48	98.43	98.39
κ (%)	93.59	93.56	97.26	96.27	96.24	97.39	97.31	97.24
AA (%)	92.80	92.65	97.18	96.18	96.23	96.93	97.26	97.21
Independent Component Analysis								
OA (%)	96.25	96.47	98.59	98.18	97.91	97.97	98.69	98.47
κ (%)	93.59	93.94	97.58	96.87	96.42	96.51	97.75	97.38
AA (%)	92.80	92.39	97.50	96.76	96.15	96.18	97.58	97.26

contain more than 99% of the total variance of the data for both the data sets. The components were rescaled to the range [0, 1000] and converted to integer in order to be processed by the attribute filters. Four EAPs were computed by considering four different attributes on the components extracted by PCA and ICA: 1) a , area of the regions ($\lambda_a = [100\ 500\ 1000\ 5000]$); 2) d , length of the diagonal of the box bounding the region ($\lambda_d = [10\ 25\ 50\ 100]$); 3) i , first moment invariant of Hu, moment of inertia [11] ($\lambda_i = [0.2\ 0.3\ 0.4\ 0.5]$); and 4) s , standard deviation of the gray-level values of the pixels in the regions ($\lambda_s = [20\ 30\ 40\ 50]$). The area and the length of the diagonal of the bounding box extract information on the scale of the objects. The moment of inertia and the standard deviation are not dependent on the size dimension, but they are related to the geometry of the objects and the homogeneity of the intensity values of the pixels, respectively. Each EAP is 36-dimensional, i.e., it is composed of four APs with nine levels computed on each component extracted. In the sequel, the notation EAP_{attr} denotes the EAP built with the $attr$ attribute. The classification maps are obtained by analyzing the features extracted by the extended profiles with an SVM classifier with a radial basis function kernel. The model selection in the training phase of the classifier was based on a gradient descent method, which proved to be computationally less demanding than the exhaustive investigation of the parameters on a grid approach, giving comparable results [12].

The thematic accuracies of the obtained maps (which are presented in Tables I and II) were assessed by computing the overall accuracy (OA), the average accuracy (AA), and the Kappa coefficient (κ) on the available reference data. The statistical significance of the classification maps obtained by PCA and

ICA and the same morphological processing was evaluated with the McNemar's test. All the results were statistically significant.

The obtained results are reported in Table I. It is clear, as in most of the cases, by including the features extracted by the EAPs in the analysis resulted in higher accuracies (up to almost 17% of OA) than those obtained by considering only the spectral features. The ICA proved to extract more informative components from the data, leading to better results than those generated by the PCA in all the experiments. When considering the contribution of the single EAPs, the EAPs built with area and the moment of inertia attributes performed the best with the PCA and ICA, respectively. This proves how it can be difficult to select *a priori* the most suitable attribute on the data. In these experiments, considering all the EAPs together, in the SVA architecture, with the ICA gives excellent results in terms of classification accuracies. As far as we know, these accuracies are higher than all those reported in the literature for this data set without postprocessing [1], [13]. In contrast, the SVA approach led to low accuracies for the PCA. This can be due to the high variation in terms of accuracy showed by the single EAPs (more than 20% of OA), which affects the overall performances of this approach. The FA is performing well in average and has a robust behavior since, in all the experiments, the accuracies obtained, when compared to those of the single EAPs, are slightly lower than the best case (less than 2% of OA) and better than all the others. The improved accuracies obtained by the proposed technique are also confirmed by the higher precision shown in the map obtained when considering the ICA and all the EAPs together [see Fig. 4(c)].

Table II reports the thematic accuracies obtained on the Center data set (the correspondent classification maps are not

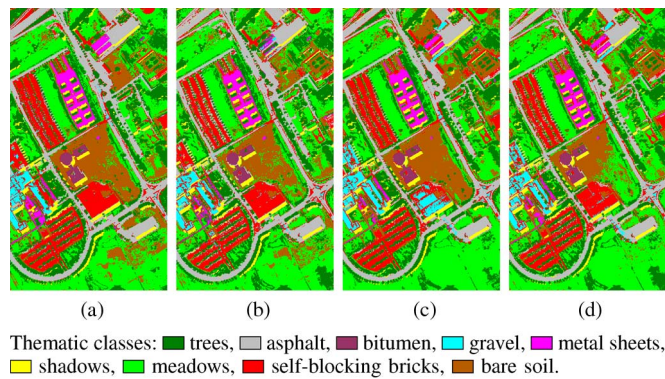


Fig. 4. ROSIS Pavia University data set. Classification maps obtained by (a) PCA with area attribute (EAP_a), (b) PCA with FA, (c) ICA with SVA, and (d) ICA with FA.

reported for space constraints). Similar considerations as for the University data set can be drawn. For this data set also, it is evident the importance of including the spatial information, which led to an increase in terms of accuracy with respect to considering the original hyperspectral data or the components obtained from the dimensionality reduction technique. The best OA obtained by using the EAPs is higher, of about 2%, than those obtained by the original spectral features and the first components. Considering the PCA and ICA transformations, the latter leads to the best results in most of the cases (except for the single components extracted and for the EAP_s). When looking at the performances obtained by considering the spatial features extracted by the EAPs, one can see that the EAP with area attribute outperformed the other single EAPs with PCA, while when considering the ICA, the choice of the standard deviation performed the best among the single EAPs. Moreover, when considering the SVA strategy resulted in the best accuracies with the ICA preprocessing (which is slightly worse than the best EAP). Again, the FA led to results over the average of the accuracies obtained by the single EAPs.

VI. CONCLUSION

In this letter, we have presented a technique based on ICA and APs for the classification of hyperspectral images. In greater details, from the hyperspectral image, some ICs are extracted, and different APs are computed for each one, leading to EAPs. The features obtained by the morphological processing are then classified with an SVM classifier. We proposed two approaches for considering the features extracted by the different EAPs, one based on the concatenation of the EAPs and one based on the fusion of the classification results obtained on the single EAPs.

The experimental analysis was carried out on two well-known hyperspectral images acquired on the city of Pavia (Italy). The results obtained on these data sets proved that the preprocessing of the hyperspectral data carried out with ICA is more suitable than the PCA for modeling the different sources of information present in the scene. Moreover, from the experiments and results, it was evident how important the

spatial features extracted by the EAPs are for classification. The concatenation of the different EAPs gave excellent results in terms of classification accuracies (with respect to other works present in the literature on these data sets). This approach did not perform well only in one case with the PCA, i.e., when significantly different results were obtained from the single EAPs with different attributes (range of difference in the overall accuracies greater than 20%). However, this effect did not occur with the ICA, where the obtained results were more uniform and all statistically significant (according to the McNemar's test). The approach based on the fusion of the classification results with the majority voting strategy proved to have a robust behavior leading to accuracies slightly lower than those of the best case obtained with a single EAPs but better than all the others.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba, University of Pavia, for providing the ROSIS data set.

REFERENCES

- [1] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [2] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Comput. Vis. Image Underst.*, vol. 64, no. 3, pp. 377–389, Nov. 1996.
- [3] 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.
- [4] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis," in *Proc. IEEE IGARSS*, Jul. 25–29, 2005, vol. 1, pp. 176–179.
- [5] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
- [6] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Nov. 2010.
- [7] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using morphological attribute filters and independent component analysis," in *Proc. IEEE WHISPERS*, Reykjavik, Iceland, Jun. 14–16, 2010, pp. 1–4.
- [8] A. Villa, J. Chanussot, C. Jutten, J. A. Benediktsson, and S. Mossaoui, "On the use of ICA for hyperspectral image analysis," in *Proc. IEEE IGARSS*, Jul. 12–17, 2009, vol. 4, pp. IV-97–IV-100.
- [9] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [10] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998.
- [11] M. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.
- [12] A. Villa, M. Fauvel, J. Chanussot, P. Gamba, and J. A. Benediktsson, "Gradient optimization for multiple kernel's parameters in support vector machines classification," in *Proc. IEEE IGARSS*, Jul. 7–11, 2008, vol. 4, pp. IV-224–IV-227.
- [13] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral & spatial classification of hyperspectral imagery based on partitionial clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.