

Estimators of the Entropy for SAR Image Classification

Julia Cassetti, Daiana Delgadino, and Alejandro C. Frery, *Senior Member*

Abstract—Remotely sensed data are successfully used for information extraction. In particular SAR imagery, which suffer the presence of speckle noise, needs special models and techniques. In this sense, the \mathcal{G}^0 family of distributions is a suitable model for SAR intensity because it can characterize areas with different degrees of texture. Information theory has gained a place in signal and image processing for parameter estimation and feature extraction. Among its tools, the entropy stands out as one of the most expressive features. In this paper, we evaluate the performance of several parametric and non parametric Shannon entropy estimators as input for supervised and unsupervised classification algorithms for SAR images. These estimators were analyzed through Kappa, overall reliability and accuracy indexes. Finally we apply these estimators to actual data.

Index Terms—Feature extraction, synthetic aperture radar, Shannon entropy, classification.

I. INTRODUCTION

IMAGES that come from coherent illumination systems, such as those acquired by synthetic aperture radar (SAR) are contaminated by speckle noise. This kind of noise corrupts the image making it difficult to analyze and interpret.

Under this perspective, statistical procedures are important tools for processing SAR data. The choice of a suitable model and appropriate measures to describe this sort of images is a fundamental point to obtain features that promote a good analysis.

In this sense, the family of distributions \mathcal{G}^0 [9] has been extensively used to model SAR data because of its ability to a wide variety of roughness targets.

Several approaches have been developed in order to obtain expressive and tractable features. In particular, entropy measures have been widely used for this purpose. Estimation parameter [10], classification [3], methodologies for constructing confidence interval and contrast measures [8], [14], and edge detection [15] are some examples of its application.

Sundry authors have tackled the segmentation and classification SAR images problem using information theory measures. Nobre et al. [16] used Rényi's entropy for monopolarized SAR image segmentation. Ferreira et al. [7] derived a closed-form expression for the Shannon entropy based on the \mathcal{G}^0 law for intensity data, and proposed a new entropy-based segmentation method. Carvalho et al. [3] employed stochastic distances to approach unsupervised classification methodology applied to Polarimetric Synthetic Aperture Radar (PolSAR) images.

Julia Cassetti and Daiana Delgadino julia.cassetti@gmail.com are with the Instituto de Desarrollo Humano, Universidad Nacional de General Sarmiento, Pcia. de Buenos Aires, Argentina.

Alejandro C. Frery alejandro.frery@vuw.ac.nz is with the School of Mathematics and Statistics, Victoria University of Wellington, New Zealand

The Shannon entropy has been applied to analyzed SAR imagery in several approaches, from inference [8] to classification [7]. Therefore, its estimation deserves attention. Vasicek [21] replaced the distribution function F by the empirical distribution function F_n and used a difference operator in place of the differential operator in the Shannon's entropy expression. Van Es [20] studied an entropy estimator based on differences between order statistics. Correa [5] proposed a new entropy estimator based on local linear regression. Al-Omari [2] and Noughabi and Noughabi [17] presented modified versions of the Ebrahimi et al. [6] estimator.

In this paper, we address the classification problem through supervised and unsupervised strategies whose input is the local estimate of the entropy. To this aim, we will evaluate estimators of the entropy, both parametric and non-parametric. In the parametric case we used the relationship between the \mathcal{G}^0 and Fisher distributions to obtain an expression of the entropy. In the non parametric case, we will evaluate these estimators in terms of bias, mean square error, computational time and accuracy when classifying a monopolarimetric SAR image.

II. THE \mathcal{G}^0 MODEL

The multiplicative model defines the return Z in a monopolarized SAR image as the product of two independent random variables: one corresponding to the backscatter X , and the other to the speckle noise Y . In this manner $Z = XY$ represents the return Z in each pixel under.

The \mathcal{G}^0 family of distributions is attractive because of its flexibility to model adequately areas with all types of roughness [12], [13]. For intensity SAR data, this family arises from considering the speckle noise Y modeled as a Γ distributed random variable with unitary mean and shape parameter $L \geq 1$, the number of looks, while the backscatter X is considered to obey a Reciprocal of Gamma law. The density function for intensity data is given by

$$f(z) = \frac{L^L \Gamma(L - \alpha)}{\gamma^\alpha \Gamma(-\alpha) \Gamma(L)} \cdot \frac{z^{L-1}}{(\gamma + zL)^{L-\alpha}}, \quad (1)$$

where $-\alpha, \gamma, z > 0$ and $L \geq 1$. The r -order moments are

$$E(Z^r) = \left(\frac{\gamma}{L}\right)^r \frac{\Gamma(-\alpha - r)}{\Gamma(-\alpha)} \frac{\Gamma(L + r)}{\Gamma(L)}, \quad (2)$$

provided $\alpha < -r$, and infinite otherwise.

Mejail et al. [12] proved a relationship between \mathcal{G}^0 distributions and the Fisher-Snedecor F law: the cumulative distribution function $F_{\alpha, \gamma, L}$ for the return Z is

$$F_{\alpha, \gamma, L}(z) = \Upsilon_{2L, -2\alpha}(-\alpha z / \gamma), \quad (3)$$

for every $z > 0$, where $\Upsilon_{2L, -2\alpha}$ is the cumulative distribution function of a Fisher-Snedekor random variable with $2L$ and -2α degrees of freedom. This relationship is useful for obtaining an entropy expression.

III. SHANNON ENTROPY

It is well known Shannon's contribution to the creation of what is known Information Theory. Shannon [19] proposed a new way of measuring the transmission of information through a channel, thinking of information as a statistical concept. In this regard Shanon's entropy (SE) is defined as

$$H[f(x)] = -E[\log f(x)] = -\int_{-\infty}^{\infty} f(x) \log f(x) dx \quad (4)$$

$$= \int_0^1 \log \frac{dF^{-1}(x)}{dx} dx, \quad (5)$$

where X is a continuous random variable with probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$.

The entropy of the \mathcal{G}^0 distribution is obtained using (3). Denote H_F to entropy under the F model, then the \mathcal{G}^0 entropy for intensity data $H_{\mathcal{G}^0}$ is

$$H_{\mathcal{G}^0}(\alpha, \gamma, L) = H_F(2L, -2\alpha) - \log\left(-\frac{\alpha}{\gamma}\right). \quad (6)$$

Using (6), the expression of $H_{\mathcal{G}^0}$ is

$$\begin{aligned} H_{\mathcal{G}^0}(\alpha, \gamma, L) = & -\log\left(-\frac{\alpha}{\gamma}\right) - (1-\alpha)\psi^{(0)}(-\alpha) \\ & + \log\left(-\frac{\alpha}{L}\right) + \frac{1}{2}(2L-2\alpha)\psi^{(0)}\left(\frac{1}{2}(2L-2\alpha)\right) \\ & + \log(B(L, -\alpha)) + (1-L)\psi^{(0)}(L), \end{aligned} \quad (7)$$

where $\psi^{(0)}$ and B are the digamma and beta functions, respectively.

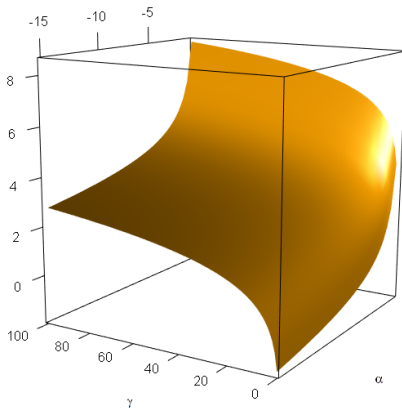


Fig. 1. $H_{\mathcal{G}^0}(\alpha, \gamma, L)$ as a function of alpha and gamma for $L=3$.

Figure 1 shows the $H_{\mathcal{G}^0}(\alpha, \gamma, L)$ theoretical entropy as a function of α and γ and $L = 3$. It can be shown that, for each fixed γ value, $H_{\mathcal{G}^0}$ is an injective function. The same behavior happens with $H_{\mathcal{G}^0}$ when α is a fixed value.

IV. SHANNON ENTROPY ESTIMATORS

Several authors have proposed entropy estimators according to (4). Most of them are based on order statistics of the sample.

Al-Omari [2] presented an overview of these estimators and also proposed a new one. From a parametric point of view, is natural to consider the maximum likelihood estimator (MV) of the entropy (HML).

In this paper we study the following entropy estimators.

A. Maximum likelihood entropy estimator

The optimal asymptotic properties of the MV estimator are well-known. Let Z_1, \dots, Z_n be an independent random sample of size n from the $\mathcal{G}^0(\alpha, \gamma, L)$ distribution. Assume L known. A maximum likelihood estimator of α and γ for L known, denoted $\hat{\alpha}_{ML}$ and $\hat{\gamma}_{ML}$ respectively, is the value (in the parametric space $\mathbb{R}_- \times \mathbb{R}_+$), that maximizes the loglikelihood function:

$$\begin{aligned} \log \Gamma(L - \hat{\alpha}_{ML}) - \hat{\alpha}_{ML} \log \hat{\gamma}_{ML} - \log \Gamma(-\hat{\alpha}_{ML}) \\ + \frac{\hat{\alpha}_{ML} - L}{n} \sum_{i=1}^n \log(\hat{\gamma}_{ML} + LZ_i). \end{aligned} \quad (8)$$

Its calculation demands numerical maximization routines; we used the L-BFGS-B version of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [11] that allows box constraints. This algorithm belongs to the family of quasi-Newton methods. The characteristic of these methods is that they do not require the Hessian matrix, only the gradient.

The ML entropy estimator (H_{ML}) [4] is, thus,

$$H_{ML} = H_{\mathcal{G}^0}(\hat{\alpha}_{MV}, \hat{\gamma}_{MV}, L). \quad (9)$$

B. Non parametric entropy estimators

We recall the nonparametric entropy estimators studied in [2]. In all of them we consider that X_1, \dots, X_n is a random sample from the distribution function $F(x)$, and that $X_{(i)}$ its order statistics, $1 \leq i \leq n$.

- Vasecek [21]:

$$H_{V(m,n)} = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right]. \quad (10)$$

- Van Es [20]:

$$\begin{aligned} H_{VE(m,n)} = & \frac{1}{n-m} \sum_{i=1}^{n-m} \left[\frac{n+1}{m} (X_{(i+m)} - X_{(i)}) \right] \\ & + \sum_{k=m}^n \frac{1}{k} + \log \frac{m}{n+1}. \end{aligned} \quad (11)$$

- Correa [5]:

$$H_{C(m,n)} = -\frac{1}{n} \sum_{i=1}^n \log \frac{\sum_{j=i-m}^{i+m} (j-i) (X_{(j)} - \bar{X}_{(i)})}{n \sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})^2}, \quad (12)$$

where $\bar{X}_{(i)} = (2m+1)^{-1} \sum_{j=i-m}^{i+m} X_{(j)}$.

- Al-Omari [1]:

$$H_{AO_1(m,n)} = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{\omega_i m} (X_{(i+m)} - X_{(i-m)}) \right], \quad (13)$$

where

$$\omega_i = \begin{cases} 1 + \frac{1}{2} & \text{if } 1 \leq i \leq m, \\ 2 & \text{if } m+1 \leq i \leq n-m, \\ 1 + \frac{1}{2} & \text{if } n-m+1 \leq i \leq n, \end{cases}$$

in which $X_{(i-m)} = X_{(1)}$ for $i \leq m$, and $X_{(i+m)} = X_{(n)}$ for $i \geq n-m$.

- Al-Omari alternative proposal [2]:

$$H_{AO_2(m,n)} = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{v_i m} (X_{(i+m)} - X_{(i-m)}) \right], \quad (14)$$

where

$$v_i = \begin{cases} 1 + \frac{i-1}{m} & \text{if } 1 \leq i \leq m, \\ 2 & \text{if } m+1 \leq i \leq n-m, \\ 1 + \frac{n-i}{2m} & \text{if } n-m+1 \leq i \leq n, \end{cases}$$

in which $X_{(i-m)} = X_{(1)}$ for $i \leq m$, and $X_{(i+m)} = X_{(n)}$ for $i \geq n-m$.

- Ebrahimi [6]

$$H_{E(m,n)} = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{\tau_i m} (X_{(i+m)} - X_{(i-m)}) \right], \quad (15)$$

where

$$\tau_i = \begin{cases} 1 + \frac{i-1}{m} & \text{if } 1 \leq i \leq m, \\ 2 & \text{if } m+1 \leq i \leq n-m, \\ 1 + \frac{n-i}{m} & \text{if } n-m+1 \leq i \leq n. \end{cases}$$

V. SIMULATION STUDY

Since most of the estimators depend on the spacing m , we first conducted a Monte Carlo study to choose the value that yields smallest bias and mean squared error (MSE).

For this study we generated 500 samples from a \mathcal{G}^0 distribution of size $n \in \{9, 25, 49, 81\}$ for each roughness value $\alpha \in \{-8, -5, -3, -1.5\}$, each non parametric estimator presented in Section IV-B, $m \leq n/2$ and also H_{ML} . The sample sizes represent different scenarios of window sizes.

We then assessed these estimators in the frame of supervised classification by applying SVM, and of unsupervised classification using k-means. We generated simulated images for several values of α and γ , and used sliding windows of different sizes to generate the entropy map. The classification was performed using these maps as features. The classification methods are evaluated using Kappa and overall accuracy. The parameter estimation methods were tested in an actual SAR image.

The details and results will be presented in the final paper.

VI. COMPUTATIONAL INFORMATION

All studies were made in the R platform and language for statistical computing [18] (version 4.0.4).

REFERENCES

- [1] A. I. Al-Omari. Estimation of entropy using random sampling. *Journal of Computational and Applied Mathematics*, 261:95–102, 2014.
- [2] A. I. Al-Omari. A new measure of entropy of continuous random variable. *Journal of Statistical Theory and Practice*, 10(4):721–735, 2016.
- [3] N. Carvalho, L. Sant’Anna Bins, and S. J. Sant’Anna. Analysis of stochastic distances and Wishart mixture models applied on PolSAR images. *Remote Sensing*, 11(24), 2019.
- [4] G. Casella and R. Berger. *Statistical Inference*. Duxbury Resource Center, June 2001.
- [5] J. C. Correa. A new estimator of entropy. *Communications in Statistics – Theory and Methods*, 24(10):2439–2449, 1995.
- [6] N. Ebrahimi, K. Pflughoeft, and E. S. Soofi. Two measures of sample entropy. *Statistics & Probability Letters*, 20(3):225–234, 1994.
- [7] J. Ferreira and A. D. Nascimento. Shannon entropy for the \mathcal{G}_f^0 model: A new segmentation approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2547–2553, 2020.
- [8] A. C. Frery, R. J. Cintra, and A. D. Nascimento. Entropy-based statistical analysis of PolSAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3733–3743, 2013.
- [9] A. C. Frery, H. J. Müller, C. C. Yanasse, and S. J. Sant’Anna. A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):648–659, 1997.
- [10] J. Gambini, J. Cassetti, M. Lucini, and A. Frery. Parameter estimation in SAR imagery using stochastic distances and asymmetric kernel. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8:365–375, 2015.
- [11] D. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer Science+Business Media, LLC, 2008.
- [12] M. Mejail, J. C. Jacobo-Berlles, A. C. Frery, and O. H. Bustos. Classification of SAR images using a general and tractable multiplicative model. *International Journal of Remote Sensing*, 24(18):3565–3582, 2003.
- [13] M. E. Mejail, A. C. Frery, J. Jacobo-Berlles, and O. H. Bustos. Approximation of distributions for SAR images: Proposal, evaluation and practical consequences. *Latin American Applied Research*, 31:83–92, 2001.
- [14] A. D. Nascimento, R. J. Cintra, and A. C. Frery. Hypothesis testing in speckled data with stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):373–385, 2010.
- [15] A. D. Nascimento, M. M. Horta, A. C. Frery, and R. J. Cintra. Comparing edge detection methods based on stochastic entropies and distances for PolSAR imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(2):648–663, 2014.
- [16] R. H. Nobre, F. A. A. Rodrigues, R. C. P. Marques, J. S. Nobre, J. F. S. R. Neto, and F. N. S. Medeiros. SAR image segmentation with Rényi’s entropy. *IEEE Signal Processing Letters*, 23(11):1551–1555, 2016.
- [17] H. A. Noughabi and R. A. Noughabi. On the entropy estimators. *Journal of Statistical Computation and Simulation*, 83(4):784–792, 2013.
- [18] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- [19] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.
- [20] B. van Es. Estimating functionals related to a density by a class of statistics based on spacings. *Scandinavian Journal of Statistics*, 19(1):61–72, 1992.
- [21] O. Vasicek. A test for normality based on sample entropy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(1):54–59, 1976.