



**HAL**  
open science

## Independent component discriminant analysis for hyperspectral image classification

Alberto Villa, Jon Atli Benediktsson, Jocelyn Chanussot, Christian Jutten

► **To cite this version:**

Alberto Villa, Jon Atli Benediktsson, Jocelyn Chanussot, Christian Jutten. Independent component discriminant analysis for hyperspectral image classification. WHISPERS 2010 - 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Jun 2010, Reykjavik, Iceland. conference proceedings. hal-00578911

**HAL Id: hal-00578911**

**<https://hal.science/hal-00578911>**

Submitted on 22 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# INDEPENDENT COMPONENT DISCRIMINANT ANALYSIS FOR HYPERSPECTRAL IMAGE CLASSIFICATION

A. Villa<sup>\*,◊</sup>, J. A. Benediktsson<sup>◊</sup>, J. Chanussot<sup>\*</sup>, C. Jutten<sup>\*</sup>

<sup>\*</sup>GIPSA-lab, Signal & Image Dept., Grenoble Institute of Technology - INPG, France.

<sup>◊</sup>Faculty of Electrical and Computer Engineering, University of Iceland, Iceland.

## ABSTRACT

In this paper, the use of Independent Component Discriminant Analysis (ICDA) for remote sensing classification is proposed. ICDA is a non-parametric method for discriminant analysis based on the application of a Bayesian classification rule on a signal composed by independent components. The method is based on the use of Independent Component Analysis (ICA) to choose a transform matrix so that the transformed components are as independent as possible. Then, a non parametric estimation of the density function is computed for each independent component. Finally, the Bayes rule is applied for classification assignment. The obtained results are compared with one of the most used classifier of hyperspectral images (Support Vector Machine) and show the comparative effectiveness of the proposed method.

**Index Terms**— Hyperspectral data, Independent Component Analysis, Bayesian Classification.

## 1. INTRODUCTION

Hyperspectral images are composed of hundreds of bands with a very high spectral resolution, from the visible to the infra-red region. The wide spectral range, coupled with an always increasing spatial resolution, allows a better characterization of materials and gives the ability to pinpoint ground objects laying on the observed surface, making hyperspectral imagery suitable for land cover classification. On the other hand, the huge amount of high dimensional data increases the computational load and can also degrade the results of the classification process, due to the curse of dimensionality [1].

Because of this, parametric classifiers such as Maximum Likelihood [2] or Bayesian classifier [2, 3], based on the assumption that each class can be represented by a multivariate normal model depending on the mean and covariance matrix of the data, lead in general to poor results. More advanced classifiers, like Neural Networks (NN) [4] and Support Vector Machines (SVM) [5], have been recently proposed for hyperspectral data classification, due to their good ability to handle high dimensional data. Neural Networks suffer basically from two main limitations. First, the number and the size of hidden layers need to be set, and this is not always being a straightforward task. Secondly, a very large number of iterations is sometimes needed to find a solution. Support Vector Machines have even shown better performances than NN, providing good results also in the case of very limited training sets [5]. The main limitations of SVM are the training time, quadratically depending on the number of training samples, and the need to find the optimal parameters for the kernel,

usually done using a cross-validation approach.

In this work we present a non-parametric method for discriminant analysis based on the application of a Bayesian classification rule on a signal composed by independent components [6]. The main characteristics of the method are the use of Independent Component Analysis (ICA) to retrieve independent components from the original data; non parametric estimation of the density function for each independent component; application of a Bayes rule for classification assignment. The obtained results are compared with one of the most used classifier of hyperspectral images (SVM) and show the comparative effectiveness of the proposed method.

## 2. INDEPENDENT COMPONENT DISCRIMINANT ANALYSIS

The proposed method is a generalization of the quadratic discriminant analysis, where the ability of ICA to retrieve independent components is exploited to estimate the marginal densities of the transformed components. These densities, which are hard to estimate when dealing with high dimensional data, can be computed in a much simpler way in an independent space.

The risk incurred when performing a classification of a measured vector  $\mathbf{x}$  into one of  $K$  possible classes is given by:

$$R(\hat{k}|\mathbf{x}) = \frac{\sum_{k=1}^K L(k, \hat{k}) f_k(\mathbf{x}) \pi_k}{\sum_{k=1}^K f_k(\mathbf{x}) \pi_k} \quad (1)$$

where  $\pi_k$  is the *a priori* probability that  $\mathbf{x}$  belongs to the class  $k$ ,  $f_k$  is the class-conditional *a priori* density of  $k$ , and  $L$  is the cost or loss incurred when assigning the sample  $\mathbf{x}$ , belonging to the class  $k$ , to the class  $\hat{k}$ . In the case of hard classification, this cost is equal to zero if  $k = \hat{k}$ , and is equal to one otherwise.

By choosing  $\hat{k}$  such that the numerator of (1) is minimized leads to the so-called Bayes decision rule. In the case of hard classification, the Bayes rule reduces to the following rule: assign  $\mathbf{x}$  to the class  $\hat{k}$  such that

$$\hat{k} = d(\mathbf{x}) = \underset{k}{\operatorname{argmax}} \{ f_k(\mathbf{x}) \pi_k \} \quad k = 1, \dots, K. \quad (2)$$

The most often applied classification rules are derived by assuming that the class-conditional densities are p-variate normal with mean vectors  $\mu_k$  and covariance matrices  $\Sigma_k$  assumed to be non singular. These two parameters are estimated from the training samples according to the following equations:

$$\hat{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_{ik} \quad (3)$$

This work was supported by the Hyper-I-Net project, funded by the European Union

$$\hat{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (x_{ik} - \hat{\mu}_k)(x_{ik} - \hat{\mu}_k)^T \quad (4)$$

where  $\{x_{ik}, i = 1, \dots, N_k\}$  are the training samples of the class  $k$ . This approach works well when the class conditional densities are approximately normal and good estimates can be obtained from the training samples. The approach is highly affected by substantial divergence from normal density and by a limited size of training set [7], as it is often the case for hyperspectral remote sensing data.

The parametric approach to discriminant analysis has been extended to the case where nothing is known about the densities  $f_k$  but only some assumptions about their general behavior [8]. Here, the idea is to apply non-parametric density estimators to the training samples and then substitute the obtained estimates in the Bayes decision rule (2). The common used procedure is to use a multivariate kernel density estimator of the form:

$$\hat{f}_k(\mathbf{x}) = \mathcal{K}\{\mathbf{x} - \mathbf{x}_k; \mathbf{H}_k\}, \quad (5)$$

where  $\mathcal{K}$  denotes a multivariate kernel function and  $\mathbf{H}_k$  is a  $p$ -dimensional vector of appropriate bandwidths. Multidimensional density estimation is highly affected by the high dimensionality of the data, and practically is not applicable when the dimension of the data is comparable with the sample size of the classes, such as in the case of hyperspectral data. In these cases, the kernel  $\mathcal{K}$  is approximated as the product of univariate Gaussian kernel functions, leading to estimates of the form [9]:

$$\hat{f}_k(\mathbf{x}) = (2\pi)^{-p/2} \mathcal{H} N_k^{-1} \sum_{l=1}^{N_k} \prod_{j=1}^p \exp\left\{-\frac{(x_j - x_{lkj})^2}{2h_{kj}^2}\right\}, \quad (6)$$

where

$$\mathcal{H} = h_{k1} \cdot h_{k1} \cdots h_{kp}. \quad (7)$$

The main drawback of this approach is that some important information for the classification process is not retrieved. When dealing with high-dimensional data where very important information for the classification process can be hidden in relatively low density regions, the tails estimation becomes crucial in order not to degrade the final results. In this case, a Gaussian kernel product estimator can be inappropriate, due to the short tailed normal density.

In [6], Amato *et al.* proposed an interesting approach to circumvent the problems of non-parametric multivariate kernel density estimators. They used ICA to enforce independence to the components of the analyzed data. In the new transformed space the components are as independent as possible. This way, the joint class-conditional density, tricky to estimate in the original data space, can be computed in the independent space in a much simpler way, as the product of marginal class-conditional densities. The results obtained are finally substituted in the Bayes rule for the class assignment. The basic steps of the proposed approach are:

1. Center the data on the  $k$ -class, for each class  $k = 1, \dots, K$ , and use the ICA to derive the optimal transform  $\hat{\mathbf{A}}_k$  according to the training samples of the class.
2. Project the data using the computed transform and use an adaptive univariate kernel density estimator to estimate the density of each component.
3. For a new observation  $\mathbf{x}$ , the joint density of  $\mathbf{Y} = \hat{\mathbf{A}}_k \mathbf{x}$  is first computed for each class as the product of the estimated marginal densities, since the components are independent. The density of  $\mathbf{x}$  can be then derived from that of  $\mathbf{Y}$  with a simple change of variable. The results are then substituted into the Bayes rule to obtain the final assignment.

Later on, we will refer to the above approach as Independent Component Discriminant Analysis (ICDA).

### 3. INDEPENDENT COMPONENT ANALYSIS

ICA consists in finding a linear decomposition of observed data into statistically independent components. Given an observation model

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (8)$$

where  $\mathbf{x}$  is the vector of the observed signals,  $\mathbf{A}$  is a scalar matrix of the mixing coefficients and  $\mathbf{s}$  is the vector of the source signals. ICA finds a separating matrix  $\mathbf{W}$  such that

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}, \quad (9)$$

where  $\mathbf{y}$  is a vector of independent components. ICA looks for a linear representation that maximizes a non-Gaussianity measure, or minimizes an objective function. A commonly used objective function in ICA algorithms is the mutual information of vector  $\mathbf{y}$ :

$$I(\mathbf{y}, \mathbf{W}) = \sum_i H(y_i) - H(\mathbf{y}) \quad (10)$$

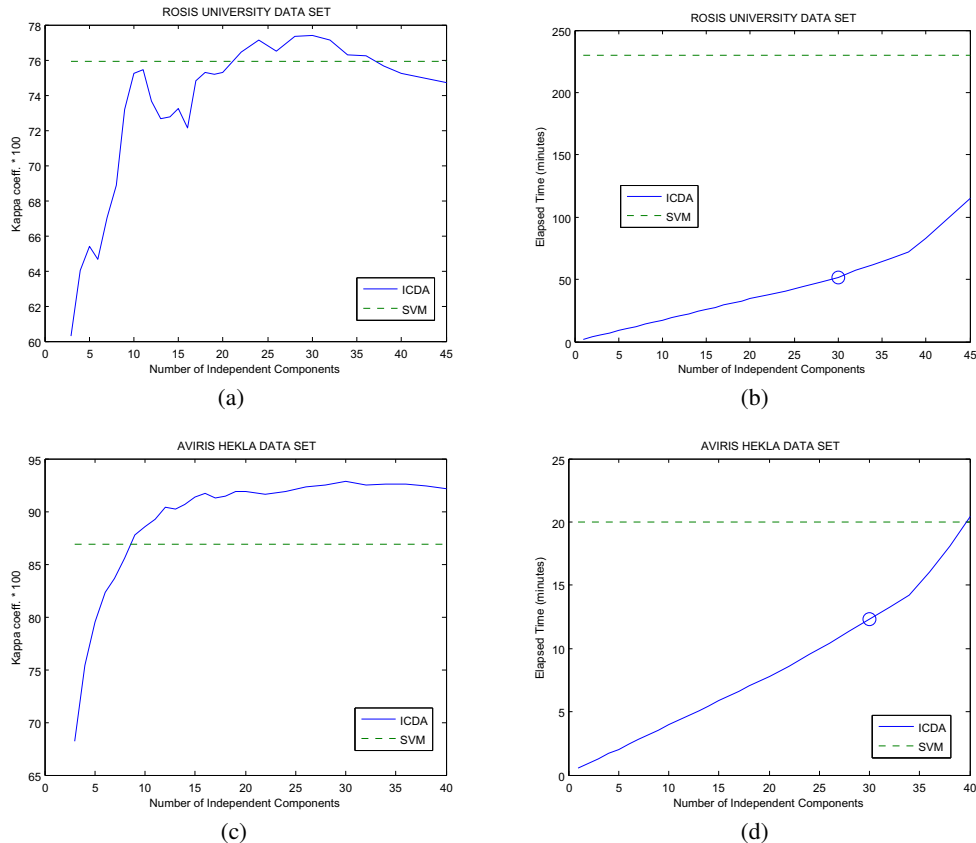
where  $H(y_i)$  and  $H(\mathbf{y})$  are the joint and marginal entropy of random variable  $y_i$  and random vector  $\mathbf{y}$ , respectively. Many algorithms have been developed in the last decades to estimate independent components. In our experiments, we will use JADE, because of the good performances shown when dealing with hyperspectral data, both for accuracy and computational burden. More details about the general framework of ICA can be found in [10, 11, 12].

### 4. EXPERIMENTAL RESULTS

Three hyperspectral data sets were considered in this work. The first one is an airborne data from the ROSIS-03 with 115 spectral bands in the spectral range from 0.43 to 0.86  $\mu\text{m}$ , acquired over the University of Pavia, Italy. The spatial resolution is 1.3m per pixel. The original data set is 610 by 340 pixels. Twelve data channels were removed due to noise, and the remaining 103 spectral dimensions were processed. Training data with nine classes are available. The training set is composed by about 10% of all the labeled samples. The second data set is a small segment of an AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) data set over the agricultural area of Indiana. The data set is 145 by 145 set of pixels. It is composed of 220 spectral channels (spaced at about 10 nm) acquired in the 0.4-2.5  $\mu\text{m}$  region [13]. All 220 bands were processed, without removing noisy channels. Sixteen reference classes were considered. Different training sets were randomly constructed from the reference data with a total of respectively 320 pixels (20 samples per class). Due to the very small size of the training set, each experiment was repeated five times to increase the statistical significance of the test and the average results are reported. Finally, the third study site is the region surrounding the central-volcano Hekla in Iceland, one of the most active volcanoes in the country. Since 1970, Hekla has erupted quite

**Table 1.** Training-test samples for the three considered data sets.

|                    | Training samples | Testing samples |
|--------------------|------------------|-----------------|
| ROSIIS University  | 3921             | 42776           |
| AVIRIS Indian Pine | 320              | 10366           |
| AVIRIS Hekla       | 600              | 9636            |



**Fig. 1.** Kappa coefficient obtained with SVM (full feature space) and ICDA and classification processing time, with respect to different number of independent components considered, for: (a) and (b) ROSIS, (c) and (d) AVIRIS Hekla data sets. The circle represents the processing time corresponding to the best classification result obtained with ICDA in terms of accuracy.

regularly every 10 years, in 1970, 1980-81, 1991 and in 2000. The volcano is located on the South-Western margin of the Eastern volcanic zone in South Iceland. Heklas products are mainly andesitic and basaltic lavas and tephra. AVIRIS data that were collected on a cloud-free day, June 17 1991, were used for the classification. The AVIRIS sensor operates in the wavelengths range from  $0.4 \mu\text{m}$  to  $2.4 \mu\text{m}$ . As on the previous case, the sensor system has 224 data channels, utilizing four spectrometers, whereas each spectral band is approximately 10nm in width. During the image acquisition, spectrometer 4 was not working properly. This particular spectrometer operates in the wavelength range from  $1.84 \mu\text{m}$  to  $2.4 \mu\text{m}$  (64 bands). These 64 bands were deleted from the imagery along with the first channels for all the other spectrometers, and the remaining 157 data channels were left.

The results obtained in our experiments with ICDA have been compared with those obtained by SVM in the full feature space, with a Gaussian kernel and 10 fold cross-validation selection of the kernel's parameter [14]. The chosen multi class strategy is One versus One. The performances have been computed in terms of Overall Accuracy, that is percentage of samples correctly classified, Average class Accuracy, which represents the average of the classification accuracies for the individual classes, and Kappa coefficient, a parameter that estimates the correct percentage classification without the amount due to chance alone. When applying ICDA, the number of components considered to compute the density estimation has an

influence both on the final classification accuracy and on the computational burden. Figure 1 shows the variation of the Kappa coefficient ((a) and (c)) and processing time ((b) and (d)) with respect to the number of independent components retained, for two of the considered data sets, ROSIS and AVIRIS Hekla. Though the number of components has a large influence on the final results, it can be seen that there is a wide region where the proposed method outperforms the SVM. The computational burden of SVM is generally much bigger than the proposed method, and the classification time in the best obtained result in terms of accuracy is, for the AVIRIS Hekla data set, about one half of time for SVM, and for the ROSIS data set, less than one fourth. The reason is that the processing time of the SVM increases quadratically with the size of the training set, and it is longer where a large number of training samples is available (as in the case of ROSIS data set), while ICDA is more affected by the size of the testing samples. The best results are presented in Table 2 and show the effectiveness of the method. In all the three considered data sets the Overall Accuracy and Kappa coefficient are significantly better than in the case of SVM classification.

## 5. CONCLUSIONS

In this work, the use of Independent Component Discriminant Analysis for hyperspectral image classification has been investigated. The proposed approach is based on the application of ICA to the

**Table 2.** Comparison of classification accuracies obtained with the proposed method ICDA (best case) and SVM in the full feature space, in the three analysed data set.

| Approach | RODIS Data Set |               | AVIRIS Indian Pine |               | AVIRIS Hekla |               |
|----------|----------------|---------------|--------------------|---------------|--------------|---------------|
|          | SVM            | ICDA          | SVM                | ICDA          | SVM          | ICDA          |
| OA       | 81.01%         | <b>82.14%</b> | 56.00%             | <b>66.58%</b> | 88.96%       | <b>93.91%</b> |
| $\kappa$ | 75.86%         | <b>77.38%</b> | 50.99%             | <b>62.50%</b> | 86.91%       | <b>92.85%</b> |
| AA       | <b>88.25%</b>  | 87.48%        | 70.10%             | <b>77.17%</b> | 89.44%       | <b>92.39%</b> |
| Class 1  | 84.93%         | 76.40%        | 84.44%             | 83.33%        | 88.36%       | 96.92%        |
| Class 2  | 70.79%         | 77.74%        | 36.79%             | 55.97%        | 87.25%       | 95.14%        |
| Class 3  | 67.16%         | 77.42%        | 40.67%             | 51.61%        | 88.24%       | 94.19%        |
| Class 4  | 97.77%         | 98.07%        | 72.31%             | 78.80%        | 84.94%       | 96.54%        |
| Class 5  | 99.46%         | 100%          | 80.40%             | 84.87%        | 93.33%       | 86.11%        |
| Class 6  | 92.83%         | 88.86%        | 78.93%             | 92.53%        | 94.24%       | 98.56%        |
| Class 7  | 90.42%         | 91.35%        | 95.38%             | 96.15%        | 87.54%       | 96.06%        |
| Class 8  | 92.78%         | 82.02%        | 76.11%             | 90.43%        | 91.69%       | 79.54%        |
| Class 9  | 98.11%         | 95.35%        | 100%               | 100%          | 85.88%       | 87.29%        |
| Class 10 | -              | -             | 53.80%             | 57.25%        | 74.20%       | 80.20%        |
| Class 11 | -              | -             | 39.73%             | 49.78%        | 100%         | 100%          |
| Class 12 | -              | -             | 49.12%             | 60.07%        | 97.59%       | 98.19%        |
| Class 13 | -              | -             | 91.42%             | 99.15%        | -            | -             |
| Class 14 | -              | -             | 81.31%             | 92.30%        | -            | -             |
| Class 15 | -              | -             | 47.05%             | 51.89%        | -            | -             |
| Class 16 | -              | -             | 94.11%             | 90.53%        | -            | -             |

data in order to retrieve independent components, the use of a kernel density estimate to obtain reliable estimation of class dependent densities and the substitution on the Bayes rule for the final assignment. Experiments have been carried out on three different real data sets. These results showed the effectiveness of the method, which provided better results than one of the state-of-the-art hyperspectral classifier, the SVM. Though the classification accuracy is influenced by the number of components retained after applying ICA, this choice is not critical, since there is a large range around the best option for which the proposed method have similar results and outperforms SVM. Moreover, the computational burden of the proposed method is much smaller as compared to SVM. Further developments of this work include a comprehensive research of the influence of the ICA algorithm used to enforce independence. Other techniques, such as FastICA or Infomax, should be investigated.

### Acknowledgments

The authors would like to acknowledge Prof. Anestis Antoniadis, IMAG Université J. Fourier, for providing the ICDA code and Prof. Paolo Gamba, University of Pavia, for providing the ROSIS data set. This work has been supported by the European Community's Marie Curie Research Training Networks Programme under contract MRTN-CT-2006-035927, HYPER-I-NET.

### 6. REFERENCES

- [1] C. Lee and D. A. Landgrebe, "Analyzing high dimensional multispectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 31, pp. 792-800, Jul. 1993.
- [2] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, second ed. John Wiley & Sons, 2001.
- [3] J. Besag, "Toward Bayesian image analysis," *J. Appl. Stat.*, vol. 16, no. 3, pp. 395-407, 1989.
- [4] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [5] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, 1351-1362, Jun. 2005.
- [6] U. Amato, A. Antoniadis and G. Gregoire, "Independent Component Discriminant Analysis", *Intern. Math. Journal*, vol. 3, no. 7, pp. 735-753, 2003.
- [7] J. Friedman, Exploratory projection pursuit, *J. Amer. Statist. Assoc.*, vol. 82, pp. 249-266, 1987.
- [8] E. Fix and J.L. Hodges, "Discriminatory Analysis - nonparametric discrimination: Consistency properties", *Int. Stat. Rev.*, vol. 57, pp. 238-247, 1989.
- [9] D.W. Scott, *Multivariate density estimation: theory, practice and visualization*, Wiley, New York, 1992.
- [10] P. Comon, "Independent component analysis, a new concept?," *Sign. Proc.*, vol. 36, pp. 287-314, 1994.
- [11] T.W. Lee, M. Girolami, A.J. Bell and T.J. Sejnowski, "An unifying information-theoretic framework for Independent Component Analysis," *Comp. Math. with Appl.*, vol. 31, no. 11, pp. 1-21, Mar. 2000.
- [12] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [13] R.O. Green, M.L. Eastwood, C.M. Sarture, T.G. Chrien, M. Aronsson, B.J. Chippendale, J.A. Faust, B.E. Pavri, C.J. Chovit, M. Solis, M.R. Olah, and O. Williams, Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, Vol. 65, No. 3, pp. 227-248, 1998.
- [14] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, 2007. Available on-line at: <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>.