CFAR HIERARCHICAL CLUSTERING OF POLARIMETRIC SAR DATA

P. Formont[1], M.A. Veganzones[2], J.M. Frontera-Pons[1], F. Pascal[1], J.-P. Ovarlez[1,3], J. Chanussot[2,4]

[1] SONDRA, Supélec, Gif-sur-Yvette, France
[2] GIPSA-lab, Grenoble-INP, Saint Martin d'Hères, France
[3] DEMR/TSI, ONERA, Palaiseau, France
[4] Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland

ABSTRACT

Recently, a general approach for high-resolution polarimetric SAR (POLSAR) data classification in heterogeneous clutter was presented, based on a statistical test of equality of covariance matrices. Here, we extend that approach by taking advantage of the Constant False Alarm Ratio (CFAR) property of the statistical test in order to improve the clustering process. We show that the CFAR property can be used in the hierarchical segmentation of the POLSAR data images to automatically detect the number of clusters. The proposed method will be applied on a high-resolution polarimetric data set acquired by the ONERA RAMSES system.

1. INTRODUCTION

In [1], authors propose a general approach for high-resolution polarimetric SAR (POLSAR) data classification in heterogeneous clutter, based on a statistical test of equality of covariance matrices. The proposed method generalizes several distance measures used in standard classification methods and can be applied to both homogeneous and heterogeneous clutter models without any a priori physical interpretation to the classification process.

The heterogeneous clutter is described by a Spherically Invariant Random Vector (SIRV) model. The Fixed Point (FP) estimate [2] of the covariance matrix is used to describe the POLSAR data. The FP estimate is independent of the texture probability density function (pdf) and is an Asymptotically Maximum Likelihood (AML) estimator for many stochastic processes obeying the SIRV model. Moreover, it is asymptotically Wishart distributed.

A statistical test of equality of covariance matrices is introduced. The distribution of the test statistic can be approximated by a χ^2 distribution, giving raise to a Constant False Alarm Ratio (CFAR) property. Such an approach provides a threshold over which pixels are rejected from the image, meaning they are not sufficiently "close" from any existing class. We take advantage of this CFAR property to modify the Hierarchical Clustering (HC) algorithm. The CFAR property allows us to decide the optimal number of clusters in the HC by setting a False Alarm threshold that forces the algorithm to stop at some level of the hierarchy. The proposed algorithm is tested on a high-resolution polarimetric data set acquired by the ONERA RAMSES system.

Section 2 reviews the statistical test of equality of covariance matrices. Section 3 introduces the CFAR hierarchical clustering algorithm. Section 4 describes the experimental design and results. Finally, in Section 5 we give some conclusions.

2. TEST OF EQUALITY OF COVARIANCE MATRICES

This section is devoted to the study of statistical test for equality of covariance matrices and its application to POLSAR data classification. A statistical approach to the problem of POLSAR data classification has many advantages:

- It can be applied to both homogeneous and heteroge-

This work was supported by the Direction Générale de l'Armement (DGA) under contract PDE2012600079.

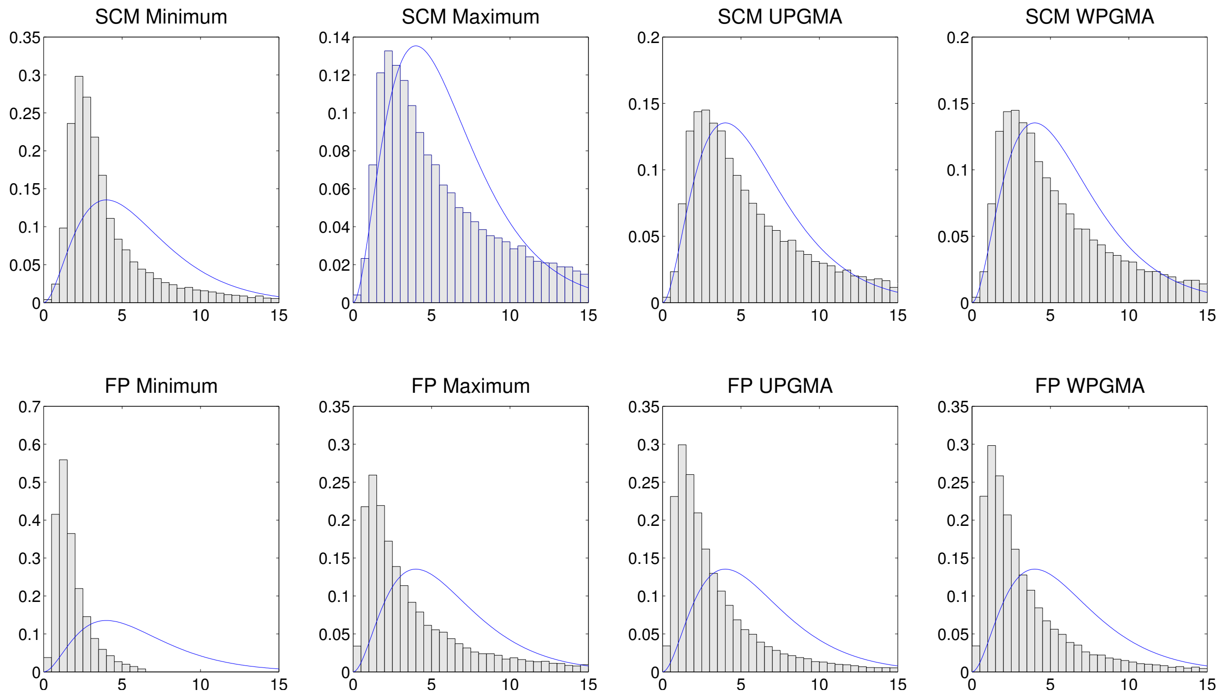


Fig. 1. Comparison of the non-parametric distribution estimation of the linkage distances respect to the actual χ^2 distribution with six degrees of freedom.

neous clutter models.

- There is no a priori physical interpretation to the classification process.

First, a Gaussian context will be assumed and later, an approach under SIRV assumption will be derived.

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent random vectors, such that $\mathbf{X}_1 \sim \mathcal{L}(\mathbf{0}, \mathbf{T}_1)$ and $\mathbf{X}_2 \sim \mathcal{L}(\mathbf{0}, \mathbf{T}_2)$ where $\mathcal{L}$ stands for any distribution with the two first moments existing. The goal is to decide if their covariance matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ are equal. The resulting binary hypothesis test can be written as:

$$\begin{cases} H_0 : \mathbf{T}_1 = \mathbf{T}_2 \\ H_1 : \mathbf{T}_1 \neq \mathbf{T}_2 \end{cases}$$

where $\mathbf{x}_1$ (resp. $\mathbf{x}_2$) is a sample of N_1 (resp. N_2) independent and identically distributed m-dimensional random vectors, and $N = N_1 + N_2$. Under Gaussian assumption, the test statistic can easily be derived as the following equation [1]:

$$\lambda = \frac{\left|\widehat{\mathbf{T}}_1\right|^{\frac{N_1}{2}} \left|\widehat{\mathbf{T}}_2\right|^{\frac{N_2}{2}}}{\left|\widehat{\mathbf{T}}\right|^{\frac{N}{2}}} \tag{1}$$

Notice the exponents are the size of the samples. Bartlett [3] proposed alternative exponents for the univariate case, replacing the samples size by the degree of freedom of the estimators $\widehat{\mathbf{T}}_i$. Eq. (1) then becomes:

$$t = \frac{|\widehat{\mathbf{T}}_1|^{\frac{\nu_1}{2}} |\widehat{\mathbf{T}}_2|^{\frac{\nu_2}{2}}}{|\widehat{\mathbf{T}}_t|^{\frac{\nu_t}{2}}} \tag{2}$$

where ν_i are the degrees of freedom of $\widehat{\mathbf{T}}_i$ and $\nu_t = N$, the degree of freedom of $\widehat{\mathbf{T}}$. As the POLSAR data is zero-mean, the degrees of freedom are equal to the size of the samples.

Box [4, 5] proposed a χ^2 approximation for the distribution of t. The statistic he proposed is:

$$u = -2(1 - c_1)\ln(t) \sim \chi^2\left(\frac{1}{2}m(m+1)\right) \tag{3}$$

where c_1 depends on some parameters of the samples and $\chi^2(a)$ denotes the χ^2 distribution with a degrees of freedom and m the size of the data vector, i.e. 3 for POLSAR data. The critical region of the test is

$$\left\{ u > \chi^2\left(\frac{1}{2}(k-1)m(m+1), P_{\text{FA}}\right) \right\} \tag{4}$$

where P_{FA} is the type I error, or false-alarm rate. In our case, we set the type I error (accepting the null hypothesis when it is not true i.e. rejecting the equality of the matrices of the two populations when they are actually equal) to be very low (typically 10^{-3} or 10^{-4}).

The same procedure can be applied in the SIRV case, considering that it was demonstrated by Pascal et al. in [6] that the FP estimate behaves asymptotically as a Wishart-distributed matrix with $\frac{m}{m+1}N$ degrees of freedom. This is a very important property of the FP estimate since all results obtained with the SCM remain valid for the FP estimate due to an asymptotical justification. The SCM estimators $\widehat{\mathbf{T}}_1$ and $\widehat{\mathbf{T}}_2$ can be replaced by the FP estimates $\widehat{\mathbf{M}}_1$ and $\widehat{\mathbf{M}}_2$ of the same N-samples $\mathbf{x}_1$ and $\mathbf{x}_2$ with the correct degrees of freedom.

3. CFAR HIERARCHICAL CLUSTERING

Hierarchical clustering [7] returns a hierarchy of clusters built by merging smaller components into bigger clusters (agglomerative clustering) or by splitting the whole image into smaller regions (divisive clustering). In order to do that, a similarity function between any two components is necessary, i.e. the Euclidean distance, to compare each pair of components. Then, a linkage function indicates which of any possible pair of components is merged into (or split from) a bigger cluster, i.e. the pair with minimum similarity. The result is then a hierarchy of clusters that can be represented as a dendrogram. Two components merging into (splitting from) any cluster, or node, in the dendrogram are linked by a similarity value. A common criterion to stop merging (splitting) is to a priori set the number k of clusters one is looking for.

In our case, we define the pairwise similarity function between any two components as the results of the test of equality of covariance matrices (3), u. The CFAR property of (4) allows us to define a stopping criterion that automatically detects the number of clusters for a given False Alarm probability, P_{FA}. We simply stop merging (splitting) once the similarity function, u, goes over (under) the threshold value, u_{FA}, corresponding to equation (4) for a given P_{FA}.

4. EXPERIMENTS AND RESULTS

4.1. Dataset

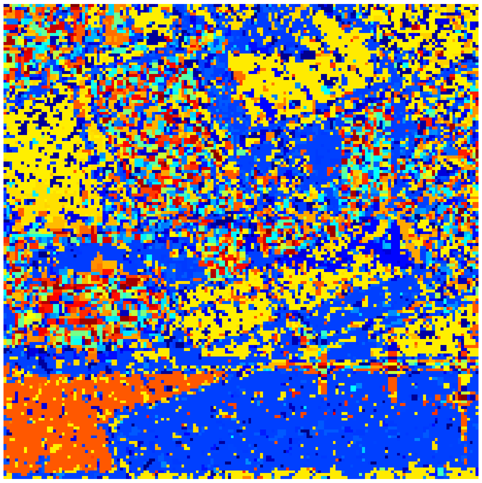
The POLSAR dataset was acquired by the ONERA RAMSES system in Brétigny, France. The acquisition was made in X-band, with a spatial resolution of 1.32m in range and 1.38m in azimuth. The resulting image is 501×501 pixels. Fig.2 shows the power representation image of the dataset.

Fig. 2. Power representation image of the POLSAR dataset.

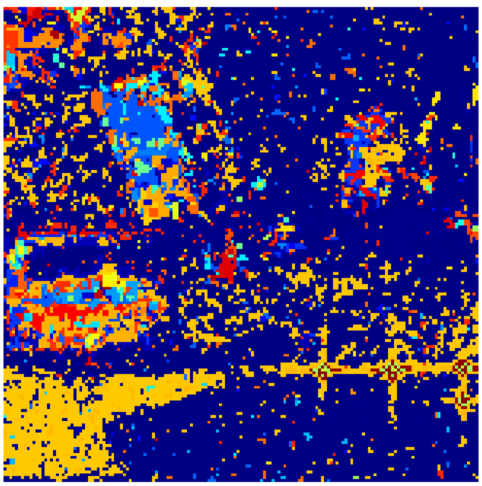
4.2. Hierarchical clustering results

In order to build the hierarchical clustering of the POLSAR dataset, we first downsample the image by taking one of each three pixels, in order to reduce the computational burden. We estimate the covariance matrices for each pixel using a 7×7 window by the sample covariance matrix (SCM) and the Fixed Point (FP) estimators. Fig.1 shows the distribution of the pairwise distances between the nodes of the hierarchical clustering, given by four different linkage functions: Minimum, Maximum, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Weighted PGMA (WPGMA)[1]. The linkage distances distribution for the SCM fits better the six degrees of freedom χ^2 distribution than the ones obtained using the FP estimator. This was expected as the χ^2 distribution assumes Gaussianity and the Wishart approximation for the FP is only asymptotical. However, the FP estimation is close enough to assume the χ^2 distribution as the basis for the

[1] http://www.mathworks.fr/fr/help/stats/linkage.html



(a)



(b)

Fig. 3. Hierarchical clustering of the POLSAR data using the average linkage function and setting $P_{\mathrm{FA}} = 10^{-4}$. Covariance matrix estimators are: (a) SCM and (b) FP.

CFAR property. The minimum linkage function fits poorly in both cases.

Fig.3 shows the hierarchical clustering maps of the POLSAR dataset using the UPGMA linkage function for $P_{\mathrm{FA}} = 10^{-4}$. It can be observed that the SCM-based map presents some oversegmentation due to the heavy tail observed in its linkage distances distribution (see Fig.1). In the FP-based map, it can be observed that the pixels are grouped in differentiated areas comprising buildings, fields, roads, etc. These preliminary results support the CFAR hierarchical clustering of POLSAR data images.

5. CONCLUSIONS AND FURTHER WORK

In this work, we have used the statistical test of equality of covariance matrices, defined for high resolution POLSAR images classification, to propose a CFAR hierarchical clustering method. The proposed method takes advantage of the statistical test CFAR property to automatically detect the correct number of clusters for a given probability of false alarm. The preliminary experimental results over a real dataset encourages us to follow this research avenue. Further work will focus on extend the benefits of the CFAR property to other clustering algorithms such as the k-Means clustering algoritm, and to the statistical classification of POLSAR data.

6. REFERENCES

[1] P. Formont, F. Pascal, G. Vasile, J.-P. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric sar images," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 567 –576, june 2011.

[2] F. Pascal, Y. Chitour, J-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis," *Signal Processing, IEEE Transactions on*, vol. 56, no. 1, pp. 34 –48, jan. 2008.

[3] M S Bartlett, "Properties of sufficiency and statistical tests," *Proceedings of the Royal Society of London*, vol. 160, pp. 268–282, 1937.

[4] G.E.P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, 1949.

[5] G.E.P. Box, "Problems in the analysis of growth and linear curves," *Biometrika*, vol. 6, pp. 362–389, 1950.

[6] Frédéric Pascal, Jean-Philippe Ovarlez, Philippe Forster, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *Signal Processing*, vol. 56, no. 6, pp. 2206–2216, 2008.

[7] Joe H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.