Chapter IV
Image Mining:
Detecting Deforestation Patterns Through Satellites

Marcelino Pereira dos Santos Silva
Rio Grande do Norte State University, Brazil

Gilberto Câmara
National Institute for Space Research, Brazil

Maria Isabel Sobral Escada
National Institute for Space Research, Brazil

ABSTRACT

Daily, different satellites capture data of distinct contexts, which images are processed and stored in many institutions. This chapter presents relevant definitions on remote sensing and image mining domain, beyond referring to related work on this field and to the importance of appropriate tools and techniques to analyze satellite images and extract knowledge from this kind of data. The Amazonia deforestation problem is discussed, as well INPE's effort to develop and spread technology to deal with challenges involving Earth observation resources. An image mining approach is presented and applied on a case study, detecting patterns of change on deforested areas of Amazonia. The purpose of the authors is to present relevant technologies, new approaches and research directions on remote sensing image mining, demonstrating how to increase the analysis potential of such huge strategic data.

INTRODUCTION

Motivation

Data acquisition and storage technology progress has led to a huge amount of data stored in repositories, which grow fast. Among increasing and relevant data acquired and processed, there is a strategic segment: satellite images, also known as remote sensing images.

The search for less expensive and more efficient ways to observe Earth motivated man in developing remote sensing satellites. They are currently the most significant source of new data about the planet, and remote sensing image databases are the fastest growing archives of spatial information. The variety of spatial and spectral resolutions for remote sensing images ranges from IKONOS 1-meter panchromatic images to the next generation of polarimetric radar imagery satellites. Given the widespread availability of remotely sensed data, many government and private institutions have built large remote sensing image archives.

The US National Satellite Land Remote Sensing Data Archieve, managed by USGS EROS Data Center, hosts 1.400 terabytes of satellite data gathered during 40 years. Satellites, like Terra and Aqua (NASA), generate 3 terabytes of images every day. The Brazil's National Institute for Space Research (INPE) holds more than 130 terabytes of image data, covering 30 years of remote sensing activities which are available on a database with free online access.

Actual society problems demand smart exploration of the vast and growing remote sensing data. There is a need for understanding relevant data and use it effectively and efficiently. Although valuable information is contained in image repositories, the volume and complexity of this data makes difficult (generally impossible) for human beings extract strategic information (knowledge) without appropriate tools (Piatetsky-Shapiro, Djeraba, Getoor, Grossman, Feldman & Zaki, 2006).

Data mining research has enabled powerful tools, new technologies and challenging techniques for relevant data domains. However, large image datasets need specific analysis resources and smart techniques and methodologies. The availability of huge remote sensing image repositories demands appropriate resources to explore this data.

A vast remote sensing database is a collection of landscape snapshots, which supplies a single opportunity to understand how, when and where changes occurred in the world. When such rich data is not analyzed, or it is done inefficiently, relevant information to understand complex processes and help solving challenging problems is wasted.

General Perspective and Objectives of the Chapter

In this chapter, which extends previous work (Silva, Câmara, Souza, Valeriano, & Escada, 2005), the authors intend to present relevant definitions on remote sensing and image mining domain, beyond presenting related work on this field and the importance of appropriate tools and techniques to explore satellite images and extract strategic knowledge from this kind of data.

They also discuss the Amazonia deforestation problem to demonstrate, through an image mining process, the strength of this approach to identify patterns and fight against the increase of affected areas in this forest. Developed technologies to support the process will be presented, providing an overview of methodologies, tools and techniques involved in research efforts.

Future trends and conclusion will bring reflection elements to consider classical and new mining resources to deal with challenging demands, citing limitations and also revealing directions to new research initiatives and relevant problems.

REMOTE SENSING AND IMAGE MINING

Broad Definitions

The first operational remote sensing satellite (LANDSAT-1) was launched in 1972, since then there has been a large worldwide experience in data gathering, processing and analysis of remotely sensed data. According to Canada Centre for Remote Sensing (2003), *remote sensing* is the science (and to some extent, art) of acquiring information about the Earth surface without actually being in contact with it. In other words, remote sensing is a field of applied sciences for information acquisition of the Earth surface through devices that perform the sensing and recording of the reflected or emitted energy, followed by processing, analysis, and application of this information. Such devices are called remote sensors, which are boarded on remote sensing aircrafts or satellites—also called Earth observation satellites. Images obtained through remote sensing acquisition and processing are used in many fields, once information from these remote sensing images is strongly demanded in many areas, including government, economy, infrastructure, and hydrology (e.g., security and social purposes, crop forecasting, urban planning, water resources monitoring).

In the image acquisition process, four concepts are fundamental: *spatial, spectral, radiometric and temporal resolution*. The *spatial resolution* defines the detail level of an image, that is, if a sensor has a spatial resolution of 20m then each pixel represents an area of 20m x 20m. The *spectral resolution* determines the sensor capability to define short intervals of wavelength; the finer the spectral resolution, the narrower the wavelength range for a particular channel or band. The *radiometric resolution* of an imaging system describes its ability to discriminate very slight differences in energy; the finer the radiometric resolution of a sensor, the more sensitive it is to detect small differences in reflected or emitted energy. The *temporal resolution* determines the necessary time for the sensor revisit a specific target and image the exact same area, that is, the time required to complete one entire orbit cycle; if a sensor is able to obtain an image of an area each 16 days, then its temporal resolution is this period (Canada Centre for Remote Sensing, 2003).

Before getting into remote sensing image mining, it is necessary to state *spatial data mining*, which refers to the extraction of knowledge, spatial relationships, or other interesting but not explicit patterns stored in spatial databases. Such mining approaches integrate spatial database and data mining issues, bringing valuable resources to understand facts and processes represented in spatial data, discovering spatial relationships, building up spatial knowledge bases, and revealing spatial patterns and processes contained in spatial repositories. Applications of the technology include, beyond remote sensing, geographic information systems, medical imaging, geomarketing, navigation, traffic control, enviromental studies, and many other areas where spatial data are used (Han & Kamber, 2001).

Remote sensing image mining deals specifically with the challenge of capturing patterns, processes, and agents present in the geographic space, in order to extract specific knowledge to understand or to make decisions related to a set of relevant topics, including land change, climate variations and biodiversity studies. Events like deforestation patterns, weather change correlations and species dynamics are examples of precious knowledge contained in remote sensing image repositories.

The *Amazonia forest*, located in South America, has 6,500,000 km^2, involving seven frontier countries. Brazil holds 63.4% of South America Amazonia, which extends to the following Brazilian states: Mato Grosso, Tocantins, Maranhão, Amazonas, Pará, Acre, Amapá, Rondônia and Roraima. Since it is the world's largest tropical forest, deforestation in the Amazonia rainforest is an important contributor to global land change. According to INPE's estimates, close to 200,000 km^2 of forest were cut in Brazilian Amazonia in the period from 1995 to 2005 (INPE, 2005). INPE uses LANDSAT and CBERS images to provide yearly assessments of the deforestation in Amazonia. Given the extent of the deforestation on tropical forests, figuring out the processes and its agents are important issues for setting up public policies that can help preserve the environment (Figure 1).

Figure 1. Amazonia deforestation (source: Isabel Escada - INPE)

Related Work

Nagao and Matsuyama (1980) developed, at Kyoto University (Japan), the first high level vision system for aerial image interpretation. The system processing modules operated on a common dataset. The analysis process was divided in the following steps: smoothing, when the images were processed to remove noise and spots on boundaries; segmentation, when elementary regions were extracted through a basic region growing algorithm; global exam of the scene, to estimate object domains using image metadata; detailed area analysis, when object detection subsystems analyzed a knowledge base to find specific objects; communication among object detection subsystems, in order to control the analysis flow managing the information on databases, resolve conflicts among detection subsystems and correct segmentation problems.

GeoMiner (Han, Koperski & Stefanovic, 1997), developed at Simon Fraser University (Canada), is a prototype of spatial data mining system, with resources to characterize spatial data through rules, compare, associate, classify and group datasets, analyze patterns and perform mining tasks in differente levels. The prototype has a language for mining tasks of spatial data (GMQL), beyond visualization tools for data and spatial mining results. GeoMiner is integrated to data warehousing technology, and it is able to access different spatial database servers.

SPIN! project (May & Savinov, 2002), developed by the Fraunhofer Institute for Autonomous Intelligent Systems (Germany), is focused on producing a spatial data mining system that integrates Geographic Information Systems and data mining in a open, extensible and tightly coupled framework. The project prioritizes issues like scalability, security, multiuser access, robustness, and platform independence. Its functionality levels include data access and management, interactive thematic mapping for statistic data visualization, detection and explanation of spatial clusters and spatial events.

ADaM, a NASA's project with the University of Alabama at Huntsville (USA), is a set of scientific data and image mining tools (Rushing, Ramachandran, Nair, Graves, Welch & Lin, 2005). Its resources include pattern recognition, image processing, optimization, association rule mining, among others. The system is a set of components that may be put together to perform complex tasks. A focus of the project is the efficient implementation of critical performance components, keeping each component of the system as independent as possible, in order to enable the use of appropriate module subsets to specific applications, including linking to third party software.

Position about the Technology

Such initiatives, among other important projects, led by institutions and researchers of different countries since 1980, demonstrates the relevance, strength and demand for efficient and robust approaches, once the mining process on image repositories demand a strong commitment with efficiency and robustness. The huge volume of the datasets need an efficient hardware and software infrastructure. The relativity of values, the spatial complexity, and the multitude of interpretations require robust implementations, competent domain specialists and experient data analysts for the mining task performances.

However, still a limited capacity is available for extracting information from large remote sensing image databases. Currently, most image processing techniques are designed to operate on a single image, and there are few algorithms and techniques for handling multitemporal images. This situation has lead to a "knowledge gap" in the process of deriving information from images and digital maps (MacDonald, 2002). This "knowledge gap" has arisen because there are currently very few techniques for image data mining and information extraction in large image data sets, and thus researchers are failing to exploit the huge remote sensing data archives.

Although there has been a large research effort in content-based image retrieval (CBIR) techniques (Rui, Huang & Chang, 1999; Smeulders, Worring, Santini, Gupta & Jain, 2000; Wang, Khan & Breen, 2002), the specific problem of mining remote sensing image databases has received much less attention. Proposals such as VISIMINE (Aksoy, Koperski, Tusk, & Marchisio, 2004) and KIM (Schröder, Rehrauer, Seidel& Datcu, 2000) are focused on clustering methods that operate on the feature space, the multidimensional space which is created by the different spectral bands of a remote sensing image. These techniques are useful for distinguishing spectral signatures of different land cover types, such as finding areas which are classified as "lakes," "cities" or "forests."

Nevertheless, in remote sensing image mining, one of the most important challenges is tracking patterns of land use change. A large remote sensing image database is a collection of snapshots of landscapes, which provide a unique opportunity for understanding how, when, and where changes take place in the world. Extensive fieldwork also indicates that the different actors involved in land cover change (e.g., small-scale farmers, large plantations, cattle ranchers) can be distinguished by their different spatial patterns of land use (Lambin, Geist & Lepers, 2003). Furthermore, these patterns evolve in time; new small farms will be created and large farms will increase their agricultural area at the expense of the forest. In these and related situations, patterns of land use change will have similar spectral signatures and image mining techniques based on clustering in the feature space will not be able to distinguish between them. Therefore, tracking the temporal evolution of patterns in remote sensing imagery requires methods that are different from standard content-based image retrieval (CBIR) systems. A typical CBIR system uses a query image as the source and images in the database as targets, and query results are a set of images sorted by feature similarities with respect to the source (Chen, Wang & Krovetz, 2003). When searching for patterns in remote sensing image databases, a different approach is necessary. Instead of similarity searches between image pairs, a system for mining remote sensing image databases must be able to do similarity searches between patterns found in different images. Therefore, mining remote sensing image databases is searching for patterns of change, not searching for internal content.

CHALLENGES AND TECHNOLOGICAL STRATEGIES ON DEFORESTATION ISSUE

Brazil's Challenge: Monitor and Decrease Amazonia Deforestation

The *land cover* describes the physical state of the land surface, which may be forest, water, buildings, and so on. Changes on this cover may be caused by climate variations, changes on river courses, and so on. However, most changes on land cover are attributed to human activities. Such modifications implies on changes on the extension (area increase or decrease) of a specific type of coverage. The *land use*, influenced by human activities and enviromental processes and features, is related to the purpose to which it is used, like agriculture, habitation, mining, leisure, among others. Land use changes occur in several spatial levels and in different periods, characterizing the environment and human dynamics on territorial segments (Briassoulis, 2004).

Desertification, climate change, biodiversity loss—among others—can imply in severe consequences to the environment and consequently to humans. The modification of forest and crop areas for urban use is an important land change, due to serious implications. The causes and consequences of land use and cover change, its social, economics and enviromental impacts have motivated different research projects. One of them is (Lambin, 1999), which emphasizes that land cover change is an important global change factor, interacting with climate, ecosystem processes, biochemical cycles, biodiversity and even with human activities. The key issues of the project deal with land cover patterns, change processes, human response to changes, integrated global and local models, development of databases about Earth surface, biophysics processes and fundamental factors. This approach aims to increase the understanding, and get new knowledge about interactive land change.

The Amazonia case is characterized by the complexity, dimension, and interests involved in the issues concerning land change (Becker, 1997). Alves (2002) presents an investigation on spatio-temporal deforestation dynamics of the Amazonia, using remote sensing images to analyze deforestation spatial patterns on 1970's, and between 1991 and 1997. This work brings valuable information: the deforested area increased from 10,000,000 ha (1970's) to 59,000,000 ha in 2000; an intensification on the deforestation rate on 1970's and 1980's was caused by the federal government politics, which included huge highway infrastructures, and a roadside colonization of 100 km along the extended highways; analyzing the images and the patterns, it is clear that beyond of the roadside deforestation along main roads and development areas, there is still the merging of little deforested areas, what originates large ones.

Once the fast deforestation process causes land degradation, social tension and irregular urbanization, faster the precise identification of areas with these tendencies, higher the chances of preventing, managing and reducing the consequences of the processes. Daily, different satellites capture data belonging to this context, which images are available to many institutions. Image mining tools can, in fact, increase the analysis potential of such huge strategic data.

Developed Technologies at INPE Concerning Image Analysis and Mining

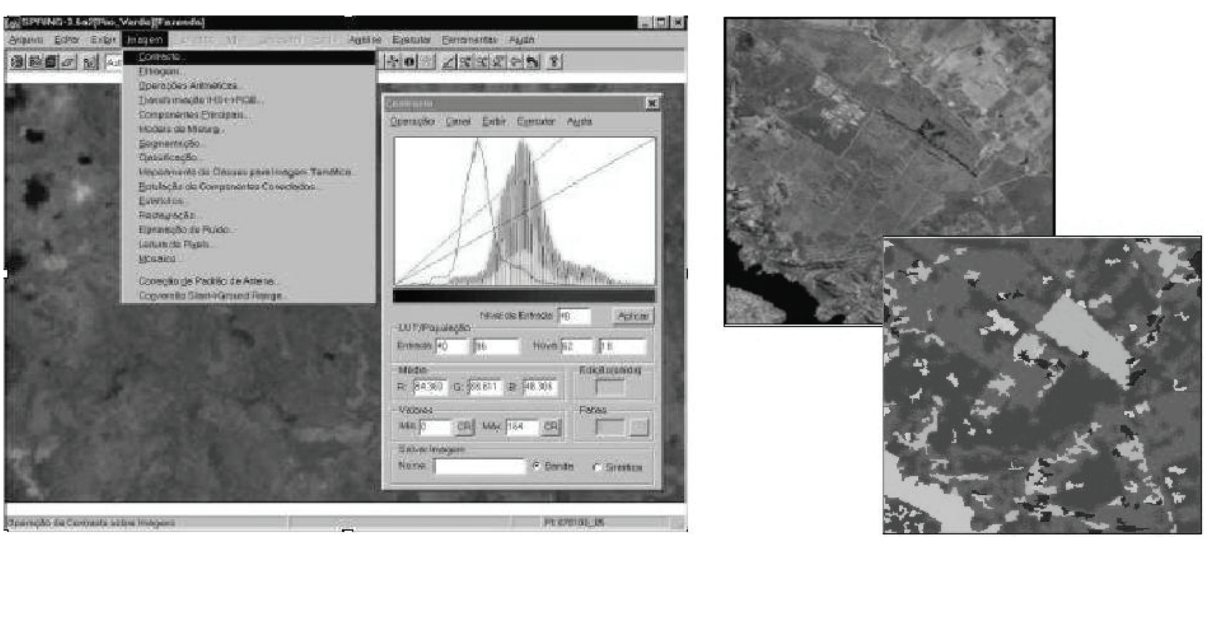
Researchers of the Brazil's National Institute for Space Research (INPE) has been studying the structural patterns on Amazonia, holding a wide know-how on the forest issues. Moreover, the historical development process is also a research topic at INPE, which maintains a rich dataset of remote sensing images that provide an extensive spatiotemporal perspective of the Amazonia territory. In addition, the Institute experience on image processing and analysis, as well the development of methodologies and software tools, supplies important elements to keep building up image analysis and mining technologies. In this context, relevant ones developed at INPE are: SPRING, TerraLib, CBERS, PRODES and DETER, which are freely available on Internet.

SPRING (www.dpi.inpe.br/spring) is a state-of-the-art geographic information system (GIS) and remote sensing image processing system with an object-oriented data model which provides the integration of raster and vector data representations in a single environment (Figure 2). SPRING main features include: an integrated GIS for environmental, socioeconomic and urban planning applications; a multiplatform system, including support for Windows and Linux; a widely accessible freeware for the GIS community with a quick learning curve. The software is a mechanism of diffusion of the knowledge developed by INPE and its partners with the introduction of new algorithms and methodologies (Câmara, Souza, Freitas & Garrido 1996).

TerraLib (www.dpi.inpe.br/terralib) is a GIS classes and functions library, available from the Internet as open source, allowing a collaborative environment and its use for the development of multiple GIS tools. Its main objective is to enable the development of a new generation of GIS applications, based on the technological advances on spatial databases. TerraLib is free software developed by INPE and its partners. The main motivation for this project is the current lack of either public or commercial GIS libraries that provide components for the diversity of GIS data and algorithms, especially when viewed upon the latest advances in geographical information science. On a practical side, TerraLib enables quick development of custom-built geographical applications using spatial databases. As a research tool, TerraLib is aimed at providing a rich and powerful environment for the development of GIScience research, enabling the implementation of GIS prototypes that include new concepts such as spatio-temporal data models, geographical ontologies and advanced spatial analysis techniques (Câmara, Vinhas, Souza, Paiva, Monteiro, Carvalho, 2001).

The *CBERS* program (http://www.cbers.inpe.br/en/index_en.htm), a joint effort of Brazil and China, embodied the development and construction of two remote sensing satellites that carry on-board imaging cameras and additionally a

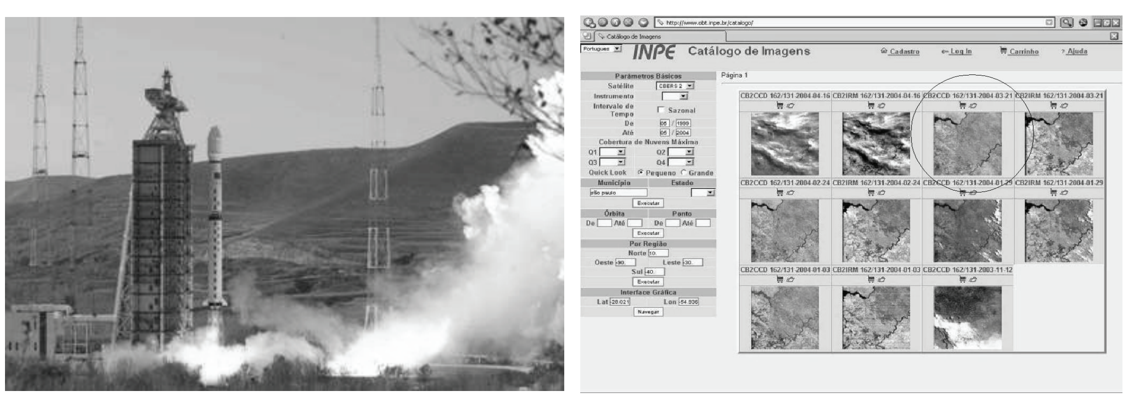
Figure 2. SPRING - image processing and geographic information system



repeater for the Brazilian System of Environmental Data Collection. CBERS-1 and CBERS-2 are identical in their technical structure, space mission and payload (on-board equipment like cameras, sensors, computers, among other equipment designed for scientific experiments). CBERS-1 was launched by the Chinese Long March 4B launcher from the Taiyuan Launch Base on October 14, 1999. CBERS-2 was launched on October 21, 2003 from the Taiyuan Satellite Launch Center in China (Figure 3). CBERS-2 was integrated and tested in the integration and test laboratory of INPE. The CBERS satellite has a set of sensors – WFI (wide field imager), CCD (charge coupled device high resolution imaging camera), IRMSS (infrared multispectral scanner) - with a high potential to meet multiple application requirements including: forestry alteration, signs of recent fires, monitoring of agricultural development, support for crop forecasting, identification of anthropic anomalies, analysis of natural recurrent events, mapping of land use, urban sprawling, identification of water-continent borders, coast studies and management, reservoir monitoring, acquisition of stereoscopic images for proper cartographic analysis, support for soil survey and geology, generation of support material for educational activities. In 2002, both governments decided to expand the initial agreement by including CBERS-3 and 4, which must be launched, respectively, in 2008 and 2012. The program objectives are: build a family of remote sensing satellites to support the needs of users in Earth resource applications, and improve the industrial capabilities of space technology in Brazil.

Since 1988 INPE has been monitoring Brazilian Amazonia using satellite images, producing estimations on annual deforestation rates of the forest through the *PRODES* project (Amazônia deforestation calculation program). From 2002 on, these estimations are being generated by image digital classification with PRODES methodology (www.obt.inpe.br/prodes/). The main advantage of this approach is the precision of georeferenced deforestation polygons, enabling a multitemporal geografic database. Using deforestation increments identified on each image, the annual rates are estimated for August 1st of the reference year. For the 2003/2004 period, the deforestation rates were obtained from 207 LANDSAT images; INPE estimates that the deforestation from August 2003 to August 2004 was 27.429 km^2. For the 2004/2005 period, the deforestation rates were obtained from 211 LANDSAT classified images; INPE estimates that the deforestation from August 2004 to August 2005 was 18.793 km^2. The Institute estimates a deforested area of 13,100 km^2 for the 2005/2006 period. PRODES digital results of 2000 until 2004 are available on SPRING databases containing LANDSAT satellite images, thematic map of the

Figure 3. CBERS-2 - Launch and Web interface of image catalog (source: www.cbers.inpe.br)



deforestation of the year, thematic map of the accumulated deforestation, and the shapefiles of the year with polygons of deforestation increment of the year, forest, total accumulated deforestation until the previous year, clouds and non-forest. From 2005 on, it is also available on the shapefile of the deforestation thematic map of the year for each LANDSAT image, and the shapefile of the mosaic of all images.

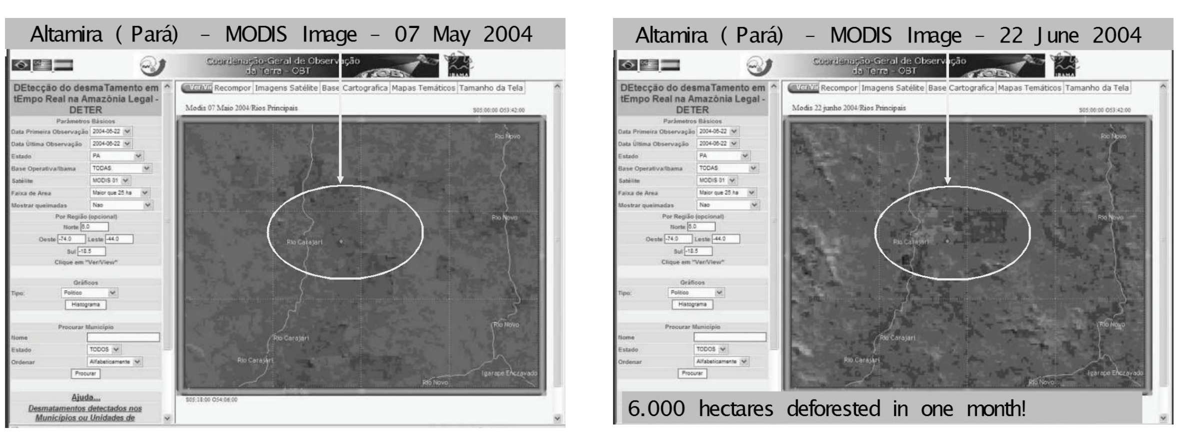
The *DETER* system (deforestation detection on real time) uses sensors with high observation frequency to reduce cloud cover limitations during the process of detecting deforestation increments (www.obt.inpe.br/deter). The instruments used are the MODIS sensor, aboard TERRA and AQUA satellites (NASA), with a spatial resolution of 250 m and temporal resolution (Brazil) of three to five days, and the WFI sensor, aboard CBERS-2, with a spatial resolution of 260 m and temporal resolution of five days. These resolutions enable the detection of recent deforested areas superior to 0,25 km^2. The results of the methodology—which produces information in almost real time about regions where new deforestation areas occur—allow DETER supplies environment surveillance institutions with periodic information about deforestation events (Figure 4). The goal of the system is not the estimation of total deforestated area in Amazonia, once estimations obtained through DETER are error-prone due to the spatial resolution of MODIS and WFI. The system is concerned on supplying recent and updated information to support government actions against the forest destruction using a higher temporal resolution of the sensors. The system is an INPE project, part of a federal plan of reducing Amazonia deforestation.

INPE's Effort to Spread Earth Observation Technologies

There is a need for global land observation advance, once the world is changing rapidly. Global land observation is a crucial need for the world, and Earth observation (EO) systems are a public good. INPE's effort on advanced policies of development of state-of-the-art software, hardware, methodologies and products relies on the need of building capacity in EO to supply the wide demand of the area.

Build capacity in Earth observation implies on removing the barriers to make all sectors of society use publically funded EO data. Three relevant obstacles are: lack of data (much EO data is expensive or unavailable), lack of tools (once good software is required to explore EO data), and lack of expertise (it is necessary to build capacity

Figure 4. Amazonia deforestation process detected by DETER



at a massive scale). INPE's approach to overcome such barriers are: make EO data free, produce good open source software for EO data handling, and provide open access to on-line training and to scientific literature.

The Internet has reduced the cost of data distribution to very close to zero, and society responds very quickly to open availability of free data and good on the Web. CBERS images received in Brazil are freely available on the Internet for Brazilian and Latin American users, and CBERS images received in China are freely available on the Internet for Chinese users (www.cbers.inpe.br). Free EO data and free EO technology create new users and new applications, increasing the need for other types of EO data. Private companies, for example, state the free CBERS data benefit: enables new business development, facilitates trial uses for new clients, creates jobs by reducing cost of data buys, increases work quality by adding data previously unavailable, and eases the planning of new applications.

Commercial EO market in many countries does not have enough income to research and development investment, once it is still a small size market. To let it grow, it is necessary to supply improvements on information extraction through high-quality software. Concerning the tool challenge, INPE developed GIS and image-processing softwares (TerraLib and SPRING) available free on the Internet, providing good software for EO data handling (www.dpi.inpe.br/spring; www.dpi.inpe.br/terralib).

The research system on EO in the developed world discourages the production of training material, once academic institutions in US and Europe graduate qualified personnel and there are good books on GIS and remote sensing (unfortunately, these books are in English and are expensive). Developing countries need innovative responses, especially good training material and on-line books. Brazilian experience is overcoming the expertise challenge releasing free books online, a three-volume set: *Introduction to GIS*, *Spatial Analysis*, and *Spatial Databases* (http://www.dpi.inpe.br/gilberto/livros.html).

INPE is focusing on the "white-box" model: *results = people + data + software*. This means support for people learning by doing and using, timely and free geospatial datasets, and adequate data analysis and integration softwares. The results: an enormous demand for remote sensing data in developing countries, a relevant increase on the number of users of Earth observation data due to free online data access, and the success of CBERS data policy that has been extremely well-received by government and society in Brazil.

DETECTING DEFORESTATION PATTERNS THROUGH SATELLITES

Patterns of Change in Remote Sensing Image Databases

Given a large remote sensing image database, researchers would like to explore the database with questions such as: What are the different land use patterns present in the database? When did a certain land use pattern emerge? What are the dominant land use patterns for each region? How do patterns emerge and change over time? The answer to these and similar questions requires the availability of data mining techniques which are able to perform searches for patterns found in different images. Silva (2006) approached this problem by using spatial patterns as a mean of describing relevant semantic features of an image.

The primary consideration is that the instruments onboard remote sensing satellites capture energy at different parts of the electromagnetic spectrum, which is then converted into digital imagery. These instruments are not designed for a specific application, but are a compromise between sensor technology and requirements from different user communities. As a result, remote sensing images have a structural description which is independent of the application domain that a scientist employs to extract information. The *image domain* and the *application* domain are distinguished, as shown in Figure 5:

- **Spatial patterns:** The geometric structures that can be extracted from the images using techniques for feature extraction, segmentation, and image classification. They must be identified and labeled according to a typology which expresses their semantics. Examples of such patterns include corridor-like regions and regular-shaped polygons representing patterns of the mined data.
- **Application concepts:** The different classes of spatial objects, which are associated to a specific domain. For example, in deforestation assessments, concepts include large-scale agriculture, small-scale agriculture, cattle ranching, and wood logging.

To associate structures found in the image to concepts in the application, there is a *structural classifier*, which is able to relate the same structures to different application domains. This strategy differs from most remote sensing image database mining systems, such as KIM (Schröder et al., 2000) and VISIMINE (Aksoy et al., 2004), which implicitly assume that there is one "best fit" for associating semantic concepts in the user domains to image-derived structures. In this approach, different structural classifiers will produce different associations between spatial patterns and the user domain concepts, and each association is valid within a given application context. In other words, there are many ways to bridge the "sensory gap" and a "best fit" should not be searched. For each type of application, there will be an appropriate structural classifier.

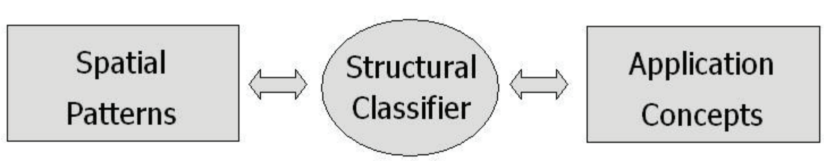
In what follows, the methodology for image mining is described and applied to the problem of mining patterns in INPE's remote sensing image database. In this context, the application domain is concerned with describing land use change in tropical forests using remote sensing satellites.

Methodology for Mining Land Use Patterns on Remote Sensing Images

The methodology for image mining in large remote sensing databases uses the application-dependent structural classifier, as outlined previously. The methodology consists of three steps:

- Definition of a spatial pattern typology according to the user's application domain (Figure 6).
- Building a reference set of spatial patterns. This reference set is built using a prototypical set of images. Landscape objects are identified and labeled: the identification employs image segmentation and the labeling is performed according to the spatial pattern typology (Figure 7).
- Mining the database using a structural classifier (guided by the application concepts of the domain), matching the reference set of spatial patterns to the landscape objects identified in images, thus revealing the spatial configurations present in each image (Figure 9).

Figure 5. Overview of pattern mining process



Defining a Spatial Pattern Typology

The first phase of the methodology calls for the definition of a spatial pattern typology which is associated to a given application domain (Escada, Monteiro, Aguiar, Carneiro & Câmara, 2005). In order to illustrate the proposal, a typology defined for mapping different types of land use change in tropical forests will be used.

When using remote sensing images for understanding the forces driving changes in tropical forests, the assumption is that the expression of change is captured by changes in land use. Extensive fieldwork also indicates that the different actors involved in land use change (small-scale farmers, large plantations, cattle ranchers) can be distinguished by their different patterns of land use (Lambin, Geist & Lepers, 2003). They propose a typology of the land use patterns in terms of deforestation processes (see Figure 6): corridor (commonly associated with riverside and roadside colonization), diffuse (generally related to smallholder subsistence agriculture), fishbone (typical of planned settlement schemes), and geometric (frequently linked to large-scale clearings for modern sector activities).

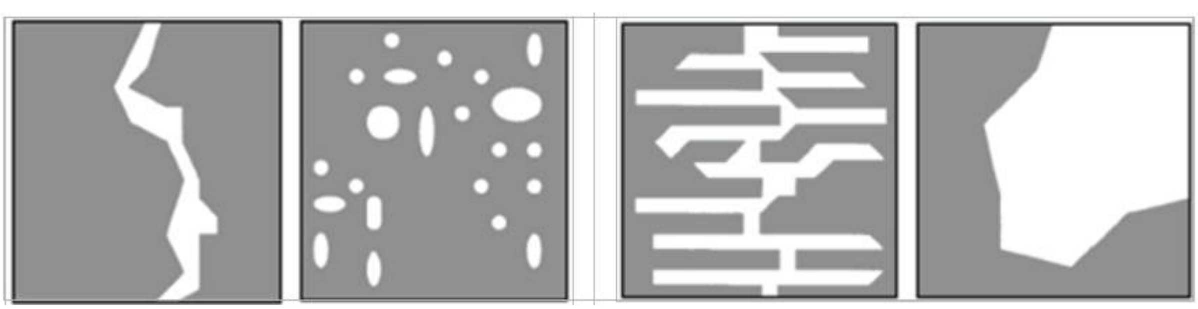
Three spatial patterns typology of Lambin will be used (corridor, diffuse, geometric), relating them to the structures of landscape objects in order to obtain the spatial patterns, through a cognitive assessment process, in which a human specialist associates landscape objects to spatial patterns typology elements.

Building a Reference Dataset of Spatial Patterns

To represent the structures detected in remote sensing images, the concept of a landscape object will be introduced. A landscape object is a structure detected in a remote sensing image by means of an image segmentation algorithm. Landscape objects can be associated to different types of spatial patterns.

To build a reference set of *spatial patterns* (Figure 7), a set of prototypical landscape objects is obtained, which are extracted from a set of sample images. Segmentation algorithms are used to partition the image into regions which are spatially continuous, disjoint and homogenous. Recent surveys (Meinel & Neubert, 2004) indicate that region-growing approaches are well suited for producing closed and homogenous regions. In this proposal, it is adopted the region-growing segmentation algorithm developed by INPE (Bins, 1996), and implemented in the SPRING software system (Câmara, 1996). This algorithm has been extensively validated for extracting land use patterns in tropical forests (Shimabukuro et al., 1998) and has been very favorably reviewed in a survey (Meinel & Neubert, 2004).

Figure 6. Spatial patterns of tropical deforestation (from left to right): corridor, diffuse, fishbone, and geometric (Source: Lambin, Geist & Lepers, 2003)



SPRING's region growing algorithm works as follows (Figure 8) (Bins, 1996): (a) the image is first segmented into atomic cells of one or few pixels; (b) each segment is compared with its neighbors to determine if they are similar or not. If similar, they are merged and the mean gray level of the new segment is updated; (c) the segment continues growing by comparing it with all the neighbors until there is no remaining joinable region, at which point the segment is labeled as a completed region; (d) the process moves to the next uncompleted cell, repeating the entire sequence until all cells are labeled. The algorithm requires two parameters: a similarity threshold value, and an area threshold value.

Mining the Database Using a Structural Classifier

Once the reference set of *spatial patterns* is built, the next phase will use it to mine *spatial configurations* from image databases. The *structural classifier* enables the association between landscape objects extracted from images and the reference set of *spatial patterns* (Figure 9).

Figure 7. Building a reference set of spatial patterns

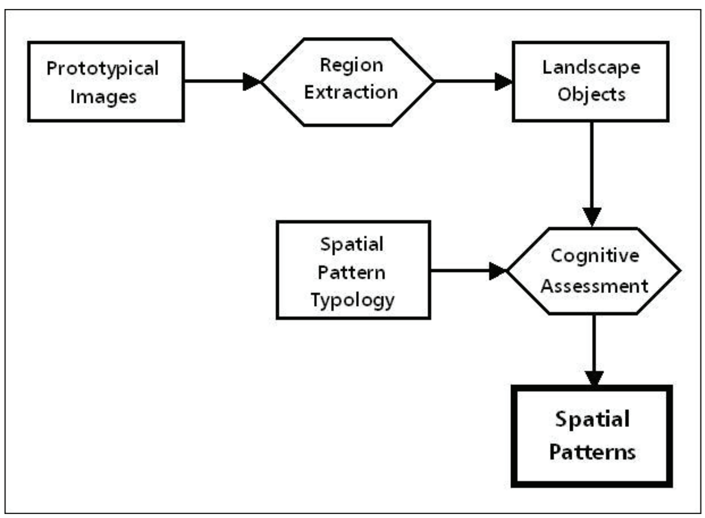


Figure 8. Example of a segmentation process

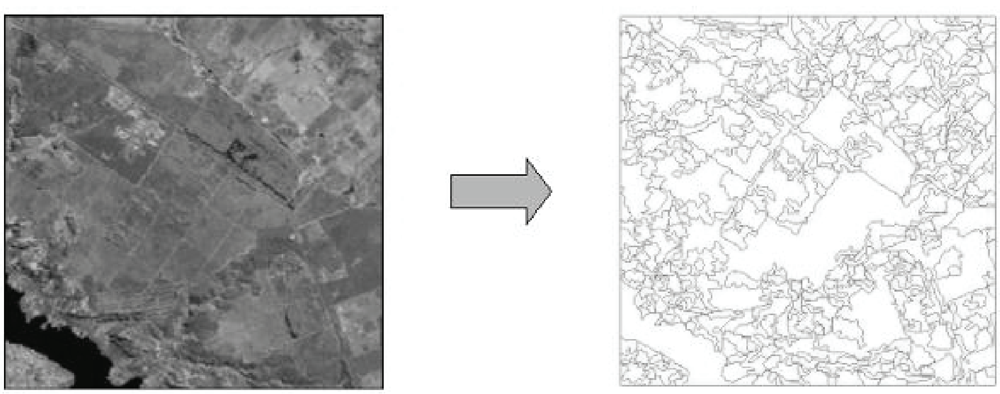
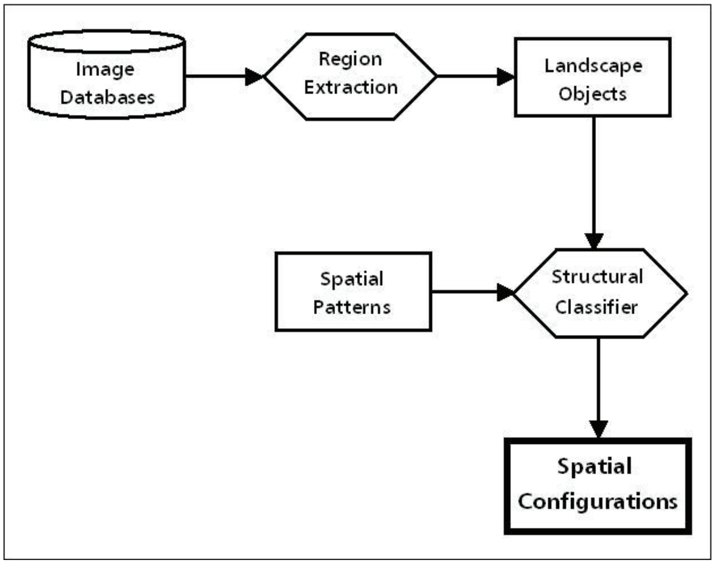


Figure 9. Obtaining spatial configurations



The structural classifier must be able to distinguish between different spatial patterns. It uses the C4.5 classifier (Quinlan, 1993), a classification method based on a decision tree. It predicts the value of a categorical attribute (Witten & Frank, 1999) based on noncategorical attributes. The categorical attribute is the pattern type and the noncategorical attributes are a set of numerical attributes that characterize each pattern.

To select the attributes that distinguish the different types of land use patterns, the concepts from landscape ecology (Turner, 1989) are used. Landscape ecology is based on the notion that environmental patterns strongly influence ecological processes. One of the key components of landscape ecology theory is the definition of metrics that characterize geometric and spatial properties of categorical map patterns (McGarigal, 2002). The pattern metrics used in landscape ecology include metrics of spatial configuration that operate at the patch level. Patches form the building blocks for categorical maps and within-patch heterogeneity is ignored. Patch metrics refer to the spatial character and arrangement, position, or orientation of patches within the landscape. The pattern metrics proposed by the FRAGSTATS (Spatial Pattern Analysis Program for Categorical Maps) software (McGarigal & Marks, 1995) are used, which include:

- *Perimeter* (m) and *area* (ha).
- *Para* (perimeter-area ratio): A measure of shape complexity.
- *Shape* (shape index): Patch perimeter divided by the minimum perimeter possible for a maximally compact patch of the corresponding patch area.
- *Frac* (fractal dimension index): Two times the logarithm of patch perimeter (m) divided by the logarithm of patch area (m^2).
- *Circle* (related circumscribing circle): 1 minus patch area (m^2) divided by the area (m^2) of the smallest circumscribing circle.
- *Contig* (contiguity index): Equals the average contiguity value for the cells in a patch.

The landscape ecology metrics are fed into the C4.5 classification algorithm to distinguish the different types of spatial patterns. After this classifier is properly trained, it can be used to label the landscape objects found in other images. Therefore, for each image in the database, this procedure identifies the number and location of the different types of spatial patterns. A specific set of spatial patterns found in an image is referred as a *spatial configuration*.

By identifying the spatial configurations of different images, the user will be able to evaluate the emergence and evolution of different types of change. Each spatial pattern is associated to a different type of land use change. Therefore, the comparison between spatial configurations of images in different locations and between spatial configurations of images at the same location in different times will allow new insights into the processes and actors that bring about change.

Case Study: Image Mining for Deforestation Patterns

Controlling deforestation on Amazon rain forest is a difficult challenge for Brazil, once the causes of deforestation include economic, social and political factors, and the current pace of land use change is substantial, with a deforested area of about 200,000 km^2 during the decade 1995-2005.

Table 1. Land use change in tropical forests

Landscape object	Land use change
Corridor pattern	Roadside colonization Riverside deforestation
Diffuse pattern	Smallholder agriculture Small deforestation increments
Geometric pattern	Large farms

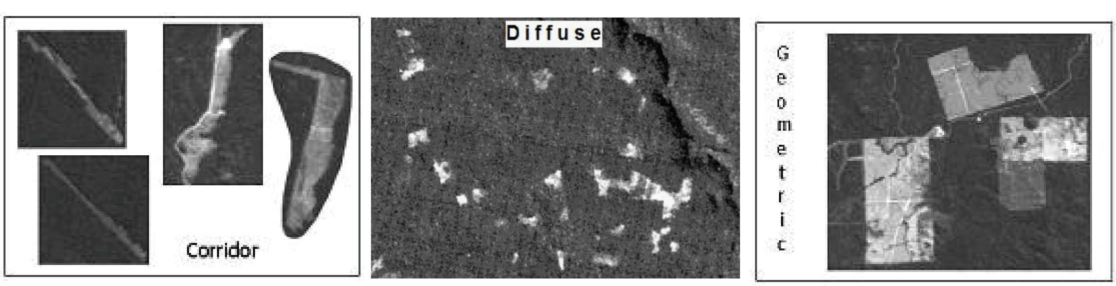
The situation demands fast and effective actions for reducing this pace of devastation. In order to monitor the extremely fast process of land use change in Amazonia, it is very important that INPE be able to use its huge data archive to the maximum extent possible. In this context, the image mining methodology was used to achieve a better understanding of the processes of land use change in Amazonia.

A case study was developed using Landsat TM images (225/64, 226/64, 226/65, 225/65) of 1997, 2000, 2001, 2002 and 2003, which cover the region of São Félix do Xingu in the state of Pará. This is a region with many violent land conflicts and one of the largest annual rates of deforestation in Amazônia (INPE, 2005). The main land use activity developed in São Felix do Xingu is cattle ranching, which holds around 10% of the cattle of Pará state (Américo, Vieira, Veiga & Araujo, in press). Deforestation in the region has two main agents: migrants, that have settled in small areas, and large cattle ranchers, many of whom have occupied land illegally (Escada, Vieira, Amaral, Araújo, Veiga, Aguiar, & Veiga, 2005). The images and deforestation data were provided by PRODES Project (INPE, 2005). The application concepts for this task are guided by the land use change domain in tropical forests (Table 1).

Building Spatial Patterns

According to the image mining methodology, landscape objects were extracted from prototypi-

Figure 10. Spatial patterns representing corridor, diffuse, and geometric patterns



cal images. Then, a human specialist, through cognitive assessment, obtained *spatial patterns* based on the spatial patterns typology of tropical deforestation (Figure 6). Spatial patterns are presented in Figure 10.

Obtaining Spatial Configurations

The *structural classifier*, using the *spatial patterns*, extracted *spatial configurations* from the set of images just mentioned. Results are presented below.

In a first case, it is necessary to answer the following question: "What's the behavior of large farmers in São Félix do Xingu during 1997-2003 period? Is the area of new large farms increasing?" Observing the evolution of the corresponding *spatial configuration* (geometric patterns - GEOM) in Figure 11, it was possible to conclude that, "in 2000, this kind of deforestation reached a peak of 55,000 ha, but decreased in the following years. In 2003, the deforestation area associated to large farms decreased to 29,000 ha. This indicates that large farms are reducing their contribution to deforestation."

Figure 11. Large farms dynamic in São Félix do Xingu

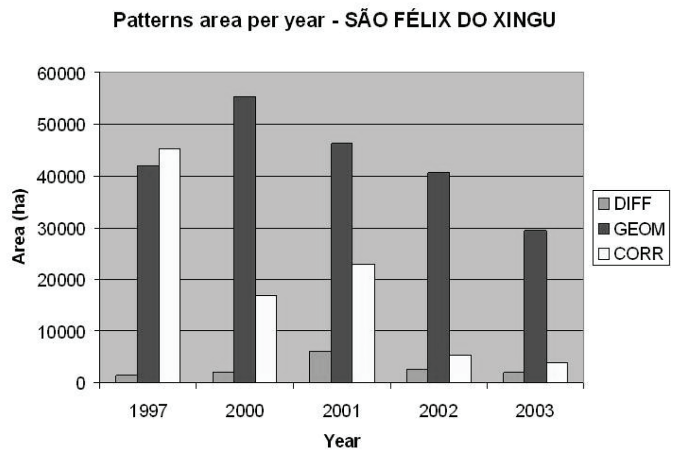
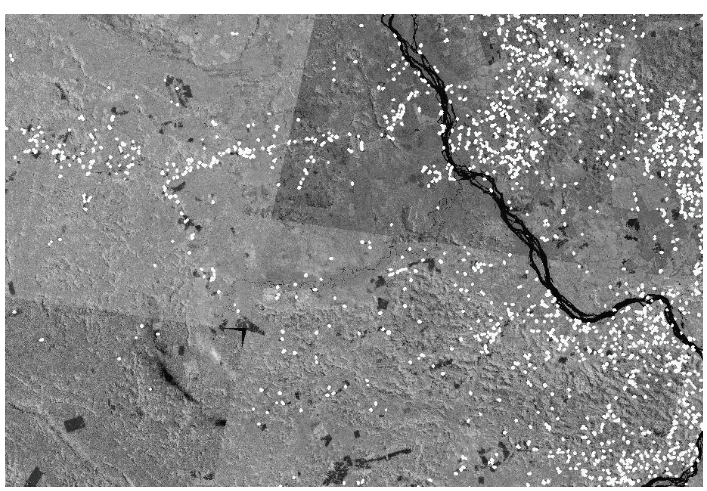


Figure 12. Diffuse pattern in São Félix do Xingu 1997-2003



There is a second question: "What's the distribution of smallholder agriculture and small deforestation increments in São Félix do Xingu area during the years 1997-2003?" Observing Figure 12, it is possible to conclude: "the distribution of this land use pattern (diffuse) in this period was mainly concentrated in the northeast and southeast of this area."

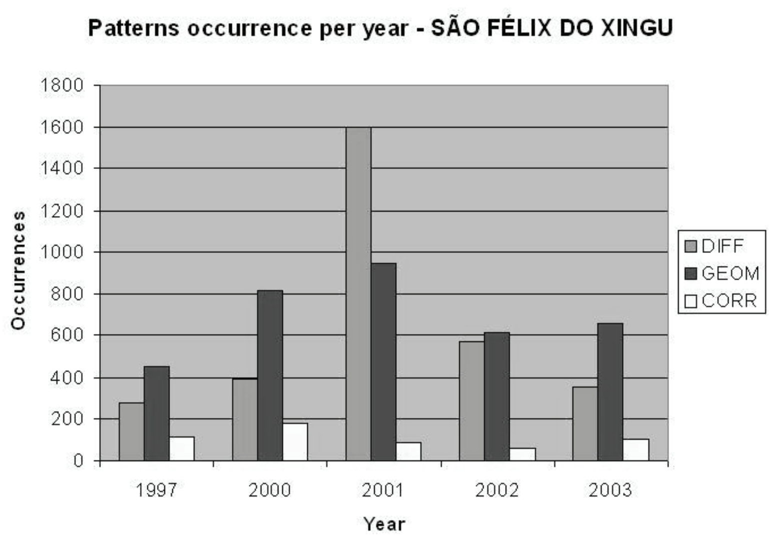
The next question is: "In São Félix do Xingu region, is there any dominant land use change pattern?" Observing Figure 13, the conclusion is: "Diffuse pattern represented 61% of total occurrences of land use changes in 2001, indicating an increase in smallholder agriculture / small increments in deforested areas in that year."

FUTURE TRENDS

A consortium of Earth observation satellites for global land monitoring, a network of cooperating ground stations, EO data free on the Internet with global weekly coverage, satellite sensor resolution improvements and the availability of web services to perform image mining tasks will provide necessary resources for new applications and a wide range of demands, specially in developing countries. Moreover, hardware and software performance increase will support mining processes on huge and improved image datasets, allowing a more intensive and extensive use of satellite image mining in strategic fields like forestry and reservoir monitoring, agricultural expansion, soil survey, analysis of natural phenomena, and urban studies.

Future research directions in remote sensing image mining include tracking individual trajectories of change. Patterns found in one map are linked to those in earlier and later maps, thus enabling a description of the trajectory of change of each landscape object. The current method aggregates landscape objects of the same type. A more sophisticated approach would be to describe the evolution of each landscape object, including operations such as merging of adjacent regions. This description would allow the image-mining tool to describe when two irregular areas of land use (associated to small settlers) were merged. It would also show when the merged region was extended with a regular pattern (suggesting that

Figure 13. Diffuse patterns in São Félix do Xingu



a large cattle ranch had been established). This description could increase even more the ability to understand the land use changes that are detectable in remote sensing image databases.

CONCLUSION

This chapter presents relevant issues on satellite image mining, describing a method for mining patterns of change that enables extracting spatial arrangements from remote sensing image databases. It addresses the problem of describing land use change. It combines techniques from data mining, digital image processing and landscape ecology to identify patterns in images of distinct dates. The method points out that patch metrics can be used to identify agents of land use change. Images of distinct dates enabled the detection of pattern changes, which are extremely valuable when assessing, managing or preventing deforestation processes.

This methodology enables associating land change objects to causative agents, and it can assist the environmental community to respond to the challenge of understanding and modeling relevant issues in a rapidly changing world. The results from the case study show that image-mining techniques are a step forward in understanding and modeling land use and cover change. The proposed method also enables a more effective use of the large land remote sensing image databases available in agencies such as USGS, ESA and INPE.

The remote sensing image-mining process is an interactive one; once it demanded the sample selection, model building and rating, context evaluation, return to specific points of the process, among others. During experiments, the result evaluation in different phases demonstrated the need of new prototype objects, better model calibration, or even adjustments on the spatial pattern typology. Once provided such topics, relevant results were obtained and validated through extensive fieldwork.

Taking into account the heterogeneity of the Amazonia context, a relevant (and expected) question is the fact that the model training and application must be performed in spatially similar regions, that is, train the structural classifier in a specific region and apply it in another region, with different spatial features, causes the generation of inconsistent results. Another methodology limitation concerns the quantity and quality of prototype objects used to generate the model for structural classification. If the number of elements or their description ability to distinguish patterns is not appropriate, the generated model (decision tree) will classify inconsistently many objects. The methodology also demands a proper spatial pattern typology, which must characterize the spatial patterns and the semantic aspects that must be detected during the process.

The mining process requires a domain specialist, due to the intense Amazonia dynamics, especially on the prototype object selection and during the spatial configuration interpretation. Further experiments are necessary to improve the method, to test alternatives for image segmentation algorithms and for pattern classifiers. The limitations of the current method are also associated to the two-dimensional nature of land use maps. An extension of the method would combine spatial information (patch metrics) with spectral information (pixel and region trajectories in multitemporal images).

Uncle Scrooge principle states that, "a penny saved is a penny earned." However, the anti-Uncle Scrooge principle reveals that, "a pixel saved is a penny wasted." Why is that so? Because "value comes from use." Coherent EO programs can supply strategic components for the enormous demand of remote sensing data, expertise, and analysis tools in developing countries. This work resources may help to leverage the power of detecting, evaluating and reducing the pace of Amazonia deforestation, once INPE holds know-how and a wide spatiotemporal coverage of the forest. Moreover, the present technology can be ported to provide solutions to a broad range of image mining applications.

REFERENCES

Américo, M. C. S., Vieira, I. C. G., Veiga, J. B. & Araujo, R. (in press) Pecuária e Amazônia: Estratégias sociais e reestruturação do território nas frentes pioneiras: Rodovia PA-279 e região da Terra do Meio no Pará [Cattle ranching and Amazonia: Social strategies and territory reorganization in new frontiers – PA-279 and Terra do Meio region in Pará state]. In R. Araujo & P. Lena (Eds.), *Alternativas de desenvolvimento sustentável na Amazônia: Experiências recentes* [Alternatives of sustainable development in Amazônia: Recent experiences].

Aksoy, S., Koperski, K., Tusk, C. & Marchisio, G. (2004). Interactive training of advanced classifiers for mining remote sensing image archives. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining* (pp. 773-782). Seattle, Washington.

Alves, D. S. (2002). Space-time dynamics of deforestation in Brazilian Amazonia. *International Journal of Remote Sensing, 23*(14).

Bins, L., Fonseca, L. & Erthal, G. (1996). Satellite imagery segmentation: A region growing approach. In *Proceedings of the 8th Brazilian Symposium on Remote Sensing* (pp.1-4).

Becker, B. (1997). *Amazonia*. São Paulo: Atica.

Briassoulis, H. (2004). *Analysis of land use change: Theoretical and modeling approaches.* Retrieved April 8, 2008, from http://www.rri.wvu.edu/WebBook/Briassoulis

Câmara, G., Souza, R., Freitas, U. & Garrido, J. (1996). SPRING: Integrating Remote Sensing and GIS with object-oriented data modelling. *Computers and Graphics, 15*(6), 13-22.

Câmara, G., Vinhas, L., Souza, L., Paiva, L., Monteiro, A., Carvalho, M. & Raoult, B. (2001). Design patterns in GIS development: The Terralib experience. In *Proceedings of the III Brazilian Symposium in Geoinformatics, GeoInfo 2001*, Rio de Janeiro.

Canada Centre for Remote Sensing (2003). *Fundamentals of remote sensing*. Remote Sensing Tutorial (pp. 5-44). Retrieved April 8, 2008, from www.ccrs.nrcan.gc.ca/ccrs/learn/tutorials/fundam/fundam_e.html

Chen, Y., Wang, J. Z. & Krovetz, R. (2003). CLUE: Cluster-based retrieval of images by unsupervised learning. In K. A. Meraim, I. Bloch (Eds.), In *Proceedings of the IEEE Seventh International Symposium on Signal Processing and its Applications* (pp. 202-231).

Escada, M. I. S., Monteiro, A. M., Aguiar, A. P., Carneiro, T. & Câmara, G. (2005). Análise de padrões e processos de ocupação para a construção de modelos na Amazônia [Analysis of land use patterns and processes for the construction of models in Amazonia]. In *Proceedings of the XII Brazilian Symposium on Remote Sensing* (pp. 2973-2983), Goiania, Brazil.

Escada, M. I. S., Vieira, I. C. G., Amaral, S., Araújo, R., Veiga, J. B. D., Aguiar, A. P. D., Veiga, I., Oliveira, M., Pereira, J. L. G., Filho, A. C., Fearnside, P. M., Venturieri, A., Carriello, F., Thales, M., Carneiro, T. S., Monteiro, A. M. V., & Câmara, G. (2005). Padrões e processos de ocupação nas novas fronteiras da Amazônia: O interflúvio do Xingu/Iriri [Land use patterns and processes in Amazonian new frontiers: The Xingu/Iriri region]. *Estudos Avançados* [Advanced Studies]*, 19*, 9-23.

Han, J., Koperski, K. & Stefanovic, N. (1997). GeoMiner: A system prototype for spatial data mining. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 553-556).

Han, J. & Kamber, M. (2001). Data mining - Concepts and tecniques. In D. Cerra & H. Severson (Eds.) (pp.405-412). San Diego, CA: Morgan Kaufmann Publishers.

INPE, National Institute for Space Research (2005). *PRODES project - Monitoring the Brazilian Amazon forest using satellites.* National Institute for Space Research. Retrieved April 8, 2008, from http://www.obt.inpe.br/prodes

Lambin, E. (1999). *Land-use and land-cover change implementation strategy*. Retrieved April 8, 2008, from http://www.geo.ucl.ac.be/LUCC/lucc.html

Lambin, E. F., Geist, H. J. & Lepers, E. (2003). Dynamics of land use and land cover change in Tropical Regions. *Annual Review of Environment and Resources, 28*(1) 205-241.

MacDonald, J. (2002). The Earth observation business and the forces that impact it. In D. Couts (Ed.), *Earth observation business network 2002*. Vancouver, CA: MacDonald Dettwiler.

May, M. & Savinov, A. (2002). An integrated platform for spatial data mining and interactive visual analysis. In *Proceedings of the International Conference on Data Mining Methods and Databases for Engineering* (pp. 90-101).

McGarigal, K. & Marks, B. (1995). *FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure*. USDA Forestry Service Technical Report PNW-351, Washington, DC.

McGarigal, K. (2002). Landscape pattern metrics. In A.H. El-Shaarawi & W.W. Piegorsch (Eds.), *Encyclopedia of environmentrics* (pp. 1135-1142). Sussex, England: John Wiley & Sons.

Meinel, G. & Neubert, M. (2004). A comparison of segmentation programs for high resolution remote sensing data. *International Archives of Photogrammetry and Remote Sensing, 35*(1), 1097-1105.

Nagao, M. & Matsuyama, T. (1980*). A structural analysis of complex aerial photographs*. New York: Plenum Press.

Piatetsky-Shapiro, G., Djeraba, C., Getoor, L., Grossman, R., Feldman, R. & Zaki, M. (2006). What are the grand challenges for data mining? - KDD-2006 Panel Report. *SIGKDD Explorations, 8*(2), 70-77.

Quinlan, R. (1993). *Programs for machine learning.* San Francisco: Morgan Kaufmann.

Rui, Y., Huang, T. S. & Chang, S. F. (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation, 10*(1), 39-62.

Rushing, J., Ramachandran, R., Nair, U. J., Graves, S. J., Welch, R. & Lin, A. (2005). ADaM: A data mining toolkit for scientists and engineers. *Computers and Geosciences, 31*(5), 607-618.

Schröder, M., Rehrauer, H., Seidel, K. & Datcu, M. (2000). Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Transactions on Geoscience and Remote Sensing, 23*(1), 2288-2298.

Shimabukuro, Y. et al. (1998). Using shade fraction image segmentation to evaluate deforestation in Landsat thematic mapper images of the Amazon region. *International Journal of Remote Sensing 19*(3), 535-541.

Silva, M. P. S., Câmara, G., Souza, R. C. M., Valeriano, D. M. & Escada, M. I. S. (2005). Mining patterns of change in remote sensing image databases. J. Han & B. Wah (Eds.), In *Proceedings of the Fifth IEEE International Conference on Data Mining* (pp. 362-369).

Silva, M. P. S. (2006). Mineração de Padrões de Mudança em Imagens de Sensoriamento Remoto [Mining patterns of change in remote sensing images] (Unpublished doctoral thesis). São José dos Campos: National Institute for Space Research (INPE).

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A. & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(1), 1349-1380.

Turner, M. G. (1989). Landscape ecology: The effect of pattern on process. *Annual Review of Ecology and Systematics, 20*, 171-197.

Wang, L., Khan, L. & Breen, C. (2002). Object boundary detection for ontology-based image classification. In *Proceedings of the Third ACM International Workshop on Multimedia Data Mining* (pp. 51-61).

Witten, I. H. & Frank, H. (1999). *Data mining: Practical machine learning tools and techniques with Java implementations*. San Francisco: Morgan Kaufmann.