Classification of hyperspectral images by using morphological attribute filters and independent component analysis

Mauro Dalla Mura, Alberto Villa, Jon Atli Benediktsson, Jocelyn Chanussot, Lorenzo Bruzzone

CLASSIFICATION OF HYPERSPECTRAL IMAGES BY USING MORPHOLOGICAL ATTRIBUTE FILTERS AND INDEPENDENT COMPONENT ANALYSIS

Mauro Dalla Mura[a,b], *Alberto Villa*[c,b], *Jon Atli Benediktsson*[b], *Jocelyn Chanussot*[c] *and Lorenzo Bruzzone*[a]

[a] Department of Information Engineering and Computer Science, University of Trento
Via Sommarive, 14 I-38123, Povo, Trento, Italy,
E-mail: dallamura@disi.unitn.it, lorenzo.bruzzone@ing.unitn.it
[b] Faculty of Electrical and Computer Engineering, University of Iceland,
Hjardarhaga 2-6, 101 Reykjavik, Iceland,
E-mail: benedikt@hi.is
[c] GIPSA-lab, Signal & Image Dept., Grenoble Institute of Technology - INP, France
961 rue de la Houille Blanche, 38402 Grenoble Cedex, France,
E-mail: alberto.villa@hyperinet.eu, jocelyn.chanussot@gipsa-lab.grenoble-inp.fr

ABSTRACT

In this paper, a technique based on Independent Component Analysis (ICA) and morphological attribute filters is presented for the classification of high geometrical resolution hyperspectral images. The ICA is computed instead of the conventional principal component analysis (PCA) in order to better model the information in the hyperspectral image. The spatial characteristics of the objects in the scene are modeled by different multi-level attribute filters. Moreover, a method for increasing the robustness of the analysis based on a decision fusion strategy is proposed. A hyperspectral high resolution image acquired over the city of Pavia (Italy) was considered in the experiments.

***Index Terms*—** Mathematical morphology, attribute filters, independent component analysis, decision fusion, remote sensing.

1. INTRODUCTION

The exploitation of the spatial information is very important for the classification of high geometrical resolution hyperspectral images [1]. In fact, when the analysis is carried out by only considering the spectral response of the single pixels without taking into account the spatial context, the classification of the image usually leads to results which are spatially inaccurate (e.g., showing a high degree of over-segmentation).

The contextual relations between spatially neighboring pixels can be exploited in several ways. A widely used technique for including the information of the spatial domain in the analysis is the family of transformations belonging to the mathematical morphology framework. In particular, 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, etc.). In [3], morphological profiles (MPs), a sequence of multi-scale connected operators, were applied to high resolution hyperspectral images by reducing the high dimensionality of the data by a Principal Component Analysis (PCA), and computing the profiles on the first principal components extracted. Due to the limitations of PCA when extracting the informative sources from the high dimensional data, in [4], the authors proposed to perform an Independent Component Analysis (ICA) before the computation of the MPs.

The characterization of the spatial information obtained by the application of a MP is particularly suitable for representing the multi-scale variability of the structures in the image but it might not be sufficient to model other geometrical features. To avoid this limitation, in [5] the use of morphological attribute filters instead of the conventional operators based on the geodesic reconstruction was proposed. The application of attribute filters in a multi-level 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, analogously to [3], the APs where applied to the first principal components extracted from a hyperspectral image, generating an Extended Attribute Profile (EAP).

In this paper, a technique based on extended attribute profiles and independent component analysis for the classification of hyperspectral high resolution images is proposed. Moreover, an architecture for dealing with the increase of dimensionality generated by the EAPs based on a decision fusion strategy is presented.

The paper is organized as follows. In Section 2 the ICA is recalled, in Section 3 the concepts of morphological attribute filters and extended attribute profiles are presented. The architecture of the proposed fusion technique is presented in Section 4 and the experimental results are illustrated in Section 5. Finally, conclusions are drawn in Section 6.

2. INDEPENDENT COMPONENT ANALYSIS

When dealing with high dimensional data, feature reduction is one of the fundamental tasks, which consists in designing a statistical model of the observed signal, trying to minimize the loss of information. In case of data classification, this pre-processing step is useful both to avoid the Hughes'phenomenon and to reduce the computational time. PCA is often used for such a task, due to its simplicity and ease of use. Nevertheless, PCA is based on the analysis of covariance matrix and second order statistics, thus some important information can be neglected, especially when few components are

978-1-4244-8907-7/10/$26.00 ©2010 IEEE

retained. In this paper, we propose the use of ICA as an effective alternative to PCA, for the purpose of feature reduction. ICA consists of finding a linear decomposition of the observed data into statistically independent components. Given an observation model:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where **x** is the vector of the observed signals, **A** is a matrix of scalars corresponding to the mixing coefficient and **s** the vector of the source signals, ICA finds a separating matrix **W** such that

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}, \tag{2}$$

where **y** is a vector of independent components (ICs).
Independence is a much stronger assumption than uncorrelation, which is obtained with common decorrelation methods such as PCA or Factor Analysis (FA). In order to ensure independence, each component should not provide any information about higher (than second) order statistics of the other components. Because of this reason, ICs are expected to contain more information than Principal Components, and to be more suitable for classification purposes, also in the case of a small number of ICs retained. There is not an unique definition for independence, and there are several methods for estimating ICA. In this paper we have used the algorithm JADE, due to good results shown when used for feature reduction of hyperspectral remote sensing data [7]. If the ICA is used as a transformation for dimensionality reduction, the feature extracted are the components that are mutually maximally independent (according to the definition of independence used). Due to the space constraints we refer the reader interested in more details about the general framework of ICA to [8].

3. ATTRIBUTE FILTERS AND EXTENDED ATTRIBUTE PROFILES

Attribute filters are a class of connected operators [2]. They process a single band image by either entirely preserving or entirely suppressing the connected components (i.e., regions of iso-intensity spatially connected pixels) that compose the image. The suppression or preservation of the flat regions is chosen according to the fulfillment of a criterion evaluated on each component. The criterion compares the value of an arbitrary attribute (e.g., area, volume, standard deviation, etc.) measured on the component against the value of a global threshold. If the criterion is verified then the connected components are kept unaffected otherwise they are completely removed.

Example of attribute filters are attribute opening and closing. An attribute opening processes the image by removing connected components brighter than their surrounding ones (while a closing processes the darker ones), which do not fulfill the criterion of the transformation. The criterion has to be increasing (if it is verified for a connected component then it will be also verified by all the regions that contain it). If the criterion does not fulfill this property, then the transformation is not increasing anymore and becomes a thinning or thickening.

Attribute openings for binary images are obtained by computing a trivial opening, Γ^T on the output of a connected opening, Γ_X, applied to all the connected components of an image X. Given a point x in the image domain and a connected component C, the connected opening is computed as

$$\Gamma_x(X) = \begin{cases} C & if \quad x \in C; \\ \emptyset & otherwise. \end{cases} \tag{3}$$

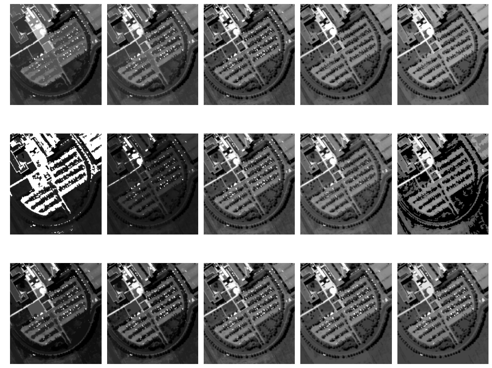


Fig. 1. Examples of APs computed with different attributes (from row one to three): area, moment of inertia and standard deviation. They model the objects in the image according to their scale, geometry and spectral homogeneity respectively.

Being a, an attribute measured on a generic connected component C of the image, the trivial opening keeps the regions for which the criterion T (e.g., $T(C) = a(C) \leq \lambda$) holds. This can be expressed as:

$$\Gamma_T(C) = \begin{cases} C & if \quad T = true; \\ \emptyset & otherwise. \end{cases} \tag{4}$$

Attribute opening is then given by:

$$\Gamma^T(X) = \bigcup_{x \in X} \Gamma_T(\Gamma_x(X)). \tag{5}$$

By duality, attribute closing is analogously computed by considering the background regions instead of the foreground ones. When the criterion is increasing, the extension of the operators from binary to gray-scale is straighforward by exploiting the threshold superposition principle [9]. Conversely, when the criterion is not increasing, the extension to grayscale of the binary concepts is not direct and a proper filtering rule has to be selected in order to implement the filtering [10].

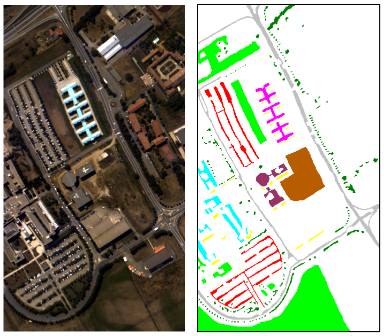
We remind that attribute filters are appealing also from the computational complexity side, since they can be efficiently computed by using the Max-Tree algorithm [10].

Attribute profiles are defined as a concatenation of L attribute openings/thinnings and L closings/thickenings with a progressively relaxed criterion $U = \{T_i(X) = a(X) \leq \lambda_i\}$ with $\lambda_i \geq \lambda_j$, $i \geq j$ and $i, j = 1, \ldots, L$ [5]. An example of APs computed with different attributes are shown in Fig. 1 where it is possible to notice as different structures are enhanced according to the type of attribute considered.

Extended attribute profiles (EAPs) were presented as an extension of the concept of AP to the hyperspectral domain [6]. EAPs are based on a feature reduction stage and are obtained by concatenating the APs computed on each of the feature extracted from the image.

4. DECISION FUSION TECHNIQUE

The *a piori* choice of the most suitable attribute for extracting the information on the geospatial objects is certainly a complex task. However, by considering in the analysis all the features extracted from several APs can lead to a consistent increase of the dimensionality, thus, leading to the Hughes phenomenon. For this reason, we



(a) (b)

Fig. 2. Pavia data set: (a) True color representation, (b) Test set.

propose an architecture for fusing the information extracted from different profiles in order to increase the robustness of the analysis but, at the same time, by keeping low the dimensionality of the data. The general scheme of the proposed technique is presented in the following. After reducing the dimension of the hyperspectral image with PCA/ICA, EAPs with different attributes are computed on the extracted components. Subsequently, each EAP is processed by a support vector machine (SVM) classifier and the results of the classification are fused according to a decision criterion in order to generate the final classification map. The multiclass strategy selected is One Against One (OAO): a binary classifiers is applied for each pair of classes, producing a vote for the class to which the sample would be assigned. Finally, each pixel is assigned to the class that gets the highest number of votes. In the fusion approach, we try to exploit the information provided by all the classifiers. A simple, yet effective criterion is to sum the votes of the classifiers applied to the four EAPs, assigning each pixel to a class, according to a majority voting scheme. Obviously, other decision criteria can be applied. Fusing the results of several independent classifiers leads to increase the robustness of the results especially if the different EAPs generate complementary errors.

5. EXPERIMENTAL ANALYSIS

The experimental analysis was carried out on a high resolution hyperspectral image acquired over the Pavia University campus by the ROSIS-03 (Reflective Optics Systems Imaging Spectrometer) hyperspectral sensor. The data are composed of 103 spectral bands (ranging from 430 to 860 nm with a step of 4 nm) of 610×340 pixels with a 1.3 m geometrical resolution. Nine thematic land cover classes were identified in the scene: Trees, Asphalt, Bitumen, Gravel, Metal sheets, Shadows, Self-blocking Bricks, Meadows, and Bare soil. A total of 3921 and 42776 pixels were available as training and test sets respectively. The true color representation of the image and the test set taken as reference are shown in Fig. 2.

From the hyperspectral image the first four components of the PCA were considered for the analysis in order to explain more than 99% of the total variance of the data. Four independent components were also extracted by the ICA from the original data. Subsequently, on the features extracted by both the PCA and ICA, four EAPs were computed with different attributes: i) a, area of the regions; ii) d, length of the diagonal of the box bounding the region; iii) i, first moment invariant of Hu (or moment of inertia) [11]; and iv) s, standard deviation of the gray-level values of the pixels in the regions. The area and the length of the diagonal of the bounding box are increasing attributes that process the objects according to their size. The moment of inertia attribute is a purely geometric descriptor that by being sensitive to the elongation of the regions, can perform an analysis based on the geometry of the objects regardless their scale. Finally, the standard deviation attribute measures the homogeneity of the intensity values of the pixels belonging to each region in the image and it gives information related to the spectral contrast of the pixels. Four reference values, used as global thresholds λs, were considered for building each of the four EAPs, leading to 36-dimensional profiles (composed by four APs of 9 levels computed on each component extracted by the original data). The thresholding values (λ) used in the filterings were arbitrarily selected and they are listed in the following:

1. EAP$_a$: $\lambda_a =$ [100 500 1000 5000];
2. EAP$_d$: $\lambda_d =$ [10 25 50 100];
3. EAP$_i$: $\lambda_i =$ [0.2 0.3 0.4 0.5];
4. EAP$_s$: $\lambda_s =$ [20 30 40 50].

A SVM classifier with a gradient descent based approach for the selection of the parameters was applied to the data. This approach provides results which are comparable to the cross-validation method, and is much more efficient from a computational point of view [12]. The classification results were quantitatively evaluated by measuring the Overall Accuracy (OA), the Average Accuracy (AA) and the Kappa coefficient (K) on the reference data.

Table 1 presents the accuracies obtained by the PCA and ICA as feature extraction techniques. It is possible to notice, when considering the PCA, how the EAP with area attribute performed the best, with more 3% of OA more than the others EAPs. By considering the proposed method of fusion, accuracies slightly smaller than the EAP area and better than all the other EAPs were obtained. Conversely, when all the EAPs are concatenated together a lower accuracy is obtained than the average of the single EAPs. When comparing the results obtained with the PCA with those of ICA, it is clear how the latter ones outperform the former set. Conversely to the PCA, by considering the independent components, the EAP with moment of inertia attribute scored the best among all the EAPs. This proves how, in this case, it is difficult to select *a priori* the most suitable attribute on the data. The use of all the EAPs together performed the best among all the single EAPs even for a high dimensionality of the data. The fusion of the results given by the single EAPs involved accuracies that are slightly worse than the best EAPs but higher than all the other ones. This confirmed what was assessed before.

The increase in accuracy previously quantified for the use of the ICA instead of the PCA can be visually noticed in the classification maps shown in Fig. 3. Moreover, one can see that the most precise map is obtained when considering the ICA and all the EAPs together (see Fig. 3.c).

6. CONCLUSION

In this paper we have presented a technique based on independent component analysis and morphological attribute filters for the classification of high resolution hyperspectral images. In greater details, from the hyperspectral image some independent components are extracted, and different attribute profiles are computed for each one, leading to extended attribute profiles. The features obtained by the morphological processing are then classified with a SVM classifier. Moreover, we proposed a robust technique for fusing the results obtained by processing the image with different attributes.

Table 1. Classification accuracies obtained according to the described scheme. The first 4 columns represent the results of SVM applied to the single EAPs. *All* means that the SVM is applied to the data obtained considering the outputs of the four filters, together. *Fusion* is obtained with the sum of the votes of the first four classifiers.

	Area	Diagonal	Inertia	Std	All	Fusion
Feat.	36	36	36	36	144	(144)
Principal Component Analysis						
OA (%)	90.00	85.42	69.80	86.56	77.81	89.21
κ (%)	87.06	81.24	63.22	82.82	71.08	86.06
AA (%)	92.04	89.55	82.48	91.15	86.84	92.04
Independent Component Analysis						
OA (%)	91.26	87.94	93.57	87.69	94.47	91.69
κ (%)	88.55	84.31	91.63	84.14	92.80	89.13
AA (%)	92.36	91.72	95.73	90.92	96.58	94.11

The experimental results obtained on a high resolution hyperspectral image acquired on the city of Pavia, Italy, proved how the ICA is more suitable to model the different informative sources present in the scene outperforming the results of the PCA of about 4% of overall accuracy. Moreover, the choice of the most suitable attribute has been confirmed to be a non trivial task. For example, when considering the PCA the EAP with area attribute performed the best, whereas for the ICA the most accurate results were achieved by the EAP with moment of inertia. The proposed method of combining the results of the single EAPs proved to be a robust approach when no *a priori* information is available on the scene. This technique has obtained accuracies slightly lower than those of the best case but better than all the others.

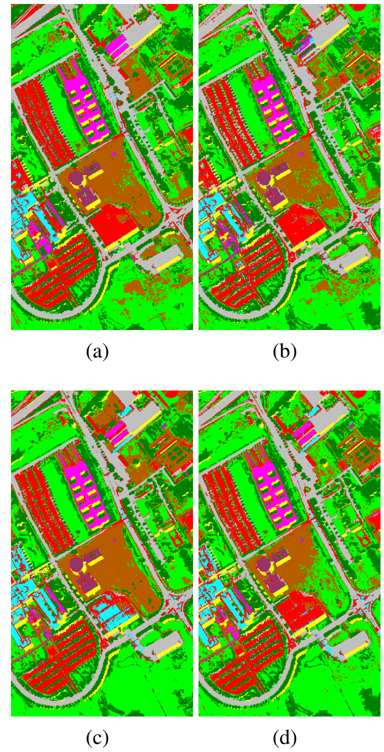


Fig. 3. Classification maps obtained by: (a) PCA with area attribute, (b) PCA fusion, (c) ICA with all the attributes, and (d) ICA fusion.

7. ACKNOWLEDGMENTS

The authors would like to thank Prof. Paolo Gamba, University of Pavia, for providing the ROSIS data set. This work has been supported in part by the European Community's Marie Curie Research Training Networks Programme under contract MRTN-CT-2006-035927, Hyperspectral Imaging Network (HYPER-I-NET) and by the Research Fund of the University of Iceland and the University of Trento.

8. 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, 2008.

[2] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Comput. Vis. Image Underst.*, vol. 64, no. 3, pp. 377–389, 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, 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 '05*, July 25–29, 2005, vol. 1, p. 4pp.

[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.*, in press.

[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.*, in press.

[7] 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 '09*, 2009.

[8] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.

[9] P. Maragos and R.D. Ziff, "Threshold superposition in morphological image analysis systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 498–504, 1990.

[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, 1998.

[11] M. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, 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 2008*, July 7–11, 2008, vol. 4, pp. IV–224–IV–227.