POLSAR REGION CLASSIFIER BASED ON STOCHASTIC DISTANCES AND HYPOTHESIS TESTS

Wagner Silva, Corina Freitas and Sidnei Sant'Anna

Instituto Nacional de Pesquisas Espaciais
Brazil

Alejandro C. Frery

Universidade Federal de Alagoas
Brazil

ABSTRACT

This work presents a region based classifier for Polarimetric SAR (PolSAR) images. The classifier uses the stochastic distances derived from the complex Wishart Model, obtained from the h-ϕ family of divergences. Adittionaly, a hypothesis test derived from the stochastic distance is also employed in the classification process. The region based classifier, using the Bhattacharyya distance, was applied to a polarimetric SIR-C image from an agricultural area in northeastern Brazil. The region based classification result significantly overperformed the a pixel based/contextual PolSAR classification based on the Maximum Likelihood/Iterated Conditional Modes. Such evidence lead us to conclude that the region based stochastic distance and hypothesis test classifier offers a good potential at identifying the land cover classes on a PolSAR image.

Index Terms— Region Based Classification, Stochastic Distances, Hypothesis Tests, PolSAR

1. INTRODUCTION

The classification of images obtained by polarimetric synthetic aperture radar (PolSAR) sensors is one of the main information extraction techniques from that kind of data. Generally, PolSAR classification falls into two categories: target decomposition [1] and PolSAR data statistical modeling.

Regarding the statistical modeling, the multiplicative model, which takes into account the contribution of backscatter and speckle noise, has been suitably employed. The return can be modeled by the complex Wishart distribution [2, 3]. Several pixel-based classifiers were developed from this statistical modeling, being one of them the maximum likelihood classifier used in [3].

Pixel based classifiers can be improved by the use of spatial context. Frery et al. [4] developed an ICM – Iterative Conditional Modes classifier which employs the maximum likelihood classification result under the complex Wishart distribution as starting point, and point wise evidence, and the Potts model as local information. This classifier quantifies the spatial information by a maximum pseudolikelihood estimator.

This work was supported by CNPq, CAPES, Fapesp and Fapeal, Brazil.

It is believed that even better PolSAR classification results can be achieved using segmented images (region-based classification). This classification strategy may use stochastic distances between the statistical distributions that model segments and training samples which represent classes. In the case of PolSAR data, these distances must be defined between pairs of complex Wishart distributions.

Salicru et al. [5] developed analytical dissimilarity measures, the so called h-ϕ family of divergences, from the convenient definition of the functions h and ϕ. Hypothesis tests based on statistics derived from h-ϕ divergences were also developed in [5]. Frery et al. [6, 7] obtained five distances between complex Wishart distributions: Kullback-Leibler, Bhattacharyya, Hellinger, Rényi and Chi-Square and their corresponding hypothesis tests were also developed and evaluated.

This work presents a new region-based classifier using analytical expressions between complex Wishart distributions and the hypothesis tests derived from them. The classifier performance evaluation is performed with SIR-C mission images. The classifier was developed using the Interactive Data Language (IDL), version 7.1.

2. CLASSIFICATION OF SEGMENTS WITH STOCHASTIC DISTANCES

The classifier assigns labels to regions previously defined (segments) in an image according to the smallest test statistic observed between each segment and each training sample. The decision rules are based on the asymptotic distribution derived in [6] for the test statistics. The classifier is, therefore, supervised, region based and statistical.

Figure 1 shows the sequence of procedures performed by the classifier. The input data consist of the PolSAR images to be classified, a segmented image, and a set of training samples for each class. The equivalent number of looks n must also be provided in advance by the user. This parameter can be estimated by one of the methods described in [8], and is assumed constant for the whole image.

The classifier is composed of three models whose use depends upon the type of input data: the Wishart distribution

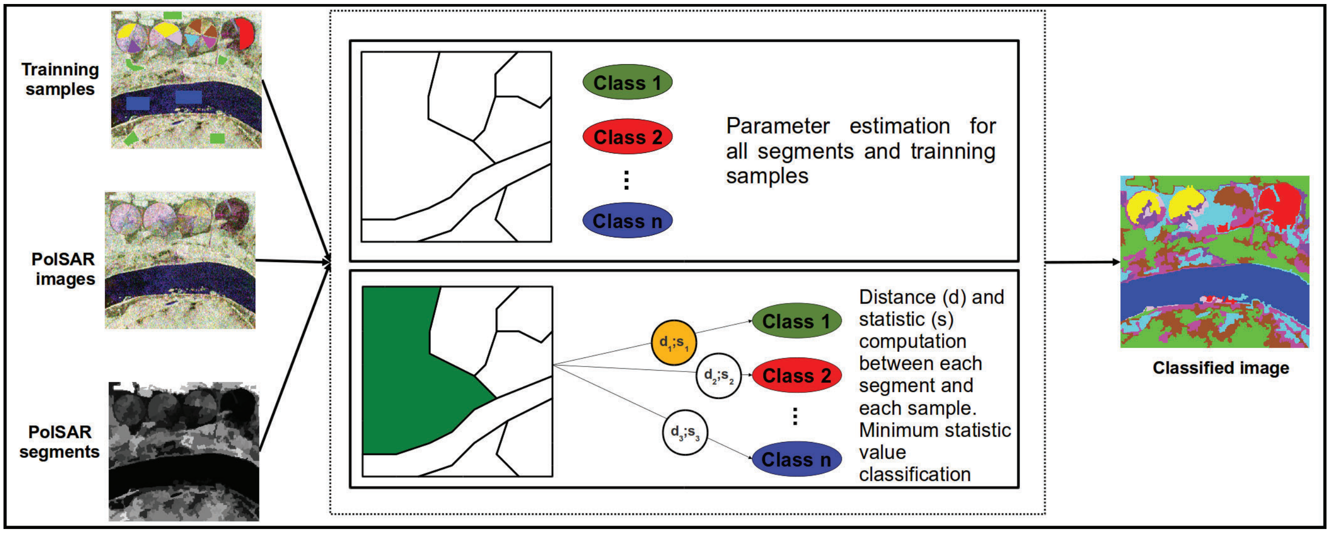


Fig. 1. Classification Process.

(adequate when the input data are in full polarimetric complex covariance matrix format), the distribution of pairs of intensities [9] (adequate when the input data are two Gamma distributed intensities), and the multivariate Gaussian distribution. Each of these models is associated to a classifier module, and different stochastic distances have to be calculated in each module. A description of these modules is presented in Table 1.

Table 1. Classifier Modules.

Statistical Modeling	Distances	Solution
Complex Wishart	Bhattacharyya Kullback-Leibler Hellinger Rényi Chi-Square	Closed form
Intensity Pair	Bhattacharyya Triangular	Numerical
Gaussian	Bhattacharyya Kullback-Leibler	Closed form

In this paper, the presented results are based only on the Bathacharyya distance derived from Complex Wishart distribution, since the data employed are from a PolSAR image in the covariance matrix format. The Bhattacharyya distance was obtained in [6]. Let $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ be the covariance matrices of order q from two distinct segments. The Bhattacharyya distance (d_B) between the two Wishart distributions is given by:

$$d_B = n\left[\frac{\log|\boldsymbol{C}_1| + \log|\boldsymbol{C}_2|}{2} - \log\left|\left(\frac{\boldsymbol{C}_1^{-1} + \boldsymbol{C}_2^{-1}}{2}\right)^{-1}\right|\right], \quad (1)$$

where $|.|$ denotes the determinant. After computing the Bhattacharyya distance between the region and each class prototype, the classifier computes the hypothesis test statistic s_B, which is used to classify the delineated image regions. According to the development by Salicru et al. [5] and the methodology proposed in [6], the test statistic is given by

$$s_B = \frac{4mn}{m+n} 2d_B, \quad (2)$$

where m and n are the size of the samples from the segment and the prototype, respectively. The parameters must be estimated by maximum likelihood.

3. APPLICATION

3.1. Description

In order to assess the region-based classifier, a classification was performed using the Bhattacharyya distance between complex Wishart distributions. The result was qualitatively and quantitatively compared with the classification obtained by the contextual Maximum likelihood/ICM classifier [4]. Three polarimetric bands (HH, HV and VV) from the SIR C/X-SAR mission were used. The images and study area details are presented in Table 2.

The land cover classes in the study area was identical to the one employed in [4], as well as the training and test samples, which are shown in Figures 2(a) and 2(b).

The segmentation was performed using the software SegSAR [10]. The SegSAR software is suitable for amplitude and intensity image segmentation, so we had to extract the intensity data from the covariance matrix to perform the exploratory segmentation. After a data analysis, the chosen

Table 2. Images and study area information.

Aquisition date	14/04/1994
Study area location	$09^o\ 07^{'} S$, $40^o\ 18^{'} W$, about 40 km northeast of the city of Petrolina-PE, Brazil.
Image size (pixels)	407x370
Frequency	L-Band - 1.254 GHz

parameters were 100 pixels of minimum area and 1 dB of similarity. The equivalent number of looks was 2.97, estimated using the method described in [4], which was also described and referred to as Fractional Moment Estimate in [8].

3.2. Results

Two classified images were obtained, one for the region based classifier using Wishart Bhattacharyya distance, and another from the contextual Maximum likelihood/ICM classifier [4]. A quantitative evaluation was performed using the test samples. The overall accuracy and the kappa coefficient of agreement were computed from the confusion matrix, as formulated in [11]. The classification results are shown in Figures 2(d) and 2(e), and the quantitative assessment of these results (Kappa Coefficient (κ) and global accuracy) are shown in Table 3. A hypothesis test of equality of kappa values showed that the region based classifier achieved a better performance than the pixel based/contextual, at any pratical significance level. These results can be checked in Table 4.

Table 3. Kappa Coefficient (κ), κ variance and global accuracy (G.A.) - Region based and contextual results.

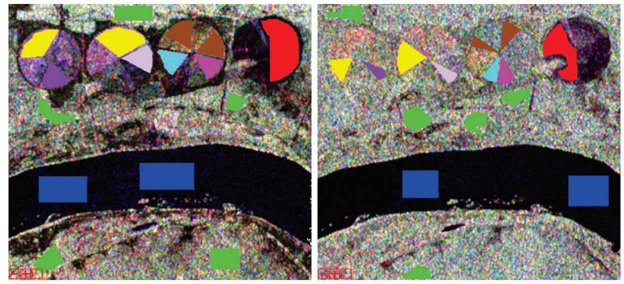
Classifier	κ	κ variance	G.A.
1	0.834	$1.25 \cdot 10^{-5}$	86.60%
2	0.803	$1.43 \cdot 10^{-5}$	83.97%

1 - Wishart Bhattacharyya Distance (region based)
2 - Maxlikellihood/ICM (pixel based/contextual)

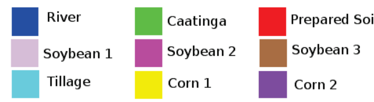
Table 4. Equality test between confusion matrices (significant level: 5%.)

Compared classifiers	*z-statistic*	p-value
1 vs. 2	6.91	≈ 0

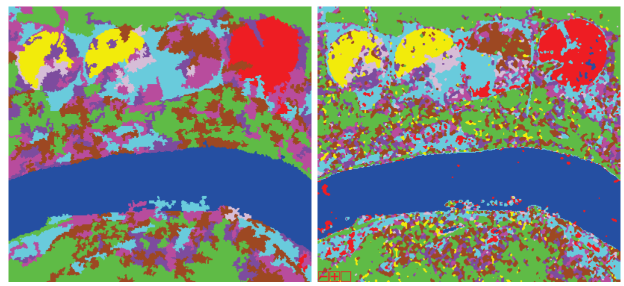
1 - Wishart Bhattacharyya Distance (region based)
2 - Maxlikellihood/ICM (pixel based/contextual)



(a) Trainning samples (b) Test samples



(c) Land cover classes legend



(d) Bhattacharyya Distance classification (e) ICM classification

Fig. 2. Trainning and test samples, region based and contextual classifications.

4. CONCLUSIONS

A new region-based classifier for PolSAR data using stochastic distances between complex Wishart distributions and derived hypothesis tests was presented. In comparison with a contextual Maximum likelihood/ICM classifier, the new classifier produced significantly better results.

From a quantitative point of view, one can conclude that the region-based classification significantly outperforms the contextual classifier, since its result reaches a kappa coefficient of 0.834 against 0.803. A statistical test rejected the equality hypothesis between the two confusion matrices at a significant level of 5%, as the computed *z-statistic* result was 6.91, yielding an approximate null p-value.

The quality improvement can also be observed qualitatively by examining the huge amount of undesirable small areas that still exist in the contextual result, while those artifacts are quite minimized by the region-based classification.

Using a region based classifier on PolSAR images segments can produce good results by avoiding small areas caused, among other factors, by the speckle noise or terrain distortions, like shadows.

It can be concluded therefore that the proposed classifier has great potential for PolSAR data classification. In the future, further investigation will be conducted using other distances and remote sensing data.

5. REFERENCES

[1] J.S. Lee and E. Pottier, *Polarimetric radar imaging*, CRC Press, 2008.

[2] J.W. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction)," *Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152–177, 1963.

[3] J.S. Lee, M.R. Grunes, and R. Kwok, "Classification of multi-look polarimetric sar imagery based on complex wishart distribution," *International Journal of Remote Sensing*, vol. 15, no. 11, pp. 2299–2311, February 1994.

[4] A.C. Frery, A.H. Correia, and C.C. Freitas, "Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3098–3109, oct. 2007.

[5] M. Salicru, D. Morales, M.L. Menendez, and L. Pardo, "On the applications of divergence type measures in testing statistical hypotheses," *Journal of Multivariate Analysis*, vol. 51, no. 2, pp. 372–391, November 1994.

[6] A C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Information theory and image understanding: An application to polarimetric SAR imagery," *Chilean Journal of Statistics*, vol. 2, no. 2, pp. 81–100, 2011.

[7] A.C. Frery, A.D.C. Nascimento, and R.J. Cintra, "Hypothesis test in complex wishart distributions," in *Proceedings of 5th International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry, POLinSAR 2011*, Frascati, Italy, january 2011.

[8] S.N. Anfinsen, A.P. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3795–3809, 2009.

[9] J.S. Lee, K.W. Hoppel, S.A. Mango, and A.R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 5, pp. 1017–1028, Sep 1994.

[10] M. A. Sousa Júnior, L. V. Dutra, and C. C. Freitas, "Segmentação de imagens JERS e TM/Landsat usando o segmentador incremental multi-níveis SegSAR," in *Proceedings...*, São José dos Campos, 2005, Simpósio Brasileiro de Sensoriamento Remoto, 12.(SBSR), pp. 4493–4500, INPE.

[11] R.G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of Environment*, vol. 37, no. 1, pp. 35 – 46, 1991.