Linear Versus Non-Linear Analysis of Relevant Scatterers in High Resolution SAR Images

Houda Chaabouni-Chouayakh* and Mihai Datcu* †
* DLR, German Aerospace Center, Oberpfaffenhofen D-82234 Wessling - Germany
e-mail: houda.chaabouni@dlr.de, mihai.datcu@dlr.de
† ENST, Ecole Nationale Supérieure des Télécommunications, 46 rue Barrault 75634 Paris Cedex 13 - France
e-mail:mihai.datcu@enst.fr

***Abstract*—With the increase of Synthetic Aperture Radar (SAR) sensor resolution, SAR images could include a large variety of interesting real man-made structures. Therefore, a more detailed analysis and a finer description of SAR images of urban areas are needed for a better understanding of the scene.**

Nevertheless, recognizing scenes using high resolution SAR images requires the capability to identify relevant signal signatures (called also descriptors), depending on variable image acquisition geometry, arbitrary objects poses and configurations. Among feature extraction methods, we propose to use Principal Components Analysis (PCA) and/or Independent Components Analysis (ICA), in order to exploit deeper the nature of SAR signatures. In this paper, both a description of our work and a presentation of our preliminary classification performance results will be provided.

I. Introduction

With the rapid development of Synthetic Aperture Radar (SAR) technologies, the need for automatic processing of large-size SAR data is getting more important. In fact, high resolution SAR images could include signatures describing structures (buildings, antennas, roads, lights, ground vehicles...), that have various shapes, materials (metallic, asphalt,...) and orientations from the sensor. These wide diversity structures generate SAR images showing different high-complexity behaviors. In some cases, the behaviors of the backscatterers may be in direct relation with the objects that exist in the scene (e.g. antennas). However, in some other cases, it could be related to some sub-parts of the scene objects such as dihedral/corner reflector (as the house wall/road...).

For better SAR Automatic Target Recognition (ATR), it is important to get reliable signatures/features, that better model the targets. A well-defined feature space is a space where signatures from different classes can be well-separated. Then, a parsimonious feature space could be generated by selecting the best discriminating and the less redundant features.

This work addresses a feature extraction/selection problem for high resolution SAR ATR, based on the Principal Components Analysis (PCA) and/or Independent Components Analysis (ICA).

The PCA transform produces an orthogonal basis (the eigenspace of the covariance matrix). Nevertheless, the covariance analysis is describing only linearly dependent structures in the SAR signals. Thus, it is more suitable for analyzing Gaussian data. However, high resolution SAR images contain edges of different shapes and sizes and could not be described only by Gaussian processes. In order to exploit deeper the nature of such complicated signatures, a combination between the PCA and ICA based methods could be more informative and more descriptive.

The organization of this paper follows: Section II is dedicated to a brief description of the feature extraction methods, that we used for our experiments (PCA and ICA), as well as a comparison between them. Then, section III introduces the concept of feature selection. In section IV, we report about the preliminary results that we obtained with our high resolution SAR database, while section V gives some conclusions.

II. Feature extraction

Feature extraction is an essential pre-processing step to pattern recognition and machine learning problems. It is a method to simplify the amount of signatures, required to describe a large set of data. Indeed, analysis with a large number of variables generally requires a large amount of memory and computation power. Feature extraction is a general term for methods of constructing combinations of variables, to get around these problems, while still describing the data with sufficient accuracy.

In the literature, the PCA and ICA based methods were widely used for feature extraction, from different kinds of signals [1]–[4]. Recently, a comparison of the performance of the two methods, on artificial and optical databases, was studied in [5].

A. Principal Components Analysis (PCA)

PCA is the most popular statistical method for feature extraction. It is based on the assumption that high information corresponds to high variance. The PCA transform is defined as follows:

$$Y = H^T X, \tag{1}$$

where X is $d \times n$ dimensional vector samples, Y is transformed $m \times n$ dimensional vector samples, and H is a $d \times m$ transform matrix.

H is calculated as the m largest eigenvectors of the $d \times d$ covariance matrix of X. In fact, it is assumed, in this case, that most of the X'information content is stored in the directions

of the maximum data variance, under the constraint of orthogonality. Since the m largest eigenvalues equal the maximal variances, the m corresponding eigenvectors are exactly the columns of the matrix H. It is also worth to note that the transformation defined in (1), gives uncorrelated components.

This method effectively represents data in a linear subspace with minimum information loss.

B. Independent Components Analysis (ICA)

ICA is a de-mixing process whose goal is to express a set of random variables as linear combinations of statistically independent component variables.

The ICA model can be expressed as:

$$x = As, \tag{2}$$

where $x = (x_1, x_2, ..., x_m)^T$ is the measured data, A is the $m \times n$ mixing matrix, and $s = (s_1, s_2, ..., s_n)$ are the n unknown independent components.

The estimation, in the model (2), is performed by trying to find a solution for the problem:

$$y = Wx, \tag{3}$$

where $y = (y_1, y_2, ..., y_n)^T$ are statistically independent (called also the estimated sources of the s_i's), and W is equal to the pseudoinverse matrix of A.

Typically, in ICA algorithms, the vectors w, are sought, such that the rows of y have as many non-Gaussian distributions as possible, and are mutually (approximately) uncorrelated. One alternative to do this, is to first whiten the data, and then to seek orthogonal non-normal projections. In the literature, PCA was considered as a solution for the whitening problem (decorrelation solution). Indeed, a zero-mean random vector is said to be white, when its elements are uncorrelated and have unit variances (covariance matrix equal to the unit matrix I). Thus, the whitening process can be accomplished by decorrelation, via the PCA technique, followed by scaling.

For ICA, the higher order statistics are usually incorporated in the estimation procedures, by means of the so-called contrast functions based on higher order cumulants. The choice of these contrast functions was analyzed in [6]. It was demonstrated that, in neural learning rules, better ICA estimation could be reached, when using tanh-like sigmoids, or functions resembling the derivative of a Gaussian function.

C. Comparison between PCA and ICA

The ICA based feature extraction method gets higher order statistics. It is, thus, able to provide a more powerful and finer data description than PCA, if we are under the assumption that the information which distinguishes images, is contained in the higher order statistics. In fact, PCA only requires that its components are uncorrelated, while ICA requires its components to be independent.

III. Feature selection

Feature selection is the technique, commonly used in machine learning, for selecting a subset of relevant features, in order to build robust learning models. In fact, by removing most irrelevant and redundant features from the data, feature selection leads to better performance results, by providing more suitable data modeling, and speeding up the learning process.

IV. Results and Discussion

A. Description of the database

For our experiments, we used a three-class high resolution SAR database (high density urban areas, low density urban areas, and vegetation). They are intensity images, collected from the same SAR image, over the city of Dresden in Germany, acquired with the Experimental SAR system (E-SAR), of the German Aerospace Center (DLR). One sample of each class is provided in Fig. 1.

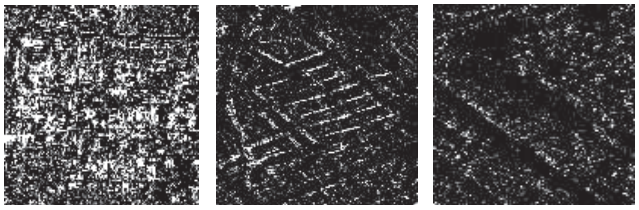

Fig. 1. Some samples of the Dresden database. From left to right: high density urban area, low density urban area, and vegetation.

More details about the Dresden database and the extracted features, are summarized in Table I.

TABLE I
Description of the Dresden database.

Number of classes	3
Number of images per class	50
Size of the images	64×64
Number of PCA features per image	7
Number of ICA features per image	7

B. Description of the experiments

A number of popular ICA algorithms exist in the literature. For our computations, we used the FastICA introduced in [7]. The first step of the algorithm consists in whitening the data by classical PCA. This means that the original data x_{old} is linearly transformed to a variable $x = Qx_{old}$, such that the correlation matrix of x equals unity. Then, to the whitened data, a fixed-point algorithm is applied to seek for the ICs. The fixed-point iteration scheme aims at finding the local extrema of a contrast function (non-linearity function) of a linear combination of the observed variables. The Matlab FastICA code is available at: http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.shtml. For our experiments, we used $\tanh$ as a non-linearity.

The feature extraction process is performed just once, using all the available examples in each class. Then, 20 repetitions run the classifier with randomly selected training and test

sets. This approach is called cross-validation, which aims at estimating how well the model, we have just learned from some training data, is going to perform on future as-yet-unseen data. By using such an approach, we avoid any possible dependency between the training data and the classification performance. In each repetition, 50% of the data is used for training and the other 50% for testing.

Four different global experiments have been performed:

- **ICA vs. PCA comparison without feature selection**
  In this first experiment, both PCA and ICA are applied to extract features from the data, and all the components are used for classifying the examples. The classifier used is a 1-Nearest Neighbor (1-NN) with Euclidean distance. The recognition rates obtained for the two feature extraction methods, without feature selection, are shown in Fig. 2.

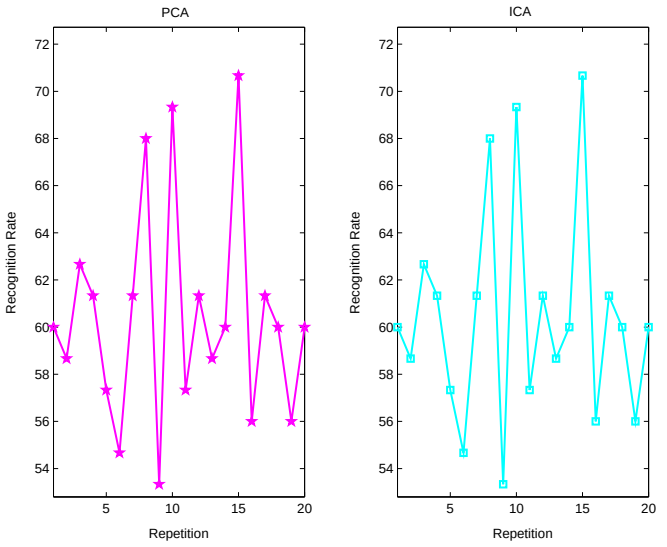

Fig. 2. PCA and ICA comparison using all the components, as a function of the repetitions.

From Fig. 2, it is clear that PCA and ICA perform equally. The two methods extract the same quantities, and project them in different subspaces (orthogonal subspace for PCA against statistically independent subspace for ICA). In fact, the PCA is the first step (whitening step) of the FastICA algorithm, and the fixed-point algorithm (second step of FastICA) transforms the PCA features in ICA feature subspace, while maintaining exactly the same global information and without changing the total number of features.

- **ICA vs. PCA comparison with feature selection**
  In this second experiment, both PCA and ICA are used to compress the data but not all the components are used for classification. The goal is to detect possible advantages of ICA over PCA when only a subset of the components are used for classification. In other words, the goal is to check whether ICA is able to extract certain components, that are particularly appropriate, for a good discrimination, between the different classes in the database.
  Fig. 3 shows the recognition rates (averages and standard deviations of 20 repetitions), that could be obtained, for each algorithm, if a perfect selection of components is carried out. The k-Nearest Neighbor (k-NN) was used for feature selection, in our case.

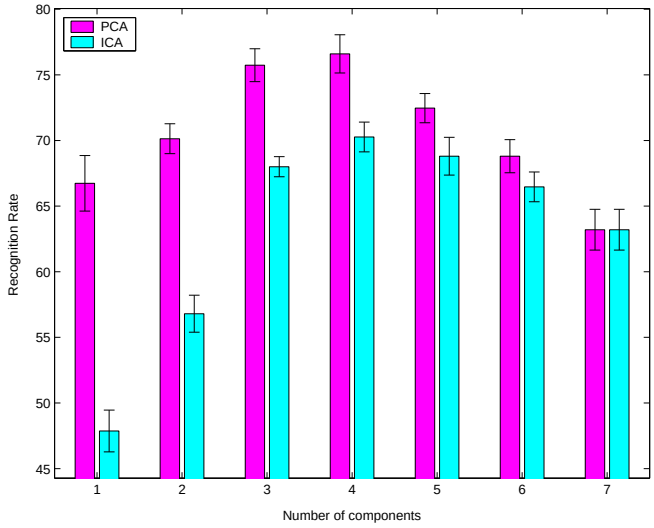

Fig. 3. PCA and iCA comparison, considering feature selection, as a function of the number of components.

From the results of Fig. 3, we can notice that, when considering feature selection, PCA and ICA no longer perform equally. The results show that ICA does not necessarily outperform PCA. The recognition rates are highly dependent on the number of selected features.
It is also easy to see how a feature selection step may help to reduce the dimensionality and improve the classification results obtained by each algorithm.

- **Combination of PCA and ICA features**
  In order to exploit both the descriptors extracted by the PCA (orthogonality properties) and the ones extracted by the ICA (independency properties), we suggest, in this part, to combine the features of the two methods and then to perform the selection on the new set of features. In the following, we call our new combination algorithm Principal and Independent Components Analysis (P-ICA).
  The flowchart of our new P-ICA algorithm is given by Fig. 4.

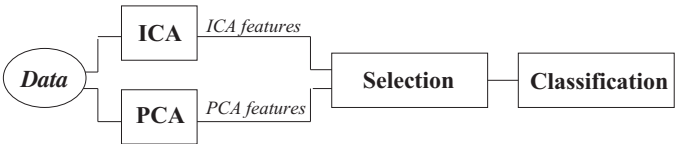

Fig. 4. Flowchart of the P-ICA algorithm.

The averages and standard deviations of 20 repetitions of PCA, ICA and P-ICA algorithms, considering feature selection via k-NN, are shown in Fig. 5.

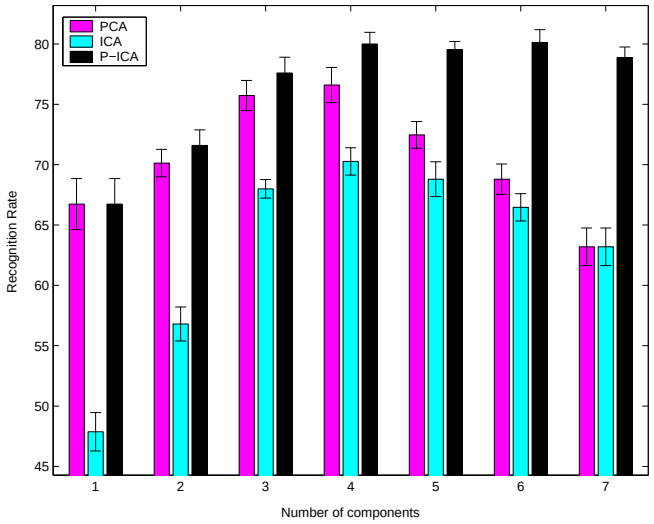

Fig. 5. PCA, ICA and P-ICA comparison, considering feature selection, as a function of the number of components.

From the classification performance presented in Fig. 5, it is clear that the combination of the PCA and the ICA features improves advantageously the classification results, for all the used numbers of components. We can also notice the gain achieved by the feature selection process. More particularly, when the number of components is very high, the recognition rates obtained by our P-ICA algorithm are clearly better than PCA or ICA, used separately.

- **How much are the PCA and ICA contributions in the P-ICA feature selection process?**
  To analyze the contribution of each method (PCA and ICA) in the P-ICA feature selection process, we illustrate, in Fig. 6, the averages and standard deviations of 20 repetitions of the PCA and ICA feature contribution rates.

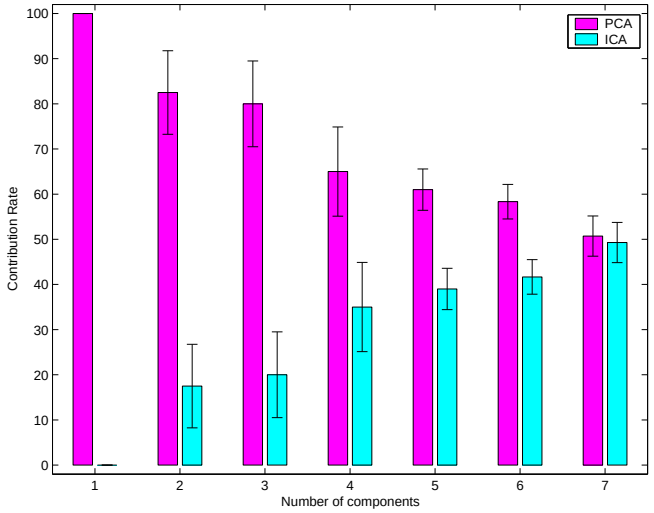

Fig. 6. PCA and ICA feature contributions in the P-ICA feature selection process, as a function of the number of components.

From Fig. 6, it is clear that, as the number of selected components increases, the contribution of ICA gets higher, while the one of PCA gets lower.

V. Conclusions

Feature extraction/selection from a three-class high resolution SAR database problem was studied in this paper. A combination of PCA and ICA (P-ICA) was proposed as a new feature extraction method. Such an approach was demonstrated to be more informative than when working with PCA and ICA separately.

The advantage of the feature selection process was also highlighted in this paper. In fact, selecting only the most relevant and the less redundant features makes the modeling of the data more appropriate, and the learning process faster.

Acknowledgments

The authors would like to thank the FastICA team of the Laboratory of Information and Computer Science in the Helsinki University of Technology, for the availability of the FastICA Matlab code.

This work was performed in the frame of the CNES/DLR/ENST center of competence on information extraction and image understanding for earth observation.

References

[1] L.M. Novak and G.J. Owirka. Radar Target Identification Using an Eigen-Image Approach. *IEEE National Radar Conference*, Atlanta, GA, 1994.

[2] B. Moghaddam and A. Pentland. A Subspace Method for Maximum Likelihood Target Detection. *IEEE International Conference on Image Processing*, Washington DC, 1995.

[3] P. O. Hoyer and A. Hyvrinen. Independent Component Analysis Applied to Feature Extraction from Colour and Stereo Images. *Network: Computation in Neural Systems*, 11(3):191-210, 2000.

[4] J.F. Cardoso. Blind Signal Seperation: Statistical Principles. *Proc. IEEE*, 9(10), 2009-2025, 1998.

[5] M.A. Vicente, P.O. Hoyer and A. Hyvrinen. Equivalence of Some Common Linear Feature Extraction for Appearence Based Object Recognition Tasks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 29, No.5, pp. 896-900, May 2007.

[6] A. Hyvrinen. One-Unit Contrast Functions for Independent Component Analysis: a Statstical Analysis. *Neural Networks for Signal Processing VII (Proc. IEEE NNSP Workshop '97* , Amelia Island, Florida), pp. 388–397, 1997.

[7] A. Hyvrinen. Fast and Robust Fixed Point Algorithms for Independent Components Analysis. *IEEE Trans. Neural Networks*, Vol.10, No. 3 pp. 626-634, 1999.