Deforestation pattern characterization in the Brazilian Amazonia

Felipe Castro da Silva[1]
Thales Sehn Korting[1]
Leila Maria Garcia Fonseca[1]
Maria Isabel Sobral Escada[1]

[1] Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{felipe, tkorting, leila, isabel}@dpi.inpe.br

Abstract. The characterization of landscape objects can vary when considering different spatial resolution and different deforestation patterns in the Amazonia. Therefore, this paper aims to evaluate the effects of spatial resolution and different deforestation patterns on the performance of landscape metrics. Deforestation maps from MODIS (250*m*) and TM (30*m*) sensors were used in this study. The experiments were performed in the region called "Terra do Meio", in São Félix do Xingu Municipality, Pará State in three different test areas for the same landscape metrics set. The results have shown, with good accuracy, that one can use similar landscape metrics sets, extracted from coarser spatial resolution images, to characterize landscape objects extracted from higher spatial resolution images and vice- versa.

Keywords: image processing, landscape metrics, data mining, patterns of changes.

1. Introduction

National Institute for Space Research (INPE) has two important projects related to Amazon deforestation monitoring, known as DETER[1] and PRODES[2]. DETER uses sensor images with coarse spatial resolution (250*m*) while PRODES estimates the deforested areas with images of higher spatial resolution (30*m*).

(Frohn and Hao, 2005) evaluated the performance of 16 landscape metrics in six different years of TM/Landsat data, in relation to spatial aggregation in a deforested area in Rondônia State. Most of them produced consistent and predictable results in relation to spatial resolution degradation. (Silva, 2006) proposed a methodology to detect land use patterns in Brazilian Amazonia region, based on PRODES database. He showed that landscape metrics, such as area and perimeter, are relevant to identify the land patterns. Many works have been studying the performance of landscape metrics to characterize deforestation (Tucker, 2002; Imbernon, 2001; Frosini et. al., 2005). Therefore, this study aims to evaluate the performance of some landscape metrics to identify land use patterns, varying spatial resolution and regions. The experiments are taken for two different sensors (MODIS: *250m* and TM: 30*m*), combining resolution changes and three different areas in the same region, to validate the landscape metrics robustness.

In order to evaluate the performance of the geometric landscape metrics, a software was developed using TerraLib Library (TerraLib, 2006). Three different modules compose this system: segmentation, metrics extraction, training and classification. The segmentation algorithm is based on region growing method (Bins et. al., 1996). In this study the segmentation module was not used since the input data are thematic maps generated by PRODES/DETER projects. For each segment in the map a certain number of geometric landscape metrics are extracted. The segments are trained and classified to identify the

[1] DETER (Real Time Deforestation Monitoring System) uses the MODIS sensor. More information at <http://www.obt.inpe.br/deter>
[2] PRODES (Estimation of Deforestation in the Brazilian Amazon) uses the TM sensor. More information at <http://www.obt.inpe.br/prodes>

deforestation patterns. The classification is performed by a structural classifier (WEKA, 2006).

2. Study Area

We performed the analysis in the region called "Terra do Meio" (**Figure 1**), São Félix do Xingu Municipality at Pará state, Brazil. This region is a large area, where much public land has been seized by illegal procedures. The deforestation rate increased strongly in the period of 2000 to 2004. The area has large farms and small settlers associated with migration (Escada et al., 2005).

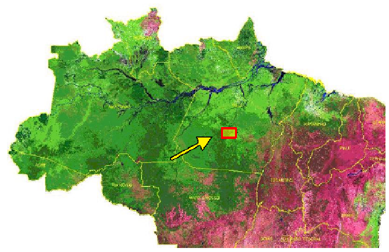

Figure 1 – Terra do Meio, city of São Félix do Xingu, PA.

Three different areas in the study area were chosen, to evaluate the robustness of the models. **Figure 2** shows the deforestation segments in the three regions for both sensors MODIS and Landsat-5 TM.

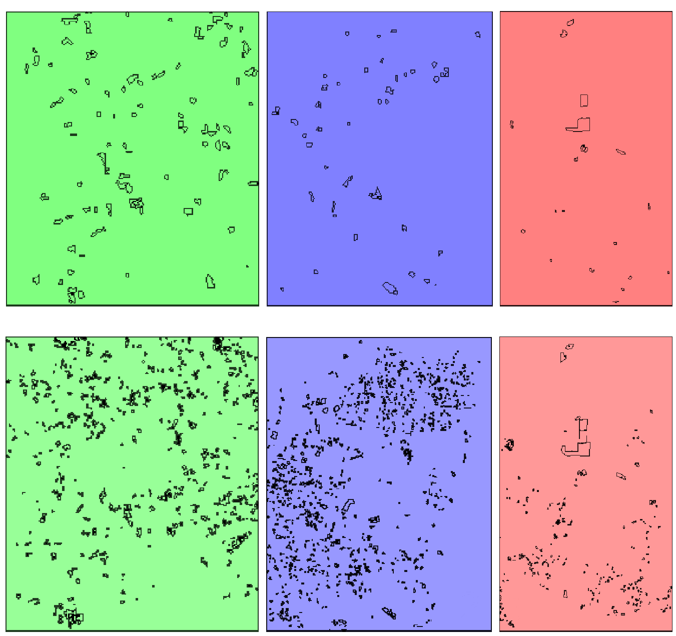

Figure 2 – Deforestation segments in the three areas: (top) MODIS and (down) TM.

Figure 3 shows the deforestation maps of the same region, in São Félix do Xingu, produced from MODIS (*250m*) and TM (30*m*) images.

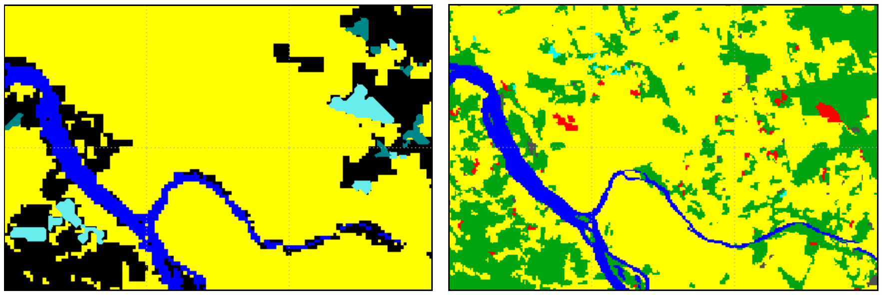

Figure 3: Deforestation in São Félix do Xingu, using MODIS (left) and TM (right)**.**

3. Methodology

The deforestation regions are represented by objects that we call landscape objects, according to (Silva, 2006). A landscape object is a structure detected in remote sensing image using an image segmentation algorithm. We characterize the landscape objects by spatial patterns (*e.g.* regular, linear, irregular), using several geometric landscape metrics. The set of landscape objects is trained using the specialist expertise. This set is used to generate a classification model, which can be used to classify the objects in the map.

To classify the objects we used the method proposed by (Silva, 2006). The method consists of three steps:

- Defining a spatial pattern typology to characterize the landscape objects.
- Building a reference set of spatial patterns, which is performed by the specialist.
- Classifying the landscape objects using a structural classifier, matching the reference set of spatial patterns to the objects identified in the images.

Our objective is to evaluate the stability of the geometric landscape metrics to characterize the objects when considering different spatial resolution and different deforestation patterns in the Amazon.

Figure 4 shows the complete system diagram. The Landscape Metrics Extraction module gets each landscape object and extracts its metrics (Section 3.1). The specialist selects some training samples and associates them to spatial patterns, resulting in a reference set of spatial patterns for the study area. Each landscape object has its metrics set, which is used to produce the classification model: a decision tree. This tree makes a binary test containing threshold for the features values, constructed by the Software WEKA[3]. Finally, the decision tree classifies all landscape objects (classification module).

For each landscape object, eight metrics are extracted and used to build the decision tree. Each landscape object belongs to one of the five patterns (**Figure 5**), namely (Silva, 2006):

- **Linear:** roadside clearings, with linear pattern following main roads matching to the earlier stages of colonization, associated to small family household;
- **Small Irregular:** found near main roads, associated to family household;
- **Irregular:** found near roads. These actors often have another incoming source from salary or commercial activities. They use family and external labor;
- **Medium Regular:** near secondary roads, associated to large farms;
- **Large Regular:** found in isolated regions, sometimes near rivers. Most of them have airstrip.

[3] WEKA is a collection of machine learning algorithms for data mining tasks. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

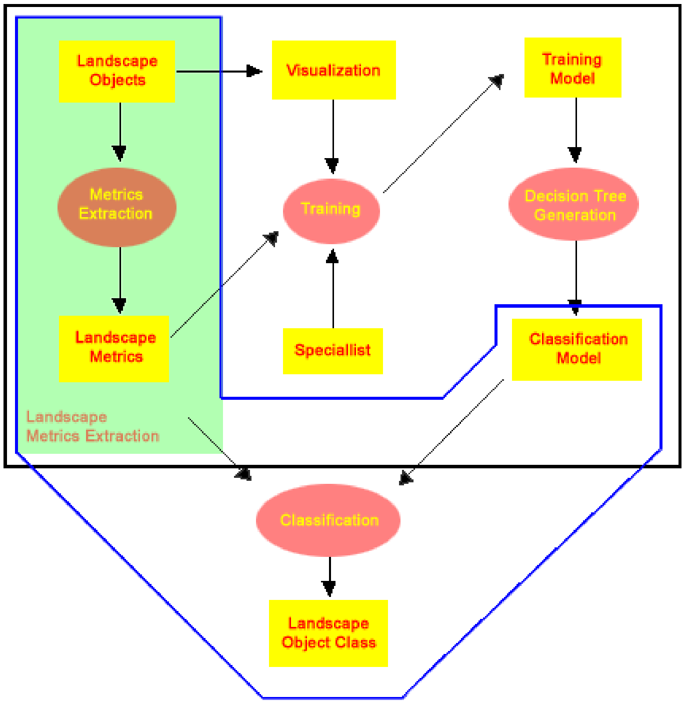

Figure 4: Application system: training (black) and classification phases (blue).

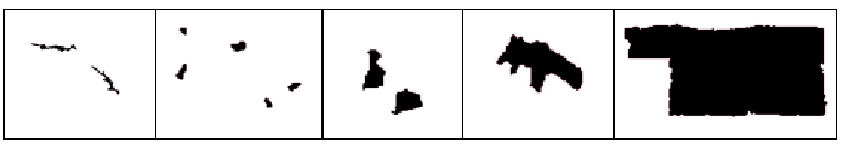

Figure 5 – Spatial patterns typologies: Linear, Small Irregular, Irregular, Medium Regular and Larger Regular, respectively.

3.1. Landscape metrics

Many landscape metrics can characterize the landscape objects, from a simple perimeter to a complex calculation of a contiguity index. **Table 1** describes eight geometric landscape metrics, their values range and their meaning. We developed an algorithm that extracts these landscape metrics from all landscape objects. To make possible the landscape metrics extraction, we used as input a set of geometric representation, obtained from deforestation thematic maps.

3.2. Training and Classification

In the training process the specialist associates some samples to their corresponding classes. These samples are used as input to the C4.5 algorithm (Wikipedia, 2006). As any training process, the number of samples of each class will determine the success or not. This algorithm is based on decision tree strategies. Thru hyperplanes, the data can be separated on the landscape metrics space.

Table 1 – Set of landscape metrics (Silva, 2006 and Fragstats, 2006).

Name	Range	Meaning
Perimeter *PERIM*	$0 < PERIM < \infty$	Perimeter of a landscape object including every internal holes
Area *AREA*	$0 < AREA < \infty$	Internal area of the landscape object
Perimeter-Area Ratio *PARA*	$0 < PARA < \infty$	Simple measure of shape complexity
Shape Index *SHAPE*	$1 \leq SHAPE < \infty$	Equals one when the landscape object is totally compact (totally square), and increases as landscape object becomes more irregular
Fractal Dimension Index *FRAC*	$1 \leq FRAC \leq 2$	Values near to one occur for objects with simple perimeters (*e.g.* squares), and come closer to two for more complex forms
Related Circumscribing Circle *CIRCLE*	$0 \leq CIRCLE < 1$	Provides a measure of overall landscape object elongation
Contiguity Index *CONTIG*	$0 \leq CONTIG \leq 1$	Represents the connectedness of the object
Radius of Gyration *GYRATE*	$0 \leq GYRATE < \infty$	Becomes greater in the ratio the extension of landscape object grows

The training module creates classification models by using the eight landscape metrics extracted in the landscape metrics extraction module. There is an alternative to perform the training stage by choosing only some landscape metrics. This allows us to choose the most suitable landscape metrics to each situation.

Classification methods based on decision trees are a sequence of boolean tests performed on the landscape metrics. Using the classification model defined in the last step, each landscape object is associated to only one class.

4. Results

The developed application performs all stages to identify the spatial pattern. It imports the landscape objects (in *shapefile* format), extracts their landscape metrics, builds a classification model (decision tree) through the training process and classifies all objects. **Figure 6** shows the system interface of the metrics extraction module.

We examined many classification models involving crossed data between different resolutions and areas to identify possible relations between distinct areas, or independence of sensors in classifying similar areas. The following sections present and discuss the results of the tests. The calculated accuracy is a rate between rightness and total number of landscape objects. With a previous classification by a specialist we can compare the classes. When a class is right, we add one to rightness. In the end, we divide rightness by total number of landscape objects to obtain the rate.

To generate the results below, first we analyzed the classification models in the same area that we got out the samples. After, we classified data from different sensors to identify what are resolution invariant landscape metrics. Finally, we analyzed the relation between different areas and same sensors.

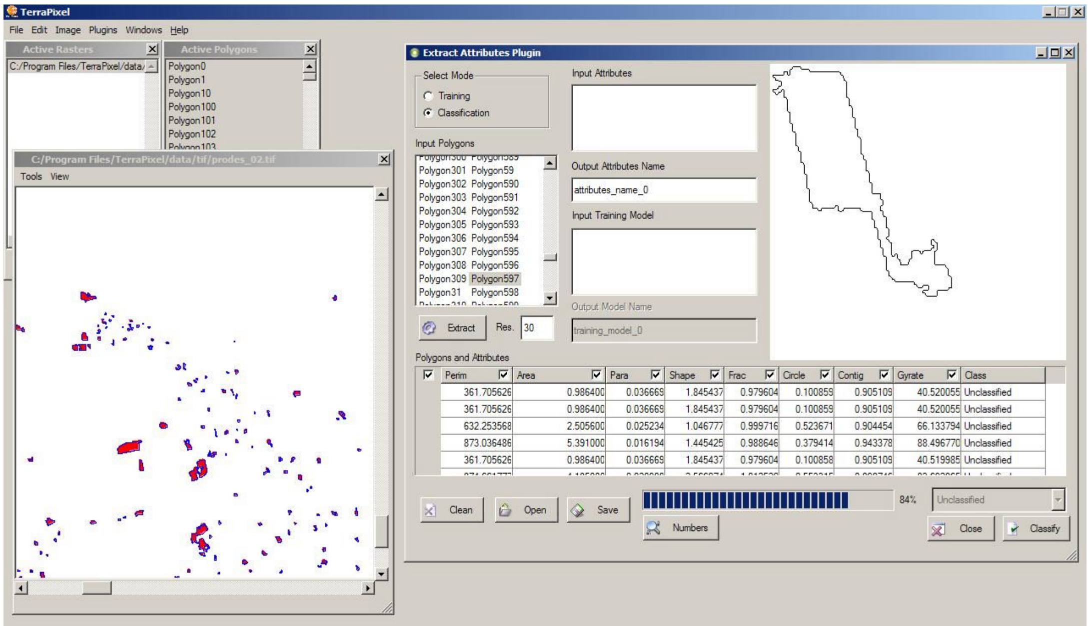

Figure 6 – Application system executing the landscape metrics extraction.

Following we will present some results of the classification analysis for different experimental tests.

4.1 Classifying the three areas using the models defined for each one

Initially, we performed a test to evaluate the classification models of each study area. In this case, the classification was performed in the same areas where the samples were selected. **Table 2** shows that the classification obtained good accuracy rates as expected.

Table 2 – Analysis about samples.

Sensors	Classification Models		
	Area 1	Area 2	Area 3
MODIS	90.91%	96.08%	92.00%
TM	95.65%	94.68%	95.10%

4.2 Classifying data from different sensors

Table 3 shows the analysis of the classification models produced by TM data and applied to MODIS landscape objects. The main diagonal presents the results when the same area is classified from data with different spatial resolutions. In this table, we can observe that good accuracy rates are hold when resolution changes, mainly when the same areas are analyzed.

Table 4 shows results obtained from the classification models using MODIS data and applied to TM landscape objects. The results also show that the main diagonal presents good accuracy rates when resolution changes for the same area. Area 2 presented the worst accuracy rate by the fact that in this case it was not possible to get many samples for each class.

Table 3 – Classification analysis for different spatial resolutions – TM to MODIS sensor.

MODIS Areas	TM Classification Models		
	Area 1	Area 2	Area 3
Area 1	**82.73%**	80.00%	73.64%
Area 2	80.38%	**78.43%**	68.63%
Area 3	72.00%	72.00%	**84.00%**

Table 4 – Classification analysis for different spatial resolutions – MODIS to TM sensor.

TM Areas	MODIS Classification Models		
	Area 1	Area 2	Area 3
Area 1	**77.64%**	73.91%	89.44%
Area 2	70.74%	**68.62%**	83.51%
Area 3	72.73%	78.32%	**84.62%**

MODIS classification model in area 3 got the best accuracy rates. A great diversity of geometries corroborates such results although it has not many landscape objects.

4.3 Spatial change analysis

A third test was carried out aiming to analyze the relation between different areas and same sensors. First, we produce classification models using MODIS data and applied to MODIS landscape objects, but in different regions. And, finally, we produce classification models using TM data and applied to TM landscape objects, also in different regions.

Table 5 and **Table 6** show the results for the MODIS and TM sensors, respectively.

Table 5 – Analysis about spatial variation – MODIS sensor.

MODIS Areas	MODIS Classification Models		
	Area 1	Area 2	Area 3
Area 1	–	84.55%	87.27%
Area 2	84.31%	–	86.27%
Area 3	80.00%	76.00%	–

Table 6 – Analysis about spatial variation – TM sensor.

TM Areas	TM Classification Models		
	Area 1	Area 2	Area 3
Area 1	–	86.96%	81.37%
Area 2	87.77%	–	81.38%
Area 3	82.52%	85.31%	–

The results show that the classification accuracy rates remain good when the areas change. One can observer that, although, the classification models have been produced from different areas, they remain robust when are applied to neighbor areas.

4.4 Landscape metrics Analysis

Some landscape metrics are not adequate when the resolution changes, so we can exclude them during the training process to avoid wrong results. However, other landscape metrics remains invariant when the resolution changes. We observed that the landscape metrics *AREA*, *SHAPE*, *CIRCLE* and *GYRATE* remained invariant when the resolution changed. To select resolution invariant landscape metrics we combined them, choosing a subset that provided the best accuracy rates. This analysis was done because statistical literature uses "curse of dimensionality" to describe difficulties associated with the density estimation feasibility in many dimensions (Warwick and Karny, 1997). Sometimes a large set of landscape metrics is not necessarily better than a subset of it.

5. Conclusion

We can use a subset of landscape metrics to train the decision tree, in images with different spatial resolutions, and even so getting high accuracy rates (more than 80%).

Taking into account that specialist analyze just a subset of the whole population, this application is useful in the deforestation classification because of the large amount of data. The software is capable to classify new regions from models created by specific knowledge, even when applied to data of different spatial resolutions, and with good accuracy rates.

Beyond the deforestation analysis this methodology can perform the data mining for other applications many different contexts such as agriculture, urban studies, etc. Some videos are available at <http://www.dpi.inpe.br/~tkorting/ index_en.html?sel=projects>.

References

Bins, L.S.; Fonseca, L.M.G.; Erthal, G.J. & Ii, F.M. Satellite imagery segmentation: a region growing approach Simpósio Brasileiro de Sensoriamento Remoto, 1996, 8, 221-223.

Escada, M. I. S., Monteiro, A. M. V., de Aguiar, A. P. D., de Senna Carneiro, T. G., and Câmara, G. Análise de padrões e processos de ocupação para a construção de modelos na Amazônia: Experimentos em Rondônia. 2005. *XII Simpósio Brasileiro de Sensoriamento Remoto*, pages 2973 – 2983.

Fragstats. Spatial Pattern Analysis Program for Categorical Maps. Available at <http://www.umass.edu/landeco/research/fragstats/fragstats.html>. Access in: 15 nov. 2006.

Frohn, R. C. and Hao, Y. Landscape metric performance in analyzing two decades of deforestation in the Amazonia Basin of Rondonia, Brazil. 2005.

Frosini de Barros Ferraz, S.; Vettorazzi, C.; Theobald, D. & Ramos Ballester, M. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondônia, Brazil: assessment and future scenarios Forest ecology and management, Elsevier, 2005, 204, 67-83.

Imbernon, J. Characterization of landscape patterns of deforestation in tropical rain forests International Journal of Remote Sensing, Taylor & Francis, 2001, 22, 1753-1765

Silva, M. P. S. Mineração de padrões de mudança em imagens de sensoriamento remoto. 2006. São José dos Campos, São Paulo, Brasil, Instituto Nacional de Pesquisas Espaciais. Tese de Doutorado.

TerraLib. TerraLib Library. Available at <http://www.terralib.org/>. Access in 15 nov. 2006.

Tucker, C. Fragmentation of a Landscape: incorporating landscape metrics into satellite analyses of land-cover change Landscape Research, Taylor & Francis, 2002, 27, 253-269.

Warwick, K. and Karny, M. Computer-Intensive Methods in Control and Signal Processing. 1997. Birkhäuser.

WEKA Software (Data Mining Software). Available at <http://www.cs.waikato.ac.nz/ml/weka/>. Access in: 20 sep. 2006.

Wikipedia. C4.5 Algorithm. Available at <http://en.wikipedia.org/wiki/C4.5>. Access in: 17 nov. 2006.