SAR Segmentation Algorithms: a Quantitative Assessment

Alejandro C. Frery
UFPE-DI

CP 7851, 50732-970 Recife, PE-Brazil
Tel: +55-81-2718430, Fax: +55-81-2718438
frery@di.ufpe.br

Eduardo V. D. Lucca - Corina da C. Freitas - Sidnei J. S. Sant'anna
INPE-DPI

Av. dos Astronautas 1758, 12227-010 São José dos Campos, SP-Brazil
Tel: +55-12-3456475, Fax: +55-12-3456468
lucca@ieav.cta.br; corina@dpi.inpe.br; sidnei@dpi.inpe.br

ABSTRACT

This paper compares the performance of two segmentation algorithms, the MUM (Merge Using Moments) and RWSEG, using simulated Synthetic Aperture Radar (SAR) data. This comparison is performed from both the qualitative and quantitative points of view, and carried out using synthetic images that contain regions with different homogeneity degrees, aiming at land-use applications. The process for obtaining simulated images consists of creating an idealized classes image (a phantom), which summarizes the main geometric and topologic characteristics of Amazonian targets. Then a statistical model of observations from each class is proposed, and a Monte Carlo experience is performed.

Keywords: SAR, segmentation; MUM; RWSEG.

INTRODUCTION

One of the first steps in image analysis consists of decomposing the input data into a collection of uniform, continuous and disjoint regions: the process known as *segmentation*. It is one of the most critical tasks in image analysis, and its importance has boosted the development of a great variety of algorithms aiming at providing an automatic way to obtain segmentations. Nevertheless, such algorithms should not be applied to all types of images and, besides that, they depend on the type of application in mind.

The evaluation and comparison methods for segmentation aims at determining their limitations and advantages. These methods can be divided into analytical and empirical: the former directly examine and assess the segmentation algorithms themselves by analysing their principles and properties, whereas the latter perform the evaluation through their results [9]. Several methods have been proposed in order to evaluate and compare segmentation algorithms, see for instance [1,3,7,9].

The purpose of this paper is to compare quantitatively two segmentation algorithms available in the software CAESAR v. 2.1: the MUM and the RWSEG [6]. These algorithms are developed specifically for SAR data, and they are applied to

simulated SAR images with 3 and 8 looks. This evaluation is performed employing empirical discrepancy methods.

SAR DATA MODELLING

The multiplicative model has been widely used in the modelling, processing and analysis of SAR images. This model assumes that the observed return Z is a random variable defined as the product between the random variables X (the terrain backscatter) and Y (the speckle noise) [4].

Different types of region and kinds of detection (intensity or amplitude format) can be associated to different distributions for the return. Regarding to region types, the homogeneous (agricultural fields, bare soil and pasture areas, for example), heterogeneous (primary forest, for example), and the extremely heterogeneous (urban areas, for example) are considered in this work. A common hypothesis is that the return in amplitude case follows the Square Root of Gamma, the K -Amplitude and the G^0 -Amplitude distribution in homogeneous, heterogeneous, and extremely heterogeneous areas, respectively [4]. The particular distributions used in this study are presented in [5], and they were chosen to be representative of the data often encountered in SAR images.

The considered homogeneity degrees also depend on sensor parameters such as wavelength, angle of incidence, polarization, etc.

SEGMENTATION ALGORITHMS

The segmentations algorithms evaluated in this paper are: the MUM [2] and the RWSEG [8]. They both produce segmentations based on gray-levels, modeled by the square root of Gamma distribution only, and they also assume that the pixel values in an image are uncorrelated.

The MUM algorithm starts with a fine segmentation (it can be assumed that each pixel is a region, for instance) and proceeds by comparing neighbouring regions. The regions that are "significantly different" are left aside, and those regions that are "similar" are tagged. The tagged regions are then sorted out and as many as possible are merged. This process of tagging, sorting and merging continues until there

are no more regions to be merged [2]. The termination criterion is controlled by the user.

The RWSEG algorithm segments an image by successive edge detection and region growing. At the end of this iterative process, a region merging is used to produce the final segmentation. During the iterative process the detected edges are used to limit region growing, and the resulting segmentation is used to generate an improved edge map [8]. After each iteration, the average contrast of segments is measured, and iteration continues until the average contrast decreases. The final stage checks if adjacent segments are statistically distinct and merges segments which are not.

QUANTITATIVE MEASURES OVER SEGMENTATIONS

The measures used in this paper are those reported in [3]. They measure the difference between a segmented and a reference image, being the latter in our case the phantom (see Figure 1 (a)).

The first step of the evaluation consists of comparing the regions detected in the segmented image with those in the reference image. This step is based on a fitting process: for each region in the reference image, there will be one in the segmented image which is selected to represent it best. This region in the segmented image is called *fitted region*. Each fitted region is evaluated by criteria of position, number of pixels, mean intensity, and shape.

Let the original and segmented (matched) images have $xpix$ columns by $ypix$ rows. A region in the reference and segmented images is referred to as $1 \leq i \leq N$ and $1 \leq f \leq M$, respectively. Averages of the original image g over a region in the reference and segmented images are denoted $\langle g_i \rangle$ and $\langle g_f \rangle$, and $N(i)$ and $N(f)$ denote the number of pixels in regions i and f respectively. Two matrices of $N \times M$ elements, denoted Gf and Fit are built with components

$$Gf(i,f) = \frac{N(i \cap f)}{N(i \cup f)} \quad \text{and} \quad Fit(i,f) = \frac{xd + yd + \frac{pd + id}{2}}{Gf(i,f)},$$

$$\text{where} \quad xd = \frac{|\langle x_i \rangle - \langle x_f \rangle|}{xpix}, \quad yd = \frac{|\langle y_i \rangle - \langle y_f \rangle|}{ypix},$$

$$pd = \frac{|N(i) - N(f)|}{|N(i) + N(f)|}, \quad id = \frac{|\langle I_i \rangle - \langle I_f \rangle|}{|\langle I_i \rangle + \langle I_f \rangle|}. \quad \text{The values}$$

$\langle x_k \rangle, \langle y_k \rangle$ and $\langle I_k \rangle$ represent, respectively, the abscissas, ordinates and return means in region k .

The Gf matrix relates pairs of regions (i, f) which have pixels in common, while the values in the Fit matrix represent a fit success measure between regions taking into account size, shape, position, and data mean intensity of regions. For each region i in the reference image its

corresponding fitted region f is taken to be at the minimum value of $Fit(i, f)$.

Once defined the fitted regions, the fit success can be evaluated through quantitative measures that take into consideration relative aspects of position, mean intensity, size, and shape of fitted regions. The measures used in this work are *Success Of Position* $Fitxy = 1 - (xd + yd)/2$, the *Success Of Intensity* $Fiti = 1 - |\langle I_i \rangle - \langle I_f \rangle| / (|\langle I_i \rangle + \langle I_f \rangle|)$, the *Success Of Size* $Fitn = 1 - (N(i) - N(f)) / (N(i) + N(f))$ and the *Success Of Shape Fit* $Gshape = N(i \cap f) / N(i \cup f)$.

Note that these measures vary between 0 and 1, where 1 represents the best fit. The tested segmentation can be quantitatively measured through the mean value of these quantities over the N regions, or through a single value, the general mean fit.

MONTE CARLO EXPERIENCE

The MUM and RWSEG algorithms were applied to simulated unprocessed images with 3 and 8 looks (Figure 1 (b) and (c) respectively), in order to determine the specific parameters that must be selected by the user. These parameters were selected in order to obtain a good deal of combinations. Seventeen MUM and ninety-six RWSEG segmentations were obtained, and the best set of parameters for each situation was obtained applying the criteria already presented for segmentation assessment.

In order to reduce the influence of a particular image over the assessment, a Monte Carlo experience was performed using images with the same statistical properties. After one hundred replications, vectors of measures are obtained for each segmentation algorithm, and compared.

EXPERIMENTAL RESULTS

The results of the Monte Carlos experience are presented in Table I, showing the mean and the standard deviation (scaled by factors of 100 and of 10^{-6} , respectively). The winning situations between algorithms are highlighted.

It is immediate that both algorithms behave better with 8 looks than with 3 looks, which is expected from the statistical modeling (see [4,5]), since in the latter case the signal-to-noise ratio is higher than in the former.

MUM is better than RWSEG according to the $Fitxy$ and $Gshape$ criteria, while RWSEG is better than MUM according to the $Fiti$ criterion. When looking at the $Fitn$ criterion, MUM is better than RWSEG when 8 looks are used and RWSEG is better than MUM when dealing with 3 looks.

The overall mean criterion points at the MUM algorithm as the best one.

Table I: Mean and standard deviation of measures of correspondence between true and obtained classifications, for two algorithms (MUM and RWSEG) and two situations (3 and 8 looks).

Looks	MUM		RWSEG	
	3	8	3	8
<i>Fitxy</i>	98.38(9)	98.68(7)	97.82(13)	98.04(12)
<i>Fiti</i>	96.90(12)	97.42(7)	97.12(11)	97.48(9)
<i>Fitn</i>	73.80(554)	77.92(402)	73.99(672)	77.37(576)
<i>Gshape</i>	61.63(475)	67.39(357)	59.09(479)	64.56(392)
Mean	82.68(142)	85.35(107)	82.01(153)	84.36(123)

CONCLUSIONS

The analysis of the MUM and RWSEG general means obtained for each replicated image in Monte Carlo experience allows to conclude that, for the proposed image set, the MUM algorithm produces quantitatively better segmentations than those produced by the RWSEG. This result is confirmed when comparing the total mean obtained for both algorithms.

The MUM is more skillful than RWSEG at producing segments at the correct position and with the correct shape. However, both algorithms are equivalent in terms of size and intensity of the produced segments. Thus, for the application type defined to compare both algorithms, the MUM is more appropriate to perform the image segmentations.

The comparison of the quantitative measures and visual analysis result over the produced segmentations reveals that even visually different segmentations have very close quantitative values. However, the quantitative measures confirm the qualitative analysis results.

ACKNOWLEDGMENTS

This work was supported by grants from PPG-7 (0808/95 and 0816/95), Convênio FINEP (6.6.96.0473.00

6.6.96.0474.00), CNPq (Proc. 523469/96-9) and FACEPE (APQ 0707-1.03/97).

REFERENCES

- [1] Caves, R.; Quegan, S.; White, R.G. "Quantitative comparison of the performance of SAR segmentation algorithms", unpublished.
- [2] Cook, R.; McConnell, I.; Oliver, C.J. "MUM (Merge Using Moments) segmentation for SAR images". In: Proc. SPIE, 2316, 92-103, 1994.
- [3] Delves, L.M.; Wilkinson, R.; Oliver, C.J.; White, R.G. "Comparing the performance of SAR segmentation algorithms". International Journal of Remote Sensing, vol. 13, pp. 21221-2149, 1992.
- [4] Frery, A.C.; Müller, H.-J.; Yanasse, C.C.F.; Sant'Anna, S.J.S. "A model for extremely heterogeneous clutter". IEEE Transactions on Geoscience and Remote Sensing, vol. 35, pp. 648-659, 1997.
- [5] Lucca, E.V.D.; Freitas, C.C.; Frery, A.C.; Sant'Anna, S.J.S. "Comparison of SAR segmentation algorithms". In: Segunda Jornada Latinoamericana de Sensoriamento Remoto por Radar - Técnicas de Processamento de Imagens, 1998. Santos, SP, INPE/ESA, pp. 123-130. (ESA SP 434).
- [6] NASoftware. Caesar: user guide. Version 2.1. Liverpool, 1994.
- [7] Quegan, S.; Rye, A.J.; Hendry, A.; Skingley, J.; Oddy, C.J. "Automatic interpretation strategies for synthetic aperture radar images". Phil. Trans. R. Soc. Lond. A, vol. 324, pp. 409-421, 1988.
- [8] White, R.G. Low-level segmentation of noise imagery. Technical Report 3900, DRA, St Andrews Road, Malvern, Works, 1986.
- [9] Zhang, Y.J. "A survey on evaluation methods for image segmentation". Pattern Recognition, vol. 29, pp. 1335-1346, 1996.

(a)

(b)

(c)

Figure 1: (a) Phantom, (b) Simulated image, 3 looks, (c) Simulated image, 8 looks.