AN ICA BASED APPROACH TO HYPERSPECTRAL IMAGE FEATURE REDUCTION

Nicola Falco [a,b], Lorenzo Bruzzone [a], Jon Atli Benediktsson [b]

University of Trento, Information Engineering and Computer Science Dept., Povo, Italy [a]
University of Iceland, Faculty of Electrical and Computer Engineering, Reykjavik, Iceland [b]

ABSTRACT

This article proposes a feature reduction technique for hyperspectral images using Independent Component Analysis (ICA). The proposed technique aims at extracting the best subset of class-informative independent components (ICs) for hyperspectral supervised classification. The selection of the most representative components is assured by the minimization of the reconstruction error, which is computed on the training samples used for the supervised classification. The searching strategy is optimized by exploiting a genetic algorithm-based approach where the fitness function is the classification accuracy obtained by using a support vector machine (SVM) classifier. The obtained results show the effectiveness of the proposed approach in providing class-informative components to improve the classification accuracy.

Index Terms— Independent Component Analysis (ICA), Genetic Algorithm (GA), Feature Reduction, Supervised Classification, Hypersepctral Images, Remote Sensing

1. INTRODUCTION

Hyperspectral images are important for numerous different applications in the Earth Observation (EO) domain. Their high spectral resolution, which is represented by hundreds of narrow spectral channels covering a wide range of the electromagnetic spectrum, offers a detailed representation of the materials on the ground, making this kind of images suitable for land cover classification. However, the classification of hyperspectral images is a complex task due to several factors. First, the high spectral dimensionality of the data makes the analysis computationally intensive. Second, the ratio between the number of available training samples (usually low) and the spectral dimension (usually high) is small, affecting the generalization capability of the classifier (this is known as the Hughes' phenomenon). This problem is usually addressed by exploiting feature selection/extraction techniques as a pre-processing step in order to retain the most useful information. Recently, ICA [1] has been exploited in hyperspectral remote sensing data analysis. ICA is a well known unsupervised blind source separation (BSS) technique, which identifies statistically independent components by considering only the observation of mixture signals. According to the aforementioned observation, ICA can be applied to hyperspectral images, which can be seen as a mixture of signals, aiming at extracting independent components that can provide information related to one or few classes.

In this paper, a new technique for feature reduction based on ICA is presented, which aims at identifying the most informative components, exploiting prior information provided by training samples. In hyperspectral image analysis literature, ICA has been applied after a dimensionality reduction pre-processing step usually performed by Principal Component Analysis (PCA). This is done to reduce the sensitivity of the detected components to the sources of noise present in the images. PCA is based on the eigenvalue analysis of the global covariance matrix in an orthogonal transformation and it has been proven to be excellent for data compression. However, it is well known that the features derived by using PCA are not optimal for the classification task. In this work, ICA analysis is optimized to address the supervised classification task based on the use of prior information provided by training samples. The ICA transformation is first applied to each class singularly, aiming at extracting features that are class-dependent. Second, the most informative endmembers of the mixing matrices obtained from the single-class ICA are selected by solving an optimization problem that minimizes the reconstruction error (of the ICA). The final subset is then optimized by applying a feature selection based on a genetic algorithm (GA) [2, 3]. Finally, the chosen endmembers are used for the transformation of the entire data set, producing a subset of features that is used for the supervised classification. The performance of the approach in retrieving class discriminant information is evaluated in terms of classification accuracies, obtained using a SVM classifier.

The rest of the paper is organized as follows. In Section 2, the ICA model is introduced. Section 3 describes the proposed technique for dimensionality reduction. In Section 4, the experimental results are presented and discussed. Conclusions are drawn and future work is discussed in Section 5.

2. INDEPENDENT COMPONENT ANALYSIS

In this paper vectors and matrices are denoted as bold lowercase and bold uppercase letters, respectively, where the elements are considered as column vectors. Using this notation, the linear mixing model

adopted for hyperspectral images is the following:

$$\mathbf{X} = \mathbf{AS} = \sum_{i=1}^{m} \mathbf{a}_i \mathbf{s}_i^T, \tag{1}$$

where $\mathbf{X}$ is an $m \times p$ matrix composed by m spectral bands and p pixels and whose elements are the mixtures $[\mathbf{x}_1, ..., \mathbf{x}_m]^T$ of the observed image, $\mathbf{A}$ is an $m \times m$ matrix and represents the unknown mixing matrix with elements $[\mathbf{a}_1, ..., \mathbf{a}_m]$ and $\mathbf{S}$ is an $m \times p$ matrix whose elements are the unknown sources $[\mathbf{s}_1, ..., \mathbf{s}_m]^T$. Following this model, the independent component analysis aims at obtaining the best possible approximation of $\mathbf{S}$ by estimating the unmixing matrix $\mathbf{W}$. The ICA model can be written as follows:

$$\mathbf{Y} = \mathbf{WX} \simeq \mathbf{S}, \tag{2}$$

where $\mathbf{Y}$ is a $m \times p$ whose component $\mathbf{y}_i^T$ are statistically independent and $\mathbf{W} \simeq \mathbf{A}^{-1}$. For the estimation of the ICA model, the following assumptions and restrictions need to be satisfied: 1) Statistical independence of the sources; 2) the independent components must have a non-Gaussian distribution; 3) $\mathbf{A}$ is assumed to be squared and fully ranked. Since $\mathbf{W}$ and $\mathbf{S}$ are unknown, three ambiguities related to the ordering, signs, and scaling of the outputs necessarily hold. In this paper, FastICA is chosen as the ICA algorithm, since it has been shown to be the most appropriate method when the entire feature space of an hyperspectral image is considered [4]. In the experimental analysis, an implementation of FastICA algorithm based on Matlab (© The MathWorks, Inc.) scripting language is used. The FastICA algorithm proposed in [5] is a very efficient and robust method for ICA. The algorithm is based on an approximation of negentropy, which is a normalized version of the differential entropy and it is used as a measurement of non-Gaussianity. FastICA implements two different approaches for the estimation of $\mathbf{W}$, deflationary orthogonalization and symmetric orthogonalization [1]. The choice of using the second approach offers two main advantages as compared to the first: it avoids the presence of cumulated errors in the estimation of subsequent vectors by the orthogonalization, and reduces the computation time since the ICs are estimated in a parallel fashion.

3. FEATURE REDUCTION BASED ON ICA

The approach presented in this paper exploits the properties of ICA aiming at extracting the most informative components for supervised classification purposes. Figure 1 shows the general scheme of the approach. Let an hyperspectral image $\mathbf{X}$ be composed of m spectral channels whose ground truth is defined by n classes of interest. The algorithm consists of the following steps:

1. Single-class ICA. The data set is first clustered based on the training samples, obtaining n clusters representing the n classes of interest. For each cluster $\mathbf{X}_{cl}$, where $cl = 1, ..., n$, the unmixing matrix $\mathbf{W}_{cl}$ and the independent components $\mathbf{Y}_{cl}$ are estimated by using FastICA.

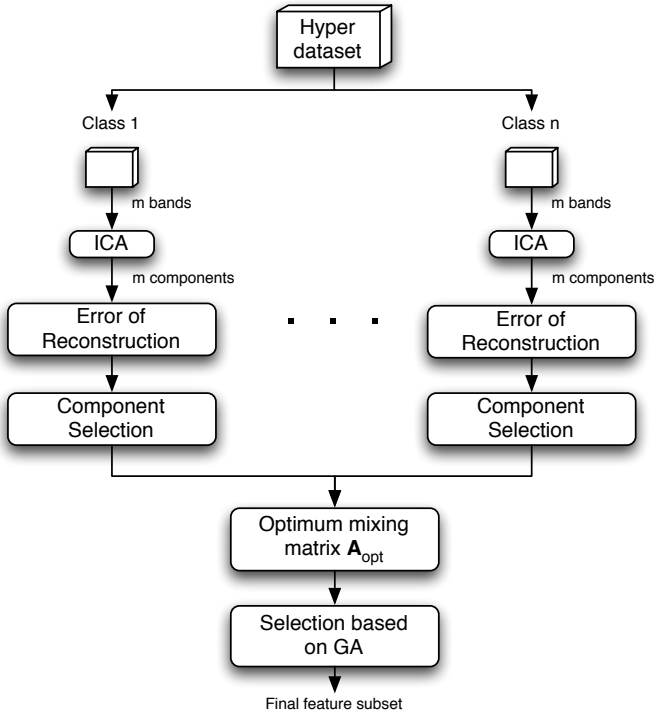


Fig. 1. General scheme of the proposed technique for feature reduction based on ICA

2. Evaluation of the reconstruction error. The reconstruction error is estimated by computing the Frobenius norm, denoted by $\|.\|_F^2$, between the original data set and the back projection of the obtained ICs. It is mathematically defined as follows:

$$E_{cl} = \|\mathbf{X}_{cl} - \mathbf{A}_{cl} Y_{cl}\|_F^2 = \left\| \mathbf{X}_{cl} - \sum_{i=1}^{m} \mathbf{a}_i \mathbf{y}_i^T \right\|_F^2, \tag{3}$$

with $\mathbf{A}_{cl} = \mathbf{W}_{cl}^{-1}$. $\mathbf{a}_i$ is a column vector of the mixing matrix $\mathbf{A}_{cl}$, representing the spectral signature related to the class, and $\mathbf{y}_i$ is a row vector of the ICA matrix $\mathbf{Y}_{cl}$ representing an independent component. Once the error is computed, the m pairs $(\mathbf{a}_i, \mathbf{y}_i)$ are ranked based on their relative contribution. The best components are the ones that minimize the reconstruction error. The indices l of the m couples are identified by applying the following iteration:

$$l^+ = \arg\min_i err(i) = \{i \mid \min_i \left\| \mathbf{X}_l^+ - \mathbf{a}_i \mathbf{y}_i^T \right\|_F^2 \}, \tag{4}$$

$$\mathbf{X}_l^+ \leftarrow \mathbf{X}_l - \mathbf{a}_l \mathbf{y}_l^T. \tag{5}$$

with $i = 1, ..., m$. The initial $\mathbf{X}_l$ corresponds to $\mathbf{X}_{cl}$, which is updated at each iteration by subtracting the contribution provided by $\mathbf{a}_i \mathbf{y}_i^T$, identified at the previous iteration. The algorithm for the computation of the reconstruction error and the l indices for a single class is shown in Algorithm 1.

3. The optimal mixing matrix A_{opt}. From the previous step,

Algorithm 1 Algorithm for the ranking of the couples $\mathbf{a}_i, \mathbf{y}_i^T$ based on the reconstruction error.

$\mathbf{X} \leftarrow \mathbf{X}_{cl}$
for $j \leftarrow 1$ to l **do**
  for $i \leftarrow 1$ to m **do**
    $E_i = \left\|\mathbf{X} - \mathbf{a}_i \mathbf{y}_i^T\right\|_F^2$
  end for
  $i_{opt} = \arg\min_i \mathbf{E}(i)$
  $E_{opt} = \min \mathbf{E}(i)$
  $\mathbf{X} \leftarrow \mathbf{X} - \mathbf{a}_{i_{opt}} \mathbf{y}_{i_{opt}}^T$
  $\mathbf{a}_{i_{opt}} \leftarrow \infty$
  $\mathbf{y}_{i_{opt}} \leftarrow \infty$
  $idx_j = i_{opt}$
  $E\_vec_j = E_{opt}$
end for
return idx, **E_vec**

we can define for each class a matrix $\mathbf{A}'_{cl}$ that is composed of the best elements $[\mathbf{a}_1, ..., \mathbf{a}_l]$, where l is the total number of endmembers selected for a given class. Then, $\mathbf{A}_{opt} = [\mathbf{A}'_1, \mathbf{A}'_2, ..., \mathbf{A}'_m]$ represents the optimal mixing matrix.

4. Endmember selection based on GA. The obtained $\mathbf{A}_{opt}$ is an $m \times (n * l)$ matrix. Based on the choice of l, the optimal matrix $\mathbf{A}_{opt}$ can have a quite high dimensionality. Moreover, since the selection is computed on each class independently, $\mathbf{A}_{opt}$ can include endmembers with redundant information. In order to find the best subset, a further selection based on GA is performed. The selection is performed on the elements of the matrix $\mathbf{A}_{opt}$, where a chromosome identifies which endmembers are selected for the transformation, and the fitness function is evaluated on the transformation $\mathbf{Y}_{GA} = \mathbf{W}_{GA}\mathbf{X}$, where $\mathbf{Y}_{GA}$ represents the new ICs that are used as input to SVM, $\mathbf{W}_{GA} = \mathbf{A}_{GA}^{-1}$ is the new unmixing matrix derived from the reduced version of $\mathbf{A}_{opt}$ and $\mathbf{X}$ is the original observed image. In our strategy, the classification accuracy obtained by using a SVM classifier with Radial Basis Function (RBF) kernel represents the fitness function to be maximized. Since the kernel parameter estimation is computationally expensive, the estimation is performed once for each population based on 5-fold cross-validation.

5. Computation of the final feature subset. The final subset is obtained by performing the linear transformation considering all the pixels that compose the data set by using $\mathbf{W}_{GA}$. For the computation of the unmixing matrix, which can represent an underdetermined system, we used the Moore-Penrose pseudoinverse. The subset is then used as an input to a SVM classifier.

It is worth mentioning that steps 1 and 2 can be performed in a parallel distributed system, as proposed in [6], since the analysis of each class is independent of the other. In this way the computational time of the ICA for the entire data set is significantly decreased and can be approximated to the computational time of a single class ICA. Also, the selection based on GA can be optimized by computing the fitness function for each population in parallel fashion.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental analysis of the proposed technique was carried out on two real hyperspectral data sets.

The first data set shows an area of the Salinas Valley, California. The image size is 512 × 217 pixels with a spatial resolution of 3.7 m. In this study, corrected data are obtained by discarding the 20 water absorption bands: [108-112], [154-167], 224. A false color composition of the image and the reference map are shown in Figures 2a and 2b, respectively, containing 16 classes of interest. The training set used in the experiments is made up of 15% randomly selected samples from each class.

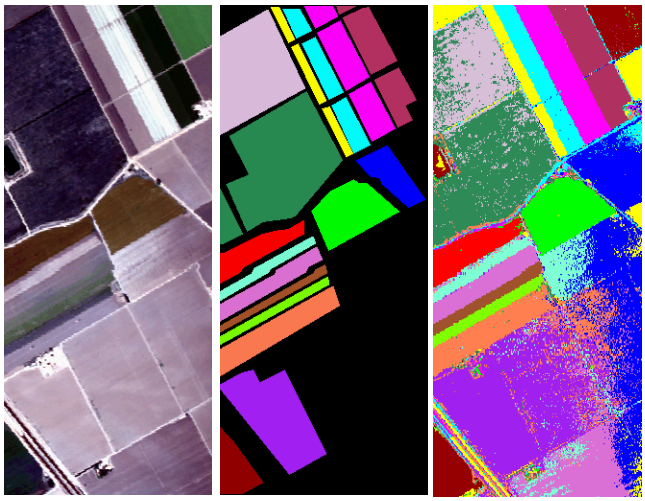
The second data set was acquired over the active Hekla volcano, which is located in south-central Iceland, by the 224-band AVIRIS sensor. After discarding the noisy and the blank channels, the final data set counted 157 spectral bands. The image has a spatial dimensions of 600 × 560 pixels with a spatial resolution of 20 m. The ground reference data contain 12 classes of interest. Figures 3a and 3b show a false color composition of the image and the reference map, respectively. For this data set, the training set used here was generated by a random selection of 50 samples from each class.

For both data sets, the presented approach is tested considering $l = 1, ..., 4$. This means that for the Salinas data set, whose reference data is made up of 16 classes, the sizes of $\mathbf{A}_{opt}$ are 16, 32, 48, 64, respectively, while for Hekla they are 12, 24, 36, 42, being composed of 12 classes. The GA parameters (rate of crossover, rate of mutation, population size, etc.) were determined empirically through a set of preliminary experiments. In this work, a uniform crossover is used, with a crossover rate of 0.80 and a mutation rate of 0.01, while the length of a chromosome varies as $n * l$ and corresponds to the length of the given subset. Table 1 shows the classification accura-

Table 1. Classification of Salinas and Hekla data sets. The table reports the number of features selected based on the reconstruction error (No. feat.), the number of features after the GA selection (No. feat. GA), the percentage overall accuracy (OA (%)) and the kappa coefficient (k). Classification results obtained by the original spectral bands are given for comparison.

	Salinas				
	Spectr.	$l = 1$	$l = 2$	$l = 3$	$l = 4$
No. feat.	204	16	32	48	64
No. feat. GA		12	18	23	31
OA (%)	94.57	94.21	95.34	95.37	95.00
k	0.91	0.94	0.95	0.95	0.94

	Hekla				
	Spectr.	$l = 1,$	$l = 2$	$l = 3$	$l = 4$
No. feat.	157	12	24	36	48
No. feat. GA		7	13	19	23
OA (%)	93.89	91.97	93.53	95.24	95.21
k	0.80	0.91	0.93	0.95	0.95



(a) (b) (c)

Fig. 2. Salinas data set: (a) Hyperspectral image (in false color) and (b) the reference map. (c) Classification map obtained by using the proposed technique ($l = 3$).

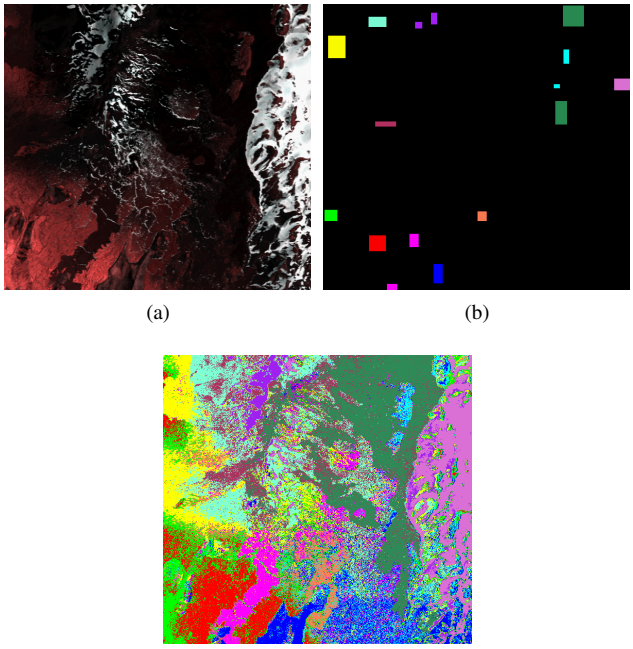
cies obtained by using the presented approach, achieving an OA of 95.37% (with 23 ICs) in case of Salinas, outperforming the OA obtained in the approach presented in [4] (which resulted in an OA of 94.12% using 35 ICs). Whereas for Hekla, the method achieves an OA of 95.24% (with 19 ICs), sharply outperforming the OA obtained in [4] (which resulted in an OA of 82.58% using 20 ICs). Figure 3c shows the resulting classification maps obtained for the Salinas and Hekla data sets by the proposed method. Overall, in both cases, the classification accuracy was improved retaining fewer components and a with significant decrease of the computational complexity.

5. CONCLUSION

In this article, a feature reduction technique for hyperspectral images using ICA was presented. The goals of this work were to extract class-informative features and at the same time to decrease the computational cost required by the ICA when applied to the entire feature space. The latter was achieved by employing a parallel implementation of ICA. Class-informative subsets of independent components were identified by analysing the error of reconstruction considering each class independently. The feature selection was then optimized by employing a GA-based approach, leading to an additional reduction. The experimental results showed the effectiveness of the proposed approach in providing informative components for classification.

6. ACKNOWLEDGMENTS

This research was in part supported by the research funds of the University of Iceland, in part by the research funds of the University of Trento and by EU FP7 Theme Space project North State.



(a) (b)

(c)

Fig. 3. Hekla data set: (a) Hyperspectral image (in false color) and (b) the reference map. (c) Classification map obtained by using the proposed technique ($l = 3$).

7. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., 2001.

[2] M. Melanie, *An Introduction to Genetic Algorithms*, MIT Press, 1996.

[3] A. B. Santos, C. S. F. de S. Celes, A. de A. Araújo, and D. Menotti, "Feature selection for classification of remote sensed hyperspectral images: A filter approach using genetic algorithm and cluster validity," in *The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV12)*, 2012, vol. 2.

[4] N. Falco, L. Bruzzone, and J. A. Benediktsson, "A Comparative Study of Different ICA Algorithms for Hyperspectral Image Analysis," in *5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS '13)*, Gainesville, Florida, 2013.

[5] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis.," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–34, Jan. 1999.

[6] S. A. Robila, "Distributed source separation algorithms for hyperspectral image processing," in *SPIE 5425, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, Aug. 2004, vol. 5425, pp. 628–635.