PCA vs. ICA Decomposition of High Resolution SAR Images: Application to Urban Structures Recognition

Houda Chaabouni-Chouayakh[a] and Mihai Datcu[a,b]

[a]German Aerospace Center (DLR), Oberpfaffenhofen D-82234 Wessling, Germany;
[b]Telecom Paris, 46 Rue Barrault, 75013 Paris, France

ABSTRACT

With the increase of the Synthetic Aperture Radar (SAR) sensor resolution, a more detailed analysis and a finer description of SAR images are needed. Nevertheless, when dealing with urban areas, the high diversity of man-made structures combined with the complexity of the scattering processes makes the analysis and information extraction, from high resolution SAR images over such areas, not easily reachable.

In general, an automatic full understanding of the scene requires the capability to identify both relevant and reliable signatures (called also features), depending on variable image acquisition geometry, arbitrary objects poses and configurations. Then, since SAR images are formed, by coherently adding the scattered radiations from the components of the illuminated scene objects, we can make the assumption that, the SAR image is a superposition of different sources. Following this approach, one alternative for a better understanding of the HR SAR scenes, could be a combination between the Principal Components Analysis (PCA) and the Independent Components Analysis (ICA) decompositions. Indeed, while the PCA exploits at most the information stored in the sample covariance matrix, the ICA is a de-mixing process whose goal is to express a set of random variables as linear combinations of statistically independent component variables. Such an approach could be useful for the recognition of urban structures, in HR SAR images. In this paper, we compare the Principal Components (PCs) to the Independent Components (ICs). Furthermore, we present some preliminary results on learning and decomposing SAR images, using PCA and ICA.

Keywords: HR SAR, relevant and reliable signatures, PCA, ICA

1. INTRODUCTION

Satellite imagery has found vast applications in a wide spectrum of areas including agriculture (e.g. detection of crop types), urbanization (e.g. tracking the development of urban areas), cartography (e.g. detection of rivers, road networks), surveillance (e.g. detection and if possible recognition of military targets), etc. This heavy demand on satellite imagery applications leads to the development of imaging systems that are alternative to optical imagery. In particular, in the last few decades, there has been an increasing interest in Synthetic Aperture Radar (SAR) imagery, since some of its properties are favorable to optical imagery. In fact, SAR is a coherent imaging mode in the microwave domain that can operate regardless of weather conditions, and whose resolution is independent of sensor height.

With the increase of the SAR sensor resolution, High Resolution (HR) SAR images could include a wide diversity of man-made structures (buildings, antennas, roads, lights, ground vehicles, etc), that have various shapes and sizes, different materials (metallic, asphalt, etc) and varying orientations from the sensor. The high diversity properties of HR SAR image contents over urban areas could be easily seen on the examples provided by Figure 1.

These wide diversity structures generate HR SAR images showing different high-complexity backscattering behaviors. Indeed, the electromagnetic scattering, in this case, is characterized by a variety of single or multiple scattering mechanisms with a wide range of scattering amplitudes.

Houda Chaabouni-Chouayakh: E-mail: houda.chaabouni@dlr.de
Mihai Datcu: E-mail:mihai.dactu@dlr.de, mihai.datcu@enst.fr

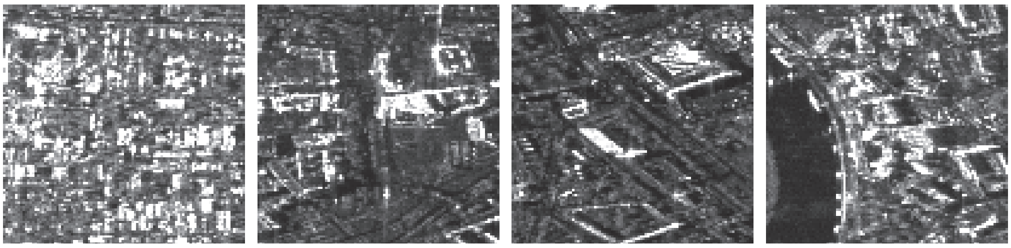

Figure 1. Some examples of HR SAR images over urban areas showing the high diversity of the contents.

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).

Moreover, HR SAR images over urban areas are strongly affected by geometric distortion effects (as layover, shadowing) due to the combination of the SAR side-looking acquisition and stepwise height variations within the scene.

All these awkward characteristics of urban areas combined together, makes the interpretation and information extraction over such areas from SAR images more complex to perform.

In order to provide powerful Automatic Target Recognition (ATR) tools in general, and HR SAR ATR tools in particular, it is extremely important to get both relevant and reliable signatures/features that better model the SAR complicated behaviors of the targets that may exist in the scene. A feature space is called well-defined if the signatures, that it provides from different classes, can be well-separated. After that, a parsimonious feature space could be generated by selecting only the best discriminating and the less redundant features. This step is called the feature selection.

All the modules that form any ATR scheme are highly dependent on the relevance of the extracted signatures. That's why the feature extraction problem was and is still a field of ongoing research. Indeed, many approaches in the literature were developed in order to propose solutions to this problem. Among them, Principal Components Analysis (PCA) and Independent Components Analysis (ICA) based methods were widely used, for the analysis and the compression of different kinds of signals and images.[1–4] Recently, a comparison between the performance of the two methods, on artificial and optical databases, was also studied.[5]

This work addresses a feature extraction/selection problem for HR SAR ATR, based on a combination between the PCA and the ICA. Such a combination should be more informative and better descriptive. In fact, the PCA transform produces an orthogonal basis (generated by the eigenvectors of the sample covariance matrix, that correspond to the largest eigenvectors). It is based on the assumption that the most relevant information corresponds to the highest variances. Nevertheless, the covariance analysis is describing only linearly dependent structures in SAR signals. Thus, it is more suitable for analyzing Gaussian data. However, some parts of HR SAR images contain edges of different shapes and sizes and could not be described only by Gaussian processes. For the interpretation of such SAR parts, the ICA could be more suitable and more informative since it gets higher order statistics.

The organization of this paper follows: Section 2 is dedicated to the methodology that we followed. Then, section 3 describes the HR SAR database on which we tested our algorithms. In section 4, we report some preliminary results, while section 5 gives some conclusions.

2. METHODOLOGY

The SAR ATR process are designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Indeed, with the rapid development of SAR technologies, the need for automatic processing of large-size SAR data is getting more and more pressing.

In the context of the ATR, and when the goal is to find a good model for the different objects and structures that may exist in the images, the terms feature extraction, feature selection, as well as, cross validation are used quite often. This section will thus, be dedicated to a short theoretical overview of the different methods that we have used in order to perform these three modules.

2.1 Feature extraction

Feature extraction is an operation consisting in extracting various image features for identifying or interpreting meaningful physical objects from images. It attempts to aggregate or combine the features in such a way to extract the common information contained in them that is most useful for building the model. In fact, since patterns in data can be hard to find in data with high dimension, where the luxury of graphical representation is not available, feature extraction methods could be seen as powerful tools for the analysis and the interpretation of the data.

In this paper, in the feature extraction step, we propose PCA and ICA based methods. In the following, a brief description of their principles, as well as the relationship (similarities and dissimilarities) between them is given.

2.1.1 Principal Components Analysis (PCA)

PCA is the most popular statistical method for feature extraction. It is based on the assumption that reliable 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. It is calculated as the m largest eigenvectors of the $d \times d$ covariance matrix $\Sigma = XX^T$ 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.

2.1.2 Independent Components Analysis (ICA)

The ICA is a statistical data analysis method that has gained popularity during the last decade. It 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 follows:

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

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, or rather, an approximative solution for the following problem:

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

by optimizing

$$max(g(y)), \tag{4}$$

where:

- $y = (y_1, y_2, ..., y_n)^T$ are statistically independent (called also the estimated sources of the s_i's);
- W is equal to the pseudoinverse matrix of A; and
- g is the so-called contrast function. It is a non-linear function measuring the independency between the n estimated independent sources (the y_i's). The choice of the suitable contrast function was analyzed previously in the literature[6] and 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.

Typically, in ICA based 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). Thus, the whitening process can be accomplished by decorrelation, via the PCA technique, followed by scaling.

A number of popular ICA algorithms exist in the literature. For our computations, we have chosen to use the FastICA algorithm,[7] because of its appealing convergence and high computational speed for high dimensional data.

The FastICA algorithm is a two-step algorithm consisting in:

- **Step 1:** 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.
- **Step 2:** a fixed-point algorithm is applied to the whitened data, in order to seek for the Independent Components (ICs). The fixed-point iteration scheme aims at finding the local extrema of the non-linear contrast function g of a linear combination of the observed variables. For our experiments, we used tanh as a contrast function.

The Matlab FastICA code is available at: `http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.shtml`.

2.1.3 PCA vs. ICA

Both PCA and ICA based methods formulate a general objective function that define the 'interestingness' of a linear representation, and then maximize that function. A second relation between PCA and ICA is that both are related to factor analysis, though under the contradictory assumptions of Gaussianity and non-Gaussianity, respectively.

However, PCA uses only second-order statistics (variances which actually, correspond to the largest eigenvectors of the sample covariance matrix), while ICA 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 the contents of the images and discriminates between the different classes, 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.

2.2 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.

It is one of preliminary stages in the ATR process when the data set includes more variables than could be included (or would be efficient to include) in the actual model building phase (or even in initial exploratory operations). In fact, it attempts to identify the best features among the (sometimes thousands of) available features, and thus leads to better performance results, by providing more suitable data modeling (good discrimination between the different classes), and speeding up the learning process (since we will work with a subset of the features).

For our experiments, we have chosen to use the k-nearest neighbors (k-nn) in order to perform the feature selection process and then the classification. In fact, the k-nn algorithm is amongst the simplest of all machine learning algorithms. It is based on collecting prototypes (typical feature vectors) representing the classes. This is the so-called training phase. Then, in the classification step, the data vectors to be classified are compared to the prototypes, and the class of a data vector is decided based on the k smallest distances from the class prototypes. An object is classified by a majority vote of its neighbors, with the object being assigned the class most common amongst its k nearest neighbors. If $k = 1$, then the object is simply assigned the class of its nearest neighbor.

2.3 Cross validation

The cross-validation 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.

For each of our experiment, we chose to make 5 repetitions with randomly selected train and test data. In each repetition, 50% of the data is used for training and the rest for testing.

3. DATABASE DESCRIPTION

For our experiments, we used a four-class HR SAR database including high density urban areas, low density urban areas, vegetation and water. One sample from each class is provided in Figure 2.

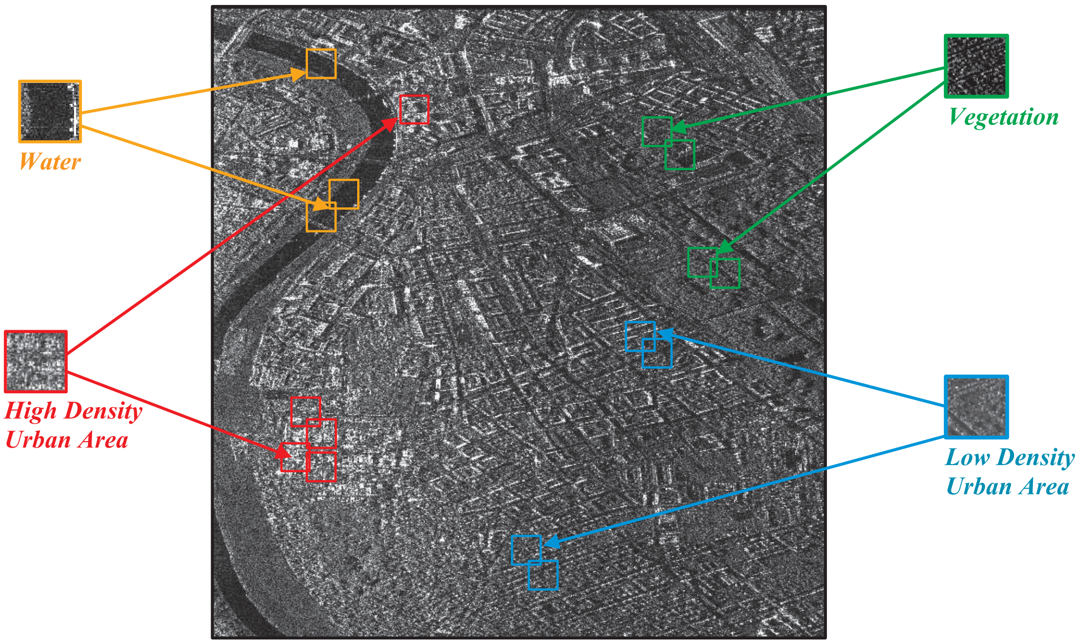

Figure 2. HR SAR image over the city of Dresden in Germany and samples of each class.

They are intensity images, collected from the same Single Look Complex (SLC) HR SAR image, over the city of Dresden in Germany (August 2000), acquired with the Experimental SAR system (E-SAR), of the German Aerospace Center (DLR).

More details about the Dresden SAR image, the database characteristics and the extracted features are stored in Table 1.

Table 1. Description of Dresden database.

Frequency domain	X-Band
Azimuth resolution	1.5m
Range resolution	0.6m
Number of classes	4
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

4. EXPERIMENTAL RESULTS

4.1 Which is the information encoded by the PCA and the ICA methods?

In order to learn the information encoded by the PCA features as well as the one by the ICA features, we summarized in Figure 3, the feature basis generated by each method.

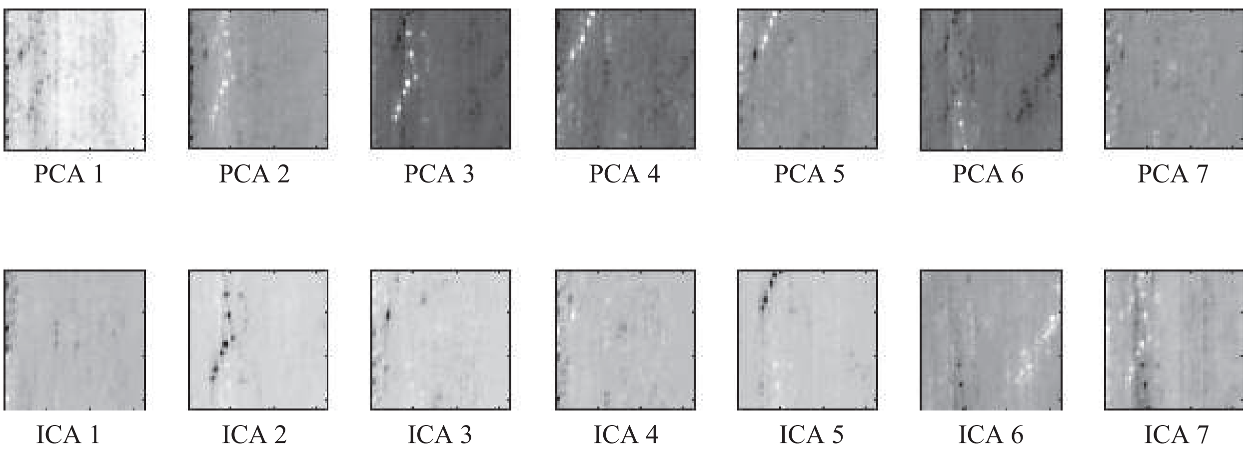

Figure 3. Feature basis: the PCA basis in the top row and the ICA basis in the bottom row.

From Figure 3, we can notice the big dissimilarity between the features extracted by the PCA and the ones extracted by the ICA. Each method is looking to the data from a different angle, and thus provides a new interpretation of the image contents (a new modeling of the similarities and the differences that may exist between the objects and the structures of the SAR scene).

Therefore, a combination between PCA and ICA could be more informative since it exploit deeper the nature of the complicated HR SAR signatures by merging the orthogonality properties of the PCA features, as well as the independency properties of the ICA features.

4.2 Combination between PCA and ICA: P-ICA algorithm

In a previous publication,[8] we have proposed a new algorithm P-ICA combining PCA and ICA to perform the feature extraction step. In this paper, the same technique will be applied. Then, a selection, using the k-nn algorithm is performed on the new set of features in order to preserve only the components whose combination gives the best discrimination between the different classes.

The flowchart of the P-ICA algorithm is given by Figure 4.

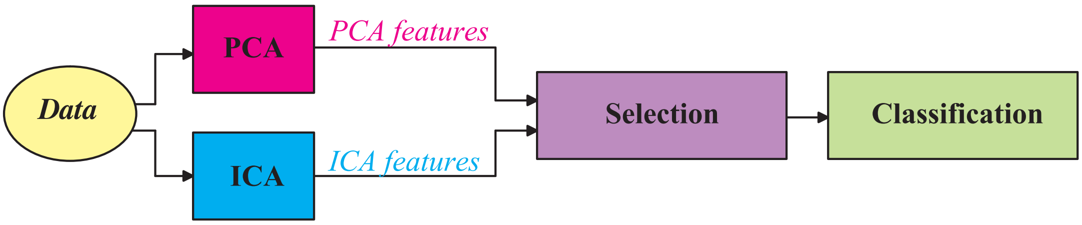

Figure 4. The P-ICA algorithm flowchart: Combination between PCA and ICA in the feature extraction step.

4.3 P-ICA classification results

The recognition rates (averages and standard deviations of 5 repetitions), that could be obtained by using PCA, ICA and P-ICA if a perfect selection of components is carried out, are shown in Figure 5.

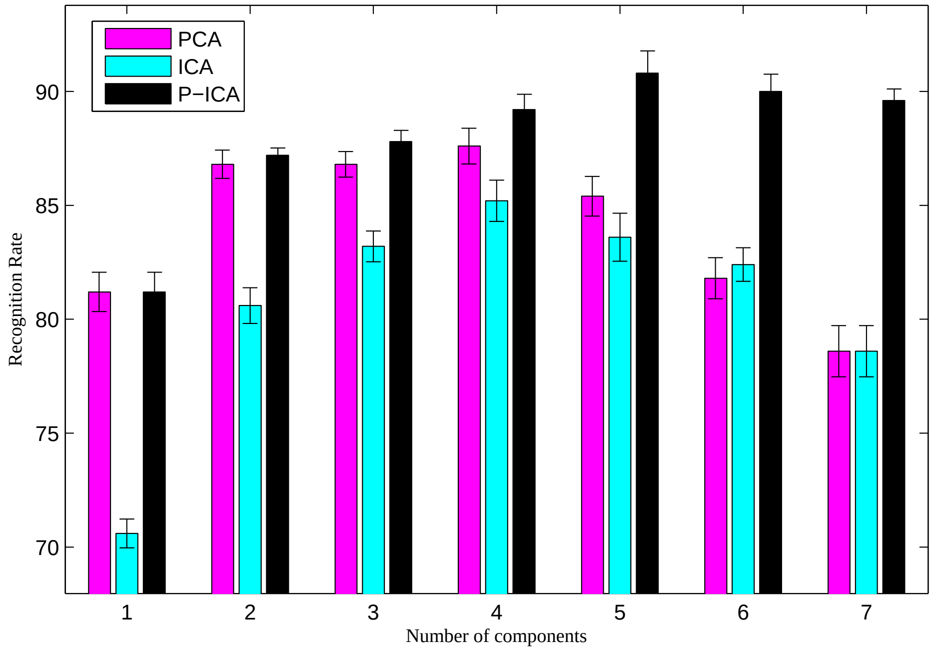

Figure 5. Classification performance of PCA, ICA and P-ICA, as a function of the number of components, when considering feature selection.

4.4 How much are the PCA and ICA contribution rates 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 Figure 6, the averages and standard deviations of 5 repetitions of the PCA and ICA feature contribution rates.

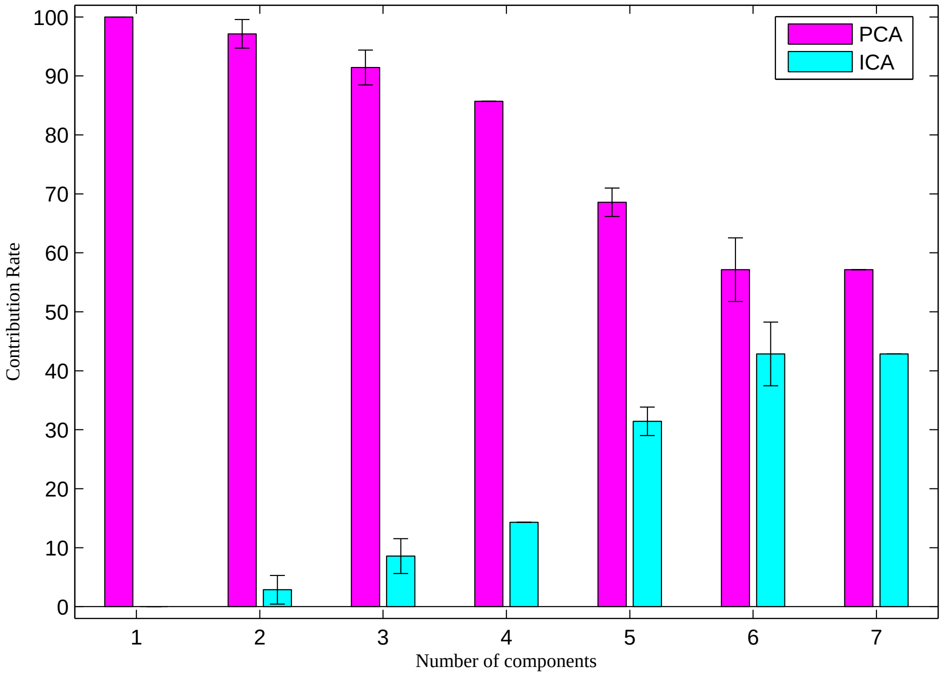

Figure 6. PCA, ICA feature contributions in the P-ICA feature selection process, as a function of the number of selected components.

From Figures 5 and 6, we could make the following observations:

- Without feature selection (number of selected components equals to 7), PCA and ICA perform equally. 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.
- When considering feature selection (number of selected components from 1 to 6), PCA and ICA no longer perform equally. The results show that ICA does not necessarily outperform PCA and vice versa.
- The combination between PCA and ICA features improves advantageously the classification results, for all the used numbers of components. More particularly, when the number of components is high, the recognition rates obtained by our P-ICA algorithm are clearly better than PCA or ICA, used separately.
- The recognition rates are highly dependent on the number of selected components. In fact, by selecting the subset of relevant features, we not only provide a more suitable data modeling (better classification results for the three algorithms), but also we speed up the learning process (less computational effort).
- As the number of selected components increases, the contribution of ICA in the P-ICA feature selection process gets higher, while the one of PCA gets lower.
- The P-ICA best recognition rates (number of selected components from 4 to 7) are obtained when both PCA and ICA contributions are high (none of the methods is dominating).

5. CONCLUSIONS

Feature extraction/selection from a four-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. In fact, the features extracted by each method describe the data from different angles, and thus provide special interpretation of the image contents.

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.

6. ACKNOWLEDGEMENTS

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. Novak and G. 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. Hyvärinen, "Independent component analysis applied to feature extraction from colour and stereo images," *Network: Computation in Neural Systems* **11(3)**, pp. 191–210, 2000.
4. J. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE* **9(8)**, pp. 2009–2026, Oct. 1998. special issue on blind identification and estimation, R.-W. Liu and L. Tong editors.
5. M. Vicente, P. Hoyer, and A. Hyvärinen, "Equivalence of some common linear feature extraction for appearance based object recognition tasks," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **29(5)**, pp. 896–900, May 2007.
6. A. Hyvärinen, "One-unit contrast functions for independent component analysis: a statistical analysis," *Neural Networks for Signal Processing VII (Proc. IEEE NNSP Workshop'97, Amelia Island, Florida)* , pp. 388–397, 1997.
7. A. Hyvärinen, "Fast and robust fixed point algorithms for independent components analysis," *IEEE Trans. Neural Networks* **10(3)**, pp. 626–634, 1999.
8. H. Chaabouni-Chouayakh and M. Datcu, "Linear versus non-linear analysis of relevant scatterers in high resolution sar images," IEEE Geoscience and Remote Sensing Symposium (IGARSS), Barcelona, Spain, July 2007.