LAND COVER DISCRIMINATION AT BRAZILIAN AMAZON USING REGION BASED CLASSIFIER AND STOCHASTIC DISTANCE

Wagner B. Silva, Luciana O. Pereira, Sidnei J. S. Sant´Anna, Corina C. Freitas, Ricardo J. P. S. Guimarães,
Instituto Nacional de Pesquisas Espaciais – INPE, Brazil

Alejandro C. Frery
Universidade Federal de Alagoas – UFAL, Brazil

ABSTRACT

Given the different nature of optical and radar data, it is reasonable the idea that each type of data can contribute in complementary ways for different applications. This paper aims at analyzing the potential joint usage of optical and Synthetic Aperture Radar (SAR) data for land use and land cover classification in a region located in the Brazilian Amazon. To achieve this objective, we evaluated region-based classifications using separated and fused optical and SAR data. Data were images from the Landsat 5/TM sensor and amplitude multipolarized images from the ALOS/PALSAR sensor. The images were classified using a region-based classifier based on the Bhattacharyya distance between Gaussian distributions. The TM data alone is better for classify land cover classes with occurrence of trees or shrubs, while SAR data contribute to improve the classification results in low vegetated areas.

Index Terms— Fusion, Region-based classification, SAR, Stochastic Distances, Brazilian Amazon.

1. INTRODUCTION

Optical data are related to the chemical and biological properties of the targets (e.g. chlorophyll and vegetation internal moisture contents) while microwave data are mainly related to their geometry and electromagnetic properties (different sizes of plant species and texture) [1]. Therefore, the combined use of optical and synthetic aperture radar (SAR) data can improve information extraction about imaged targets. Some of the most common methods of data combination are fusion techniques. These techniques applied to optical and SAR data provide information about the targets, which, otherwise, would not be available since these techniques aim to preserve the joint characteristics of optical and microwave data. The fused data can then be used as input for several image processing techniques such as segmentation or classification.

In recent years, several specific classifiers for radar have been developed. Among them, maximum likelihood classifier using adequate probability density functions and classifiers based on target decomposition methods [3]. However, these classifiers are pixel-based and, therefore, due to speckle noise, which is inherent to radar data, the results obtained from these type of classifiers are usually very noisy. Contextual classifiers, as the Iterative Conditional Modes (ICM) [2], incorporate context. Another approach is the use of region-based classifiers, which employ collections of similar pixels to make decisions. Region classifiers can use of stochastic distances, which measure the separability between the probability distributions of the data from two different regions on the images.

The objective of this paper is to analyze the joint usage of optical and radar data for land use and land cover classification purposes, in a region located in the Brazilian Amazon. To achieve this objective, region-based classifications using separated and fused optical and SAR data were performed, and the results were evaluated with overall accuracy, kappa coefficient of agreement and by visual inspection.

2. METHODOLOGY

The study area is located to the east of the Tapajós National Forest, Pará State, Brazil and it is characterized by having hot

and humid climate and the predominance of humid tropical rainforest. There are areas of degraded forest (mainly due to burning), and forest areas at different stages of regeneration, as a result of human actions in the proximity of the BR-163 highway. Agricultural activities are frequent in the study area, characterized mainly by the production of soybeans and other grains such as rice, maize, sorghum and beans. The agricultural calendar starts in December, at the beginning of the rainy season, with one or two planting activities during the year, being the harvests carried out respectively in March/April and July/August.

The land use and land cover classes used in this work were defined based on field works done on dates close to those of image acquisitions. The classes of interest are presented in Table 1. After a careful exploratory analysis, the degraded forest (DF), old regeneration (OR) and intermediate regeneration (IR) classes were grouped into a single class, totalizing eight land use and land cover classes. The number of pixels used for training and testing the classifier, and the color associated with each class are also presented in Table 1.

Table 1 – Land Use Classes and it's properties

Land use Class	Description	Training Samples	Test Samples	Color
Primary Forest - PF	Intact natural forest which has not been disturbed or modified by human activity	13858	7258	
Degraded Forest - DF	Primary forest which has suffered fire actions	5480	2846	
Old Regeneration -OR	Secondary forest older than 20 years of regrowth			
Intermediate Regeneration - IR	Secondary forest with 8 to 20 years of regrowth			
New Regeneration - NR	Secondary forest newer than 8 years of regrowth	1049	546	
Clean Pasture - CP	Land used for grazing animals with low presence of trees, shrubs and invasive species	1623	292	
Dirty Pasture - DP	Land used for grazing animals with presence of trees, shrubs and invasive species	510	495	
Fallow Agriculture - FA	Resting agricultural areas after harvest	1252	450	
Soybean1 – S1	Soybean ready for harvest (after 100 days of seeding approximately)	562	306	
Soybean2 – S2	Soybean with approximately 40 days of seeding	142	68	

The data consisted of multi-spectral images from Landsat 5/TM sensor with six bands, acquired on 06/29/2010, and two amplitude multipolarized images (HH and HV) from the ALOS/PALSAR sensor, Fine Beam Dual mode (FBD), 1.5 processing level, from 06/21/2010.

Classifications were performed using an in-house region classifier. This classifier provides several distance measures between probability distributions (stochastic distances) to

classify images modeled by Gaussian distributions, pairs of intensity SAR images modeled by bivariate Gamma distributions, and polarimetric SAR images, modeled by Wishart distributions, using the covariance matrix. This classifier will, in a near future, use hypothesis tests derived from the distances to provide information about the quality of classification. The module of the classifier used in this work is based on Bhattacharyya distance and assumes the data are

modeled by multivariate Gaussian distribution. The distance calculation follows the methodology described in [4].

Data were fused by Principal Component Analysis (PCA) applied to the six Landsat TM 5 bands. The first principal component was then replaced by the HH component of SAR data, followed by inverse transformation.

The SAR and TM images were segmented using the SegSAR software [5] for the SAR data and ENVI EX for optical data. The results of these two segmentations were matched, producing a final segmented image, which was used as an input for the region classifier. Three classifications were performed, all with the same segmented image, using the following data sets: a) Multispectral TM/Landsat 5 images, with six bands, b) pair of ALOS/PALSAR amplitude images (two bands) and c) fused image - PCA transformation with TM and HH ALOS/PALSAR images.

Classifications were evaluated with test samples (see Table 1) from all classes, generating a confusion matrix for each classification. The overall accuracy, kappa coefficient of agreement (overall class and per class) [6] and their variances were computed from the confusion matrices.

3. RESULTS

The values of overall accuracy, kappa coefficient of agreement and kappa variance, obtained by the three classification schemes are showed in Table 2.

Table 2 – Classification results

Classification scheme	Overall Accuracy (%)	Kappa Coefficient
TM/Landsat 5	95.65	0.926
ALOS/PALSAR (HH and HV)	51.90	0.305
PCA TM - PALSAR(HH)	95.42	0.922

From the values of kappa coefficient and kappa variance, it can be concluded that the classifications using TM imagery

and the fused image are statistically equal under a confidence level of 95%.

These results show that such classifications are, in general, much better than the classification using only PALSAR amplitude pair (HH and HV).Besides the equality of the overall kappa, which can lead to the conclusion that the usage of SAR data does not contribute to improve the optical data classification, a detailed analysis of the confusion matrices and per-class kappa coefficients (not shown here) indicates that different data types outperform the others, as illustrated in Figure 1 and exposed in the following analysis:

- TM data achieved the best performance in classifying primary forest (see Figures 1a, 1b and 1c), DF+OR+IR (see Figures 1d, 1e and 1f), new regeneration (see Figures 1g, 1h and 1i) and dirty pasture classes;

- PCA Fusion data contributed to minimize the confusion between the clean pasture and soybean2 classes and also between soybean1 and fallow agriculture classes (Figures 1j, 1k and 1l). These data contributed to improve the classification of the soybean2 class (Figures 1p, 1q and 1r).

- PALSAR data achieved the best performance in classifying the soybean1 class, as can be seen in Figures 1m, 1n and 1o.

4. CONCLUSIONS

According to the results and analysis performed, it can be concluded that the region based classifier offers a good potential in identifying land cover classes in the area studied here. The analysis for individual classes show that SAR data can help to decrease the misclassification in some pairs of classes, and can even outperform the optical data classification regarding at least one of the classes (soybean1), showing the contribution of this kind of data. It is also possible to conclude that TM data alone is better to classify land cover classes with occurrence of trees or shrubs, such as primary and degraded forest, regenerations and dirty pasture.

SAR data contribute to improve the classification results in low vegetated areas, such as clean pasture and fallow agriculture, and different stages of soybean growth. It would be interesting in following works to perform a classification which takes into account each feature separately, such as a decision tree classifier, due to the observed fact that some classes, such as soybean1 or primary forest are almost completely identified by the usage of only one data type.
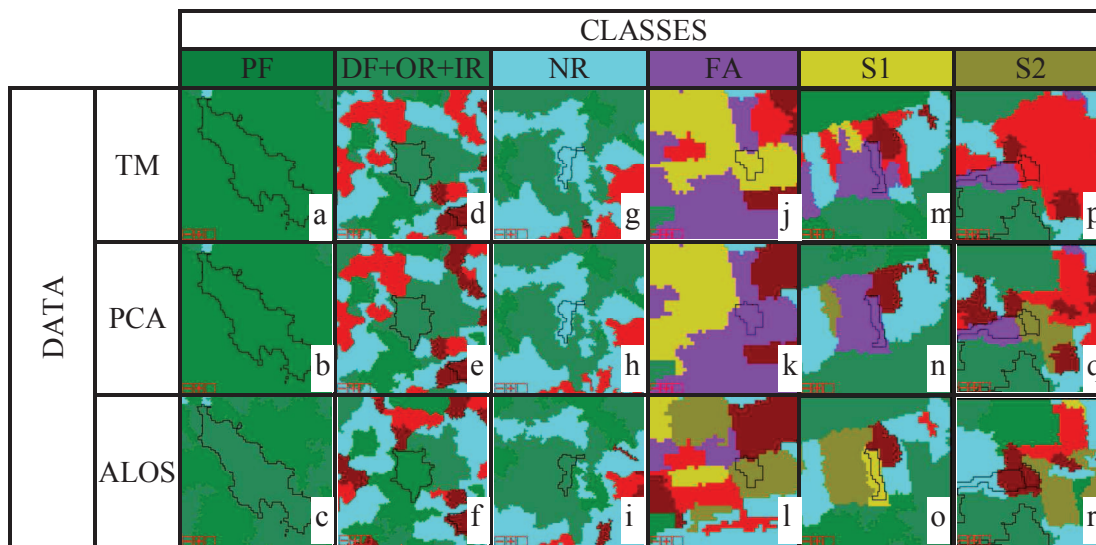
CLASSES						
	PF	DF+OR+IR	NR	FA	S1	S2
TM	a	d	g	j	m	p
PCA	b	e	h	k	n	q
ALOS	c	f	i	l	o	r

Figure 1: Specific classes results examples

5. REFERENCES

[1] Huang, S.; Potter, C.; Crabtree, L.D.; Hager, S.; Goss, P. Fusing optical and radar data to estimate sagebrush, herbaceous, and bare groundcover in Yellowstone. **Remote Sensing of Environment**, v. 114, n. 2, pp. 251-264, February 2010.

[2] Frery, A.C.; Correia, A.H.; Freitas, C.C. Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information. **IEEE Transactions on Geoscience and Remote Sensing**, v. 45, n. 10, p. 3098–3109, 2007.

[3] Lee, J.S., Pottier, E. **Polarimetric Radar Imaging: From Basics to Applications.** CRC Press, Boca Raton, Florida, USA, 2nd edition, 2009.

[4] Nascimento, A.D.C.; Cintra, R.J.; Frery, A.C. Hypothesis Testing in Speckled Data With Stochastic Distances. **IEEE Transactions on Geoscience and Remote Sensing**, v. 48, n.1, p.373-385, January 2010.

[5] Sousa Júnior, M.A. **Segmentação multi-níveis e multi-modelos para imagens de radar e ópticas**. 2005. 131 p. PhD Thesis – National Institute for Space Research, São José dos Campos. 2005. (in portuguese).

[6] Congalton, R.G. and Green, K. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC Press, Boca Raton, Florida USA, 2nd edition, December 2009.