CLUSTERING ANALYSIS APPLIED TO NDVI/NOAA MULTITEMPORAL IMAGES TO IMPROVE THE MONITORING PROCESS OF SUGARCANE CROPS

L. A. S. Romani[1], *R. R. V. Gonçalves*[3], *B. F. Amaral*[2], *D. Y. T. Chino*[2], *J. Zullo Jr.*[3], *C. Traina Jr.*[2], *E. P. M. Sousa*[2], *A. J. M. Traina*[2]

[1]Embrapa Agriculture Informatics, Campinas, PB 6041- 13083-886, Brazil
[2]Computer Science Department, USP at São Carlos, Brazil
[3]Cepagri, University of Campinas, Campinas, Brazil

ABSTRACT

This paper discusses how to take advantage of clustering techniques to analyze and extract useful information from multitemporal images of low spatial resolution satellites to monitor the sugarcane expansion. Additionally, we introduce the SatImagExplorer system that was developed to automatically extract time series from a huge volume of remote sensing images as well as provide algorithms of clustering analysis and geospatial visualization. According to experiments accomplished with spectral images of sugarcane fields, this proposed approach can be satisfactorily used in crop monitoring.

Index Terms— Time series, NOAA/AVHRR, productivity, K-means, K-medoids

1. INTRODUCTION

In the last decade, the improvements in data collection methods and sensor technology, as well as the large number of new remote sensing satellite launchings, have fostered an increase in the capabilities of acquiring spatial data. Specifically, satellites of polar orbiting such as NOAA/AVHRR, which are scheduled to accomplish two daily passes focusing each target on Earth, have generated an important archive of multitemporal images. As a consequence, the volume of satellite images stored in institutions in the whole world exceeds our human analysis capability, and we only can expect to grasp the potential information it embodies counting on the support of automated data analysis tools. The potential of satellite multitemporal images to support research of agricultural monitoring has increased according to improvements in technological development, especially in analysis of large volume of data available for knowledge discovery.

NOAA/AVHRR is one of the satellites that have been widely used as source of spectral information about agricultural regions, in particular sugarcane fields. As sugarcane crops are cultivated on large fields, researchers have used satellite-based images from sensors of medium and low spatial resolution such as NOAA/AVHRR to identify areas for sugarcane expansion [1], enabling to assess its social and economic impacts [2], such as to predict its yield [3], among other applications. In fact, this agricultural commodity has a strategic importance to the Brazilian economy, with increasing demand due to the ethanol production (one of its derivative) used as renewable energy source to replace fossil fuels.

In this context, this paper proposes a methodology based on data clustering to analyze NDVI (Normalized Difference Vegetation Index) multitemporal images in order to monitor the expansion of sugarcane crops. The presented experiments show that our approach can identify regions of sugarcane fields and areas with spectral mixture or noise. The whole process, including extraction of NDVI profiles, clustering methods and geospatial visualization, was implemented in a new system named SatImagExplorer [4]. The results confirm that NOAA/AVHRR images of sugarcane fields can be satisfactorily used in crop monitoring.

This paper is organized as follows. After introducing the problem and motivation in the first Section, we detail the material and methods in Section 2. Results and experiment analysis are presented in the Section 3. Finally, we conclude the work in Section 4.

2. MATERIAL AND METHODS

The study area is located in São Paulo, an important state of northeastern Brazil ($54°$ 00' to $43°$ 30' W and $25°$ 30' to $19°$ 30' S) which is responsible for the major sugarcane production in the country. NDVI multitemporal images were obtained from NOAA/AVHRR satellites from 2001 to 2009.

Figure 1 presents a flowchart with the process we propose to assess multitemporal satellite images. The proposed methodology involves georeferencing raw images and generating NDVI images in a process executed in batch by the NAVPRO system [5]. The process corrects raw images to eliminate distortions and noise. We have used NOAA-16 and

Thanks to Embrapa, FAPESP, CNPq, CAPES, Santander and Microsoft Research for funding.

 978-1-4577-1203-6/11/$26.00

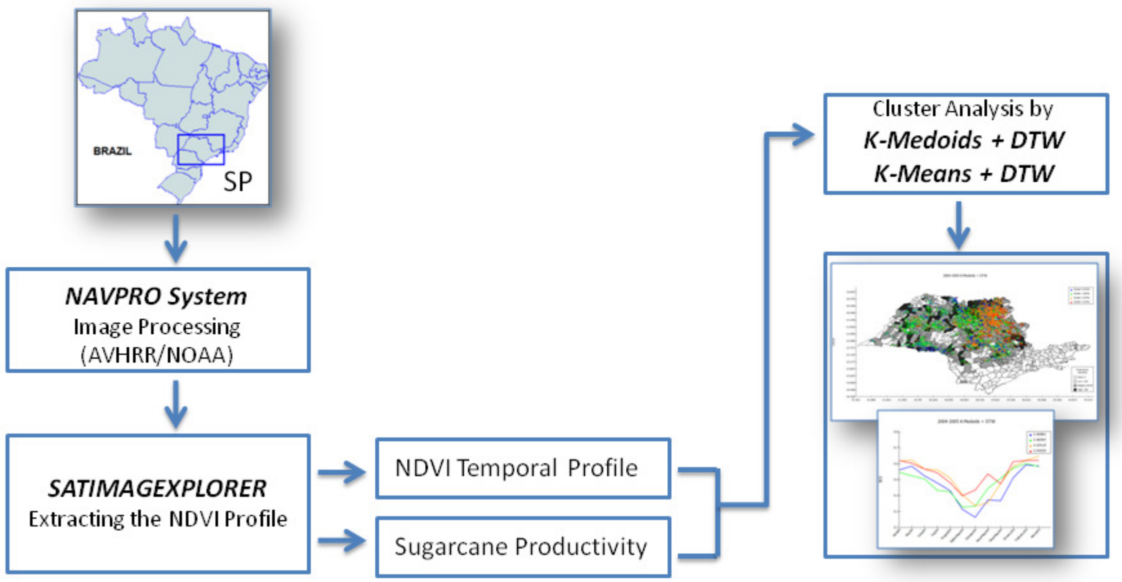

Fig. 1. Flowchart with the main steps of the methodology employed in this work.

NOAA-17 images to generate the NDVI multitemporal images from April, 2001 to March, 2009. In this work, we consider complete sugarcane cycles of approximately one year, each starting in April and ending in March. The effect of shadows, aerosols and water vapor are minimized by the generation of Maximum Value Composite (MVC) for NDVI. One MVC NDVI image was generated for each month of the eight year/crop season period.

The analysis process is performed using the SatImagExplorer system, which has been developed to handle different satellite multitemporal images, making it easier to extract time series from images. Figure 2 shows snapshots of SatImagExplorer, its interface is simple and intuitive allowing the user to extract time series from satellite images by point or region. If the user needs a time series that represents a specific region, the system generates one time series for each pixel in the defined area and the average value considering all pixels. All time series extracted from the images are stored in the database and presented in a graphic format. Besides the direct interaction with the system interface, users can also extract time series using a vector of coordinates that defines the desired region, as shown in Figure 2.

To guarantee that only the pixels classified as sugarcane fields would be processed, we generated masks to eliminate urban areas, soil and other kinds of vegetation. The mask for the 2004/2005 crop season was used to isolate pixels that are potentially sugarcane. Then, we extracted NDVI values associated to each pixel in all images, generating a time series for each pixel covered by the study area.

The NDVI time series are submitted to the K-Means and K-Medoids clustering methods associated to the Dynamic Time Warping (DTW) distance function [6]. Both K-Means and K-Medoids are partition-based clustering methods [7]. Results from the two clustering methods are presented in a map with municipal division. Additionally, SatImagExplorer shows a graph with the spectral signature of the centroid element of each cluster. In this work, these signatures represent the NDVI profiles for sugarcane crop in different regions of the study area.

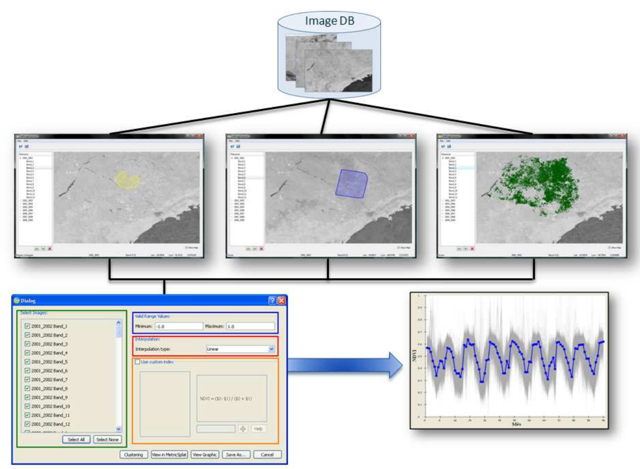

Fig. 2. Snapshot of SatImagExplorer system illustrating the extraction mode with regions of interest and mask generated by a coordinate file (on the left). On the bottom the time series graph generated.

Sugarcane productivity data were also used to generate a layer that can be compared to the cluster analysis results. Both cluster analysis and the plot of the results on a multi-layer map were developed by our group in the SatImagExplorer system.

3. RESULTS

In general, both clustering methods have presented satisfactory results regarding the identification of pixels correspondent to sugarcane fields, spectral mixture and noise. As the K-Means technique considers the mean of elements (series) to calculate the cluster centroids, it wasn't able to spot NDVI profiles following different trends (outliers). On the other hand, K-Medoid technique groups NDVI series that present curves distinct from the standard NDVI profile into different clusters. Thus, the results generated by the K-Medoids method were more effective to separate pixels that probably do not correspond to sugarcane.

Figure 3 shows two examples of the maps and graphs generated by the execution of K-Means and K-Medoids in the SatImagExplorer system for 2005-2006 and 2006-2007 crop seasons. The orange and red clusters correspond to NDVI profiles with higher mean values, what indicates larger sugarcane production. As it can be seen in the figure, the results of both clustering methods show that the majority of pixels in the Northeast region of São Paulo belongs to these clusters. This indeed corresponds to the reality, as these regions are know to be highly productive, as indicated in the map of sugarcane productivity (right column), where black indicates high productivity (>86 tons/ha), gray indicates medium productivity (66-85 tons/ha), light gray is low productivity (1-65 tons/ha) and white is equivalent to none productivity.

The blue cluster corresponds to the lowest value of NDVI time series. Signature of this group is less similar to the standard NDVI profile, as it can be seen in the graph at Figure 3. This cluster consists of pixels with spectral mixture, since NOAA/AVHRR has a low spatial resolution of 1km X 1km, noise caused by clouds and other problems in the satellite image. Negative peaks in the NDVI profile signature during December and January correspond to images eliminated in the georeferencing process due to the large number of clouds during the rainy season in this region of Brazil. Assessing the sequence of maps generated with the cluster analysis results (Figures 3(a) and (b)), we can identify the expansion of sugarcane production from the Northeast to West of the São Paulo state, observing the increasing of the clusters in red and orange from 2005-2006 to 2006-2007 crop seasons. This evidence can be confirmed by official data obtained at Brazilian government reports (IBGE - http://www.ibge.gov.br) that is illustrated in the sugarcane productivity maps (Figure 3).

4. CONCLUSIONS

The data mining approach to assess multitemporal satellite images presented satisfactory results considering two clustering methods. Regions of high sugarcane production were identified as well as the expansion of sugarcane to different areas of the state. Since the whole process of satellite images handling was developed in an automatic system (SatImagExplorer), the whole analysis process becomes more efficient in terms of time and ability to extract data from images, and also to visualize the results. As a consequence, specialists are able to better analyze a huge volume of images in a feasible manner, improving their work, accelerating final results and anticipating possibilities of new research work.

5. REFERENCES

[1] B. F. T. Rudorff, M. Adami, D. A. de Aguiar, A. Gusso, W. F. da Silva, and R. M. de Freitas, "Temporal series of evi/modis to identify land converted to sugarcane," in *IEEE International Geoscience and Remote Sensing Symposium - IGARSS 2009*, Cape Town, South Africa, 2009, vol. IV, pp. 252–255, IEEE.

[2] L. A. Martinelli and S. Filoso, "Expansion of sugarcane ethanol production in brazil: environmental and social challenges," *Ecological Applications*, vol. 18, no. 4, pp. 885–898, 2008.

[3] C. R. Nascimento, R. R. V. Gonçalves, J. Zullo Jr., and L. A. S. Romani, "Estimation of sugar cane productivity using a time series of avhrr/noaa-17 images and a phenology-spectral model," in *MultiTemp 2009 - The Fifth International Workshop on the Analysis of Multi-temporal Remote Sensing images*, Groton, Connecticut, 2009, pp. 365–372.

[4] D. Y. T. Chino, L. A. S. Romani, and A. J. M. Traina, "Construindo séries temporais de imagens de satélite para sumarização de dados climáticos e monitoramento de safras agrícolas," *Revista Eletrônica de Iniciação Científica*, vol. 10, pp. 1–16, 2010.

[5] J. C. D. M. Esquerdo, J. F. G. Antunes, D. G. Baldwin, W. J. Emery, and J. Zullo Jr, "An automatic system for avhrr land surface product generation," *International Journal of Remote Sensing*, vol. 27, pp. 3925–3942, 2006.

[6] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the Knowledge Discovery in Databases - KDD Workshop (KDD'1994)*, Seattle, Washington, USA, 1994, pp. 359–370, ACM Press.

[7] J. Han and M. Kamber, *Data Mining - Concepts and Techniques*, Morgan Kaufmann Publishers, New York, NY, USA, 2001.

978-1-4577-1203-6/11/$26.00

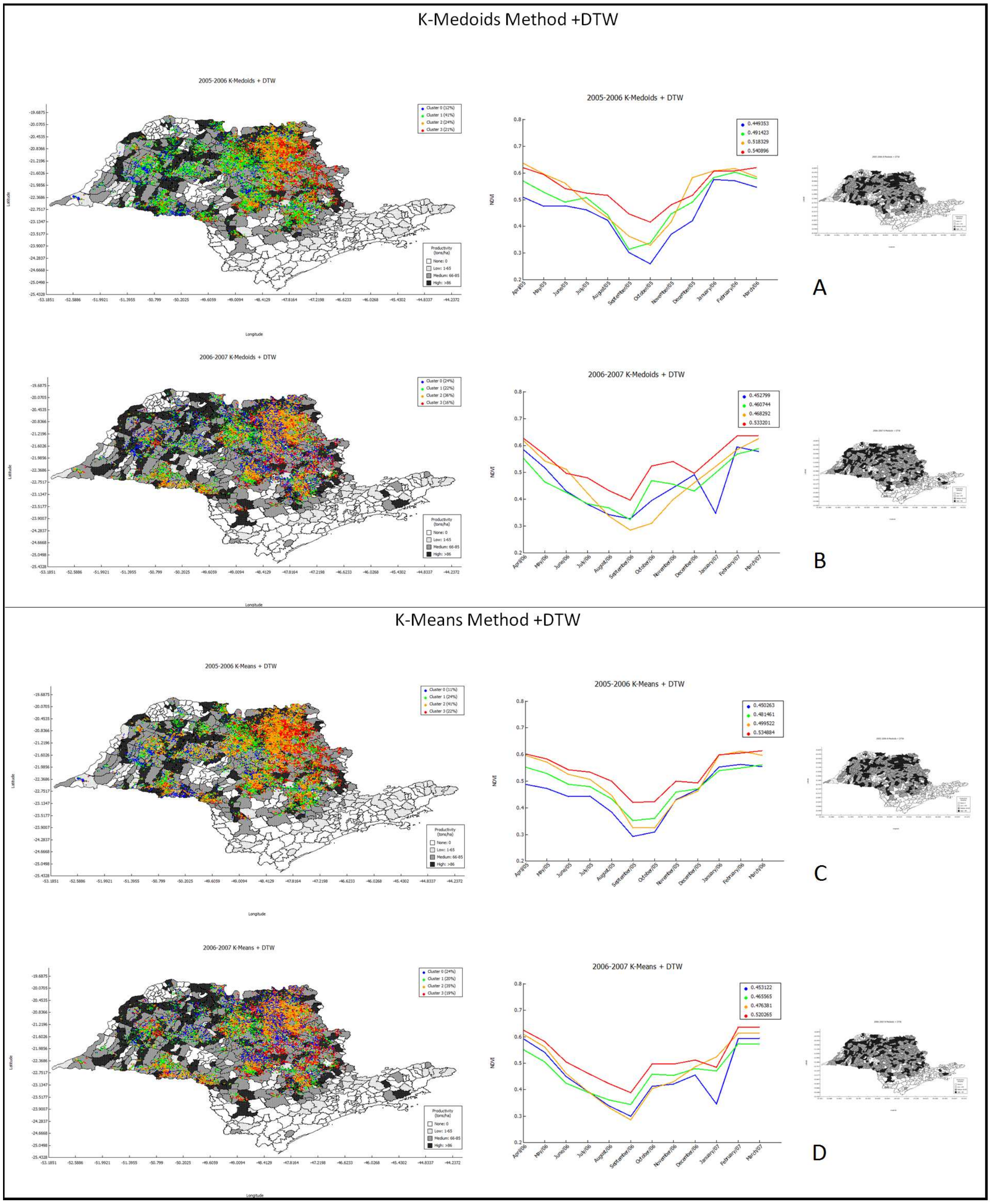

Fig. 3. Maps containing clusters of NDVI series and the corresponding plots showing the spectral signature for the centroids elements and maps with the sugarcane annual productivity.