PCA-ICA NEURAL NETWORK MODEL FOR POLSAR IMAGES ANALYSIS

Salim Chitroub

Signal and Image Processing Laboratory
Electronics and Informatics Faculty, U. S. T. H. B.
P. O. Box 32, El-Alia, Bab-Ezzouar, 16111, Algiers, ALGERIA
Fax: (213) (21) 24-76-07, email: s_chitroub@hotmail.com

ABSTRACT

The POLSAR images are modeled by a mixture model that results from the product of two independent models, one characterizes the target response and the other characterizes the speckle phenomenon. For the scene interpretation, it is desirable to separate between the target response and the speckle. For this purpose, a PCA-ICA neural network model is proposed. Based on its rigorous statistical formulation, a neuronal approach for the simultaneous diagonalisation of the signal and noise covariance matrices using PCA transform is proposed. The PC images are uncorrelated and having an improved SNR. However, the speckle is a non-Gaussian multiplicative noise, the higher order statistics contain an additional information about it. ICA method is used to separate the speckle from the PC images and providing new IC images that have an improved contrast. The method has been applied on real POLSAR images. The extracted features are quite effective for the scene interpretation.

1. INTRODUCTION

Recent advances in the remote sensing POLarimetric Synthetic Aperture Radar (POLSAR) systems provide a rich set of data for the same scene. This set of data brings knowledge on the nature of targets [1]. The amount of information is scattered in many images that are correlated. Pack information, decorrelate the data and reduce the noise for an efficient scene interpretation are then necessary. However, the POLSAR images are corrupted by speckle that appears as a granular signal-dependent noise. It has the characteristics of a non-Gaussian multiplicative noise [2].

Supposed that the speckle is a non-stationary noise, a generalized form of the PCA is suggested [3]. The method is the noise-adjusted principal components (NAPC) transform [4] adapted to the POLSAR data. According [1], the method appears to give degenerate results in the context of the multiplicative model. Applying PCA in the logarithmic domain of POLSAR images has been suggested in the literature. The speckle is then transformed to an additive noise and hence PCA can be applied. However [4], the loss of information is clearly proved via the reconstitution process of the original POLSAR images. Recently, blind source separation by ICA has received attention because of its potential applications in various domain. What distinguishes ICA from other methods is that it looks for components that are both statistically independent and non-Gaussian by using some form of higher-order statistics, which means information not contained in the covariance matrix. It reduces higher-order statistical dependencies, attempting to make the signals as independent as possible. ICA is suitable for neural network implementation and different theories recently proposed for that purpose lead to the same iterative learning algorithm. Lee et al. [5] review those theories and suggest that information theory can be used to unify several lines of research. Different neural-based blind source separation algorithms are reviewed in [7]-[11].

POLSAR images contain a lot of information in the higher order statistics [2]. This paper demonstrates the usefulness of ICA for POLSAR data analysis. A PCA-ICA neural network model is proposed (Fig. 1). We show that the Independent Component (IC) images present a lower Contrast Ratio (CR) compared to the Principal Component (PC) images. This implies that the speckle is reduced. The IC images have a good Signal-to-Noise Ratio (SNR) and are quite effective for scene interpretation. Before detailing the two parts of the model in sections 3 and 4, respectively, we give in section 2 the POLSAR images model and the statistics to be used later. We conclude the paper in the last section.

2. MODEL AND STATISTICS

Let x_i be the content of the pixel in the ith SAR image, s_i the noise-free signal response of the target, and n_i the speckle. Then, we have the following multiplicative model:

$$x_i = s_i . n_i \qquad (1)$$

By supposing that the speckle has unity mean, standard deviation of σ_i, and is statistically independent from the observed signal x_i [2], the multiplicative model in (1) can be rewritten as:

$$x_i = s_i + s_i.(n_i - 1) \qquad (2)$$

The term $s_i.(n_i - 1)$ represents the zero mean signal-dependent noise and characterizes the speckle noise variation. Thus, we have converted the multiplicative model into the additive model without using any transform operator such as the logarithm operator often suggested in the literature. In fact, applying a logarithm operator tends to depress the information, since the logarithm operation compresses the image dynamic range [4]. The statistics computed in the logarithm domain are different to those computed from the original data. Now, let $\mathbf{x}$ be the stationary random vector of input POLSAR images. Its covariance matrix $\Sigma_{\mathbf{x}}$ can be written as:

$$\Sigma_x = \Sigma_s + \Sigma_n \tag{3}$$

Σ_s and Σ_n are the covariance matrices of the noise-free signal vector and the signal-dependent noise vector, respectively. These two matrices are used in the design process of the transformation matrix of the proposed neural network model.

3. PC IMAGES EXTRACTION

The PCA-based part (Fig. 2) is devoted to extract the PC images. It is based on the simultaneous diagonalisation concept of the two matrices Σ_x and Σ_n, via one orthogonal matrix A [4]. This means that the PC images (vector y) are uncorrelated and have an additive noise that has a unit variance. This step of processing allows us making our application coherent with the theoretical development of ICA. In fact, the constraint to have whitening uncorrelated inputs is desirable in ICA algorithms because it simplifies the computations considerably [5]. These inputs are assumed non-Gaussian, centered, and have unit variance. It is ordinarily assumed that x is zero-mean, which in turn means that y is also zero-mean, where the condition of unit variance can be achieved by standardizing y. For the non-Gaussianity of y, it is clear that the speckle, which has non-Gaussianity properties, is not affected by this step of processing since only the second-order statistics are used to compute A. The criterion "*C*" for determining A is: "*Finding* A *such as the matrix* Σ_n *becomes an identity matrix and the matrix* Σ_x *is transformed, at the same time, to a diagonal matrix*". This criterion can be formulated in the constrained optimization framework as: *Maximize* $A^T.\Sigma_x.A$ *subject to* $A^T.\Sigma_n.A = I$, where I is the identity matrix. Based on the well-developed aspects of the matrix theories and computations, the existence of A is proved in [4] and a statistical algorithm for obtaining it is proposed. Here, we propose a neuronal implementation of this algorithm [6] with some modifications (Fig. 2). It is composed of two PCA neural networks that have a same topology. The lateral weights c_j^1, respectively c_j^2 forming the vector C_1, respectively C_2, connect all the first *m-1* neurons with the *m*th one. These connections play a very important role in the model since they work toward the orthogonalization of the synaptic vector of the *m*th neuron with the vectors of the previous *m-1* neurons. The solid lines denote the weights w_i^1, c_j^1, respectively w_i^2, c_j^2, which are trained at the *m*th stage, while the dashed lines correspond to the weights of the already trained neurons. Note that the lateral weights asymptotically converge to zero, so they do not appear between the already trained neurons.

The first network of Fig. 2 is devoted to whitening the noise in (2), while the second one is for maximizing the variance given that the noise is being already whitened. Let x_1 be the input vector of the first network. The noise is whitened, through the feed-forward weights $\{w_{ij}^1\}$, where *i* and *j* correspond to the input and output neurons, respectively, and the superscript *1* designates the weighted matrix of the first network. After convergence, the vector x is transformed to the new vector x' via the matrix $u = w_1.\Lambda^{-1/2}$, where w_1 is the weighted matrix of the first network, Λ is the diagonal matrix of eigenvalues of Σ_n (variances of the output neurons) and $\Lambda^{-1/2}$ is the inverse of its square root. Next, x' be the input vector of the second network. It is connected to M outputs, with $M \leq N$, corresponding to the intermediate output vector noted x_2, through the feed-forward weights $\{w_{ij}^2\}$ Once this network is converged, the PC images to be extracted (vector y) are obtained such as: $y = A^T.x = u.w_2.x$ where w_2 is the weighted matrix of the second network. The activation of each neuron in the two parts of the network is a linear function of their inputs. The *k*th iteration of the learning algorithm, for both networks, is:

$$\begin{aligned} w(k+1) &= w(k) + \beta(k)\big(q_m(k)P - q_m^2(k)w(k)\big) \\ c(k+1) &= c(k) + \beta(k)\big(q_m(k)Q - q_m^2(k)c(k)\big) \end{aligned} \tag{4}$$

P and Q are, respectively, the input and output vectors of the network. $\beta(k)$ is a positive sequence of learning parameter. The global convergence of the PCA-based part of the model is strongly dependent on the parameter β. The optimal choice of this parameter is well studied in [6].

4. IC IMAGES EXTRACTION

Speckle is non-Gaussian, the higher order statistics of the data contain additional information about it that is not affected by the PCA-based part of the model. To facilitate the scene interpretation, the speckle presence should be reduced to the minimum as much as possible in the PC images, without additional prior knowledge of their statistical properties. This is the purpose of the ICA-based part (Fig. 3). The M inputs of the network are the PC images. The M output neurons correspond to the IC images (vector z). We have then $z = B.y$, where B is the separating (or de-mixing) matrix that we want to determine.

ICA can be carried out by using many different methods. The JADE algorithm [7] is a cumulant-based method that uses joint diagonalization of a set of fourth-order cumulant matrices and exploits their algebraic properties to define a contrast function. Algebraic-based contrast functions typically require extensive batch computations using estimated higher-order statistics. One very straightforward neural learning method is based on the nonlinear PCA learning rule [8]. However, this algorithm is restricted to the separation of sub-Gaussian sources, because of stability requirements. FastICA algorithm is based on a fixed-point iteration and uses a deflation scheme to calculate components sequentially [9]. It has contributed to the application of ICA to large-scale problems due to its computational efficiency. Infomax algorithm maximizes the joint entropy of the components that is an invertible monotonic no-linearity function of the outputs [5], [10]. It has a simple architecture and can deal with either sub-Gaussian or super-Gaussian components by adaptively switching between two non-linearities. The switching is possible by using the stability analysis given in [11]. In this paper, we have adapted this algorithm to learn the matrix B.

Using the concept of differential entropy and the invertible transformation of $z = B.y$, the mutual information between the outputs is: $I(z) = \sum_{i=1}^{M} H(z_i) - H(y) + log(|det\, B|)$, where $H(z_i)$ are the marginal entropies of the outputs and $H(z)$ is the joint entropy of z. By constraining z_i to be uncorrelated and of unit variance, this implies that: $det\, E(z.z^T) = 1$. As the negentropy is a measure of non-Gaussianity, that is: $J(z) = H(z_{Gaussian}) - H(z)$. So the mutual information and negentropy differ only by a constant that does not depend on B and the sign, that is: $I(z) = C - \sum_{i=1} z$, which means that finding an invertible transformation B that minimizes the mutual information is approximately equivalent to finding directions in which the sum of non-Gaussianities of z_i is maximized. Maximizing the joint entropy $H(z)$ can approximately minimize the mutual information among the output components $z_i = g_i(v_i)$, where $g_i(v_i)$ is an invertible monotonic non-linearity and $v = B.y$. If the mutual information among the outputs is zero, the mutual information before the non-linearity must be zero as well since the nonlinear transfer function does not introduce any dependencies. The relation between z_i, v_i, and $g_i(v_i)$ is such as: $p(z_i) = p(v_i)/|\partial g_i(v_i)/\partial v_i|$. By this relationship, $g(v)$ must be chosen so that its derivative approximately forms a probability distribution function for the sources to be recovered. The only remaining parameters to adapt are the synaptic weights that can be found by maximizing $H(z)$ with respect to B. The weight update rule will then be a gradient descent in the direction of maximum joint entropy. More computationally efficient approaches have been proposed in [11]. If we define the term score function $\varphi(v)$ as: $\varphi(v) = (\partial p(v)/\partial v)/p(v)$, then an efficient weight update is:

$$\Delta B \propto (I - \varphi(v)v^T)B \tag{5}$$

The form of $\varphi(v)$ plays a crucial role because it is function of the transfer and therefore a function of the source estimate. For the sub-Gaussian sources, the form of $\varphi(v)$ is such as: $\varphi(v) = v - tanh(v)$, where $tanh(.)$ is the hyperbolic tangent. For the super-Gaussian sources, $\varphi(v)$ takes the form: $\varphi(v) = v + tanh(v)$. The switching between the sub-Gaussian and super-Gaussian learning rule is [10]:

$$\Delta B \propto (I - K.tanh(v)v^T - v.v^T)B \tag{6}$$

K is a N-dimensional diagonal matrix with elements $sign(k_4(v_i))$. $k_4(v_i)$ is the kurtosis of the source estimate v_i. The switching parameter $k_4(v_i)$ can be derived from the general stability analysis of separating solutions [9], [11].

5. EXPERIMENTAL RESULTS

A real POLSAR data provided by the SIR-C system [4] are used to evaluate the proposed model. The data were acquired over the Orgeval site (east of Paris, France, 329x329pixels) during summer 1994 and correspond to bands C and L with HH and HV polarizations for each. The four bands are shown in Fig. 4. The extracted PC images are given in Fig. 5. Most of the information contained in the original images is now concentrated in the first PC image, which is an image of quality. Second and third PC images contain mainly noise more than information. The fourth PC image is very noisy and no information can be extracted from it. The SNR values of the original and PC images are given in Table 1. The original images have ratios ranging from 4.20 to 18.84. While the SNR value in the first PC image is improved to 26.62; this corresponds to a factor of 1.41 compared with the best original image (L band-HV). Note that the gray levels in the first PC image are reversed compared to the L band-HV. We have used as input for the ICA-based part only the first PC image. The obtained IC image is shown in Fig. 6, which is an image of very high quality and better contrasted than the first PC image. In fact, when all PC images are used as input vector, the results are not significative. This can be justified by the fact that the Infomax algorithm is efficient only in the case where the input data have low noise [10], [5]. We quantify the speckle level by computing the CR value, which is the average value of the standard deviation to mean ratios calculated in small homogeneous areas of the observed scene. The speckle reduction is quite evident when comparing the IC image with the first PC image. The first PC image has a CR of 0.58, while the IC image reduces the speckle level to 0.39.

6. CONCLUSION

The particularity of the suggested model for POLSAR image analysis lies in the exploitation of the proper advantages of both PCA and ICA. PCA is used to provide PC images that are uncorrelated and having an improved SNR. The dimensionality reduction can be made at this stage by retaining only the first PC image. The speckle information as well as the mutual independence with respect to the higher-order statistics are treated by ICA. The speckle is separated from the first PC image and a new IC image that has an improved contrast is provided. Thus, the IC image is quite effective for scene interpretation. The method is more general than PCA or ICA approaches separately used, and it is a powerful tool for interpreting and analyzing the complex scene acquired by POLSAR sensors.

REFERENCES

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Artech-House, London, 1998.

[2] S. Chitroub, A. Houacine, and B. Sansal, "Statistical characterisation and modelling of SAR images", *Signal Processing,* Vol. 82, No. 1, pp. 69-92, February 2002.

[3] J. S. Lee and K. Hoppel, "Principal component transformation of multifrequency polarimetric SAR imagery", *IEEE Transactions on Geoscience & Remote Sensing,* vol. 30, no. 4, PP. 686-696, July 1994.

[4] S. Chitroub, A. Houacine, and B. Sansal, "A New PCA-based method for data compression and enhancement of multi-

frequency polarimetric SAR imagery", *Intelligent Data Analysis, International Journal*, Vol. 6, No. 2, 2002.

[5] T. W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "A Unifying information-theoretic framework for independent component analysis", *Neural Networks*, Vol. 39, pp. 1-21, 2000.

[6] S. Chitroub, A. Houacine, and B. Sansal, "Neuronal principal component analysis for an optimal representation of multispectral images", *Intelligent Data Analysis, International Journal*, Vol. 5, No. 5, pp. 385-403, 2001.

[7] J. F. Cardoso, "High-order contrasts for independent component analysis", *Neural Computation*, Vol. 11 No. 1, pp. 157-192, 1999.

[8] J. Karhunen and J. Joutsensalo, "Representation and separation of signals using nonlinear PCA type learning", *Neural Networks*, Vol. 7, pp. 113-127, 1994.

[9] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis", *IEEE Transactions on Neural Networks*, Vol. 10, No. 3, pp. 626-634, May 1999.

[10] T. W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources", *Neural Computation*, Vol. 11, pp. 417-441, 1999.

[11] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation", *IEEE Transactions on Signal Processing*, Vol. 45, No. 2, pp. 434-444, 1996.

[12] T. Cover and J. Thomas, *Elements of Information Theory*, Vol. 1, John Wiley & Sons, New York, 1991.

[13] S. Amari, T. P. Chen , and A. Cichocki, "Stability analysis of adaptive blind source separation", *Neural Networks*, Vol. 10, No. 8, pp. 1345-1352, 1997.

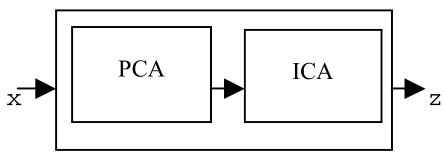


Fig. 1. PCA-ICA neural network model.

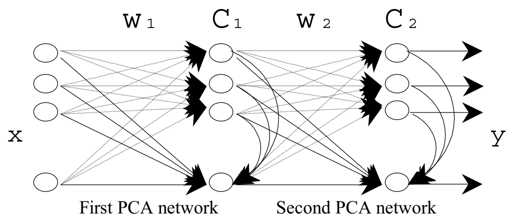


Fig.2. PCA-based part of the model.

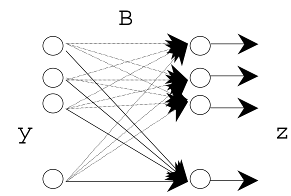


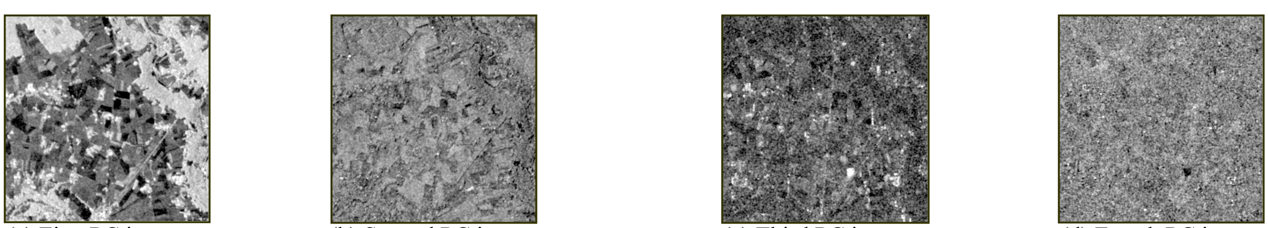
Fig. 3. ICA-based part of the model.

Table 1. Signal-to-Noise Ration (SNR) values in the original POLSAR images and PC images.

Images	C Band-HH	C Band-HV	L Band-HH	L Band-HV	First PC	Second PC	Third PC	Fourth PC
SNR values	4.205	8.529	12.575	**18.840**	**26.620**	3.978	3.247	2.022

(a) C Band-HH (b) C Band-HV (c) L Band-HH (d) L-Band-HV

Fig. 4. The four original POLSAR images.



(a) First PC image (b) Second PC image (c) Third PC image (d) Fourth PC image

Fig. 5. The four extracted principal component (PC) images using the PCA-based part of the model.

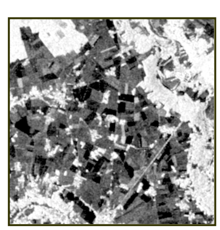


Fig. 6. The extracted independent component (IC) image using the ICA-based part of the model.