2014-07

Land use temporal analysis through clustering techniques on satellite image time series

IEEE International Geoscience & Remote Sensing Symposium, 2014, Québec City.
http://www.producao.usp.br/handle/BDPI/48642

LAND USE TEMPORAL ANALYSIS THROUGH CLUSTERING TECHNIQUES ON SATELLITE IMAGE TIME SERIES

R. R. V. Gonçalves[1], J. Zullo Jr[1], B. F. Amaral[2], P. P. Coltri[1], E. P. M.Sousa[2], L. A. S. Romani[3]

[1]Center of Meteorological and Climate Researches Applied to Agriculture (Cepagri),
University of Campinas (Unicamp), Campinas, Brazil
[2]Institute of Mathematics and Computer Science,
University of São Paulo, São Carlos, Brazil
[3]Embrapa Agricultural Informatics, Campinas, Brazil

ABSTRACT

Satellite images time series have been used to study land surface, such as identification of forest, water, urban areas, as well as for meteorological applications. However, for knowledge discovery in large remote sensing databases can be use clustering techniques in multivariate time series. The clustering technique on three-dimensional time series of NDVI, albedo and surface temperature from AVHRR/NOAA satellite images was used, in this study, to map the variability of land use. This approach was suitable to accomplish the temporal analysis of land use. Additionally, this technique can be used to identify and analyze dynamics of land use and cover being useful to support researches in agriculture, even considering low spatial resolution satellite images. The possibility of extracting time series from satellite images, analyzing them through data mining techniques, such as clustering, and visualizing results in geospatial way is an important advance and support to agricultural monitoring tasks.

Index Terms - K-means, multivariate, vegetation index, albedo, surface temperature

1. INTRODUCTION

Numerous changes in Earth system have already significantly modified natural systems. Changes of atmospheric environment, in particular, have drawn the attention of decision makers and scientists, being at the center of global change discussions [1]. In agriculture, particularly, several studies have shown the effects and impacts of possible climate changes and global warming on Brazilian agricultural crops production. Considering the huge volume of data, researches in climatology, agriculture and environment increasingly need aid of computational techniques, since remote sensors data, weather stations and climate models generate databases in the order of terabytes, with prospect of exponential growth.

For several decades, short and long-term changes have been monitored using satellite data. Geographical information systems (GIS) and remote sensing tools have aided in studying and understanding better each aspect of our planet, including the relationship between land use and climate change. It is possible, thus, to use geo-technologies to monitor the land use [2]. Data from satellite sensors have proved to be suitable for identification of agricultural crops, urban and natural vegetation areas, as well as water resources. Sensors of land surface imaging produce multi-temporal image time series, such as MODIS/TERRA (from 2000), SPOT VEGETATION (from 1998) and AVHRR/NOAA (from 1970). These satellites allow generation of vegetation indices images [3] that enable the identification of targets, since they indicate vegetation biomass, albedo [4] and surface temperature [5]. Furthermore, the daily acquisition allows the usage of composite images, which largely removes undesirable effects of cloud cover. In addition, these techniques associated to computational tools can be used to automatically process, analyze and visualize remote sensing data that are very important to dynamically support agricultural, urban and environmental monitoring [6, 7]. One technique that can be used is clustering [8]. However, when we work with complex data as satellite images, one variable (one dimension) is not always enough to generate a satisfactory result from clustering techniques. Sometimes multidimensional dataset can improve the quality of clustering results allowing the algorithm to generate clusters with objects inter-clusters more distinct and objects intra-clusters more similar.

Based on this context, the main goal of this work was to use clustering techniques on three-dimensional (multivariate) time series of NDVI, albedo and surface temperature images from AVHRR/NOAA satellite to map the variability of land use in the state of São Paulo (Brazil).

The clustering algorithm used was K-Means, combined with Dynamic Time Warping (DTW) distance function.

2. MATERIAL AND METHODS

The study area was the state of São Paulo (Figure 1), located in Southeastern Brazilian Macro-Region (54°00’ to 43°30’W and 25°30’ to 19°30’S). Figure 2 shows a flowchart of the methodology used involving: (1) processing of AVHRR/NOAA images, (2) extraction of NDVI, albedo and surface temperature time series, and (3) clustering.

Fig. 1. Study area: state of São Paulo, Brazil.

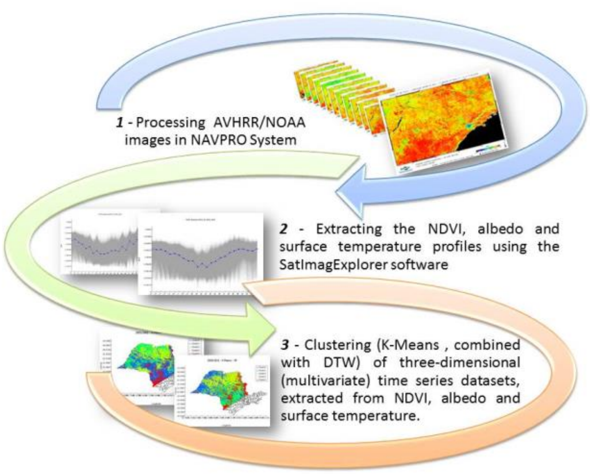


Fig. 2. Flowchart with main steps of methodology used in this work.

It was used a time series with 324 monthly images of NDVI, albedo and surface temperature, based on AVHRR/NOAA-16 and 17 sensors that have low spatial resolution (1km x 1km by pixel). Time series comprehend the period from April 2001 to March 2010, and were obtained in a database of AVHRR/NOAA images available at Center of Meteorological and Climate Researches Applied to Agriculture (Cepagri), of the University of Campinas (Unicamp). AVHRR/NOAA images were processed automatically performing radiometric calibration, geometric correction (accurate geo-referencing) and generation of products (NDVI, albedo and surface temperature).

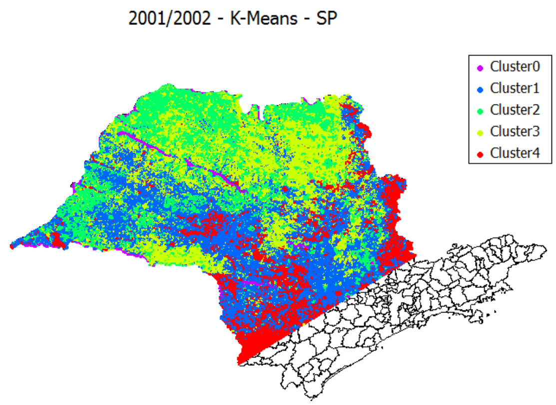
Whole dataset had 220,238 data series, being each observation a triplet of NDVI, albedo and surface temperature values of study area in a given month, with 108 values per time series. Clustering algorithm used was K-Means, combined with DTW distance function, implemented and incorporated in SatImagExplorer [9] image processing system. Clustering is a process of grouping sets of objects based on their similarity, so that one object is more similar to another in the same cluster, and less similar to another in a different cluster, according to a given distance function [10]. To perform time series clustering, use of a distance function compatible with time series were required. DTW distance function [11] was used to calculate similarity between two time series by performing the alignment among different pairs of data points. Because of this approach, if two time series have similar shapes but are not aligned in time axis, DTW can still recognize its similarity. Since DTW calculates the distance between pairs of data points using Euclidean distance, DTW method can be applied to multivariate time series.

3. RESULTS

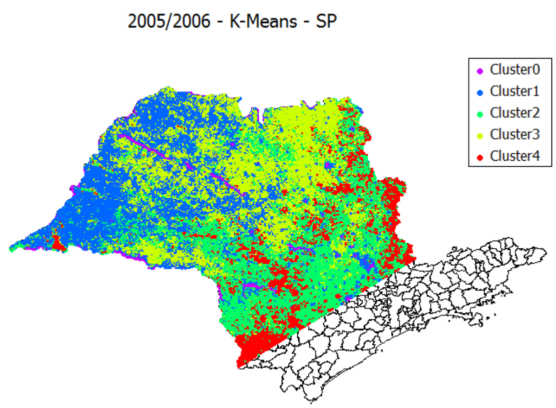
Dataset with more than 220,000 series, extracted from satellite images in the period 2001-2010, were clustered into five clusters (#0 - 4) by K-Means method with DTW distance function. The identified areas were sugarcane and other targets, such as water, urban area, agriculture crops, grasslands and forests in the state of São Paulo. Each cluster was formed according to the characteristics of NDVI, surface temperature and albedo from AVHRR/NOAA.

The resulting clusters represent the same targets at the time series of NDVI, surface temperature and albedo, and the temporal profiles showed homogeneity for each target. The cluster 0 (magenta) corresponds to water; cluster 1 (blue) to the urban area, and areas where the soil is exposed or have low vegetation and pasture; cluster 2 (green) represents areas of agriculture crops; cluster 3 (yellow) corresponds to sugarcane; and cluster 4 (red) represents forest areas (Figures 3A, B and C).

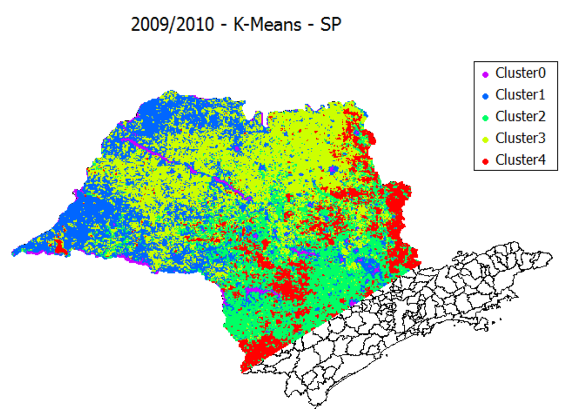
For example, as forests have high concentration of vegetation and biomass, these areas normally present high values of NDVI during the whole season, as shown by red colored representative time series, in profile visualization (Figure 4A). On the other hand, urban and water areas, represented by blue and purple profiles, usually present low NDVI values during the year due to lack of vegetation concentration. Clustering results for agriculture crops and grassland were less accurate, probably because different crops present similar NDVI values in some phenological phase during vegetative crop cycle, but are useful to separate agricultural from non-agricultural areas, such as water, urban areas and forest.



(A)

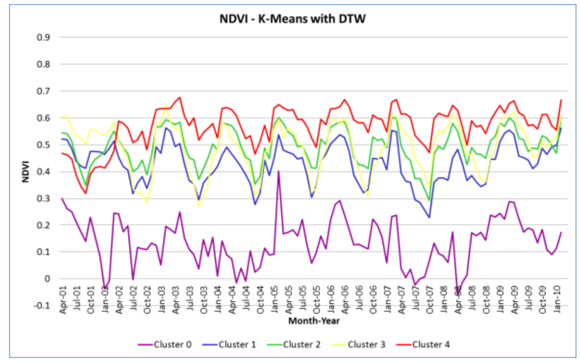


(B)

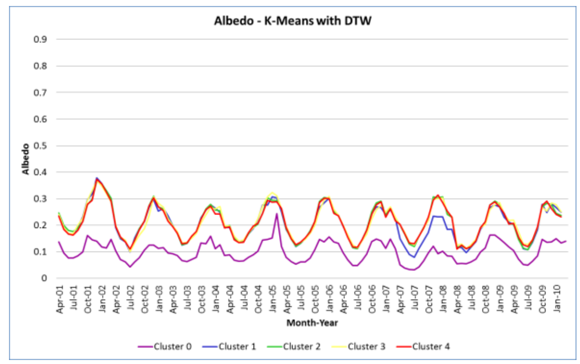


(C)

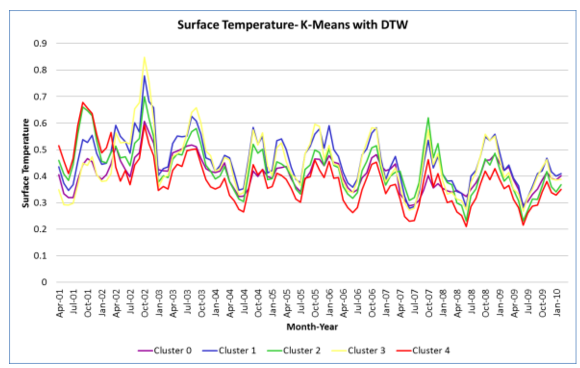
Fig. 3. Geographic spatial of 2001/2002 (A), 2005/2006 (B) and 2009/2010(C) of clustering results.



(A)



(B)



(C)

Fig. 4. Profile visualization (2001-2010) of NDVI (A), albedo (B) and surface temperature (C) of clustering results.

Albedo variable was useful to separate water areas from other targets, but was not enough to distinguish areas having different levels of vegetation cover (Figure 4B). Clustering of other areas was defined mainly by surface temperature, being higher for targets with lower canopy, such as urban areas and exposed soil, and lower for woodland.

In this context, the water represented by cluster 0, was well clustered, since the NDVI values and especially the albedo values are different from other clusters, as shown in the temporal profile of NDVI (Figure 4A) and albedo (Figure 4B). The albedo and NDVI values are lower (less than 0.1), since there is no presence of vegetation in the water, or when there is minimal.

Cluster arrangement varied from year to year because weather also varied during last decade (2001-2010), affecting mainly values of surface temperature. For example, when NDVI is higher, surface temperature is lower (Figures 4A and C). The forest areas represented by cluster 4, in Figures 3A, B and C, have high NDVI values (Figure 4A) and temperature values lower surface (Figure 4C), as they are very shady and dense vegetation coverage areas.

In general, the clustering defined by the K-Means distinguish targets with less accurate for agriculture and pasture, probably because different crops have similar NDVI values at any phenological stage during its growth cycle, but the NDVI values are useful for separating agricultural land and non-agricultural areas, such as water, forest and urban areas.

However, considering the most important agricultural crop of the state of São Paulo, having 60% of national production, sugarcane fields were well clustered over the crop years. This fact is due because the sugarcane has a typical behavior (long seasonal cycle) than other crops. In Figures 3A, B and C, it is possible to observe the dynamic of this agricultural crop, represented by cluster #3 (yellow), throughout the decade in which the crop year 2001/2002 the acreage was low, with higher production and planted in the northeast area of the state, and in the end of the crop years 2008/2009 and 2009/2010 there was a significant increase in the planted area towards the western of the state.

4. CONCLUSIONS

The clustering technique on three-dimensional time series applied in this work was suitable to accomplish the temporal analysis of land use, demonstrating that this methodology can be used to identify and analyze dynamics of land use and cover. The possibility of extracting time series from satellite images, analyzing them through data mining techniques, such as clustering, and visualizing results in geospatial way is an important advance and support to agricultural monitoring tasks. Moreover, the potential of this analysis can be highlighted by the use of low spatial resolution images, even this kind of image taking less detail as compared to the high resolution one. Finally, the main contribution of this research is the association of low spatial resolution satellite images with clustering techniques to extract useful information in a regional scale improving the work of farmers and decision makers.

5. REFERENCES

[1] J. A. Marengo, S. C. Chou, G. Kay, L. Alves, J. F. Pesquero, W.R. Soares, D.C. Santos, A. A. Lyra, G. Sueiro, R. Betts, D. J. Chagas, J. L.Gomes, J. F. Bustamante, and P. Tavares, "Development of regional future climate change scenarios in South America using the ETA CPTEC/HadCM3 climate change projections: Climatology and regional analyses for the Amazon, São Francisco and and the Parana River Basins", Climate Dynamics. v. 38, n 9-10, pp. 1829-1848, 2012.

[2] J. Sundaresan, K. M. Santosh, A. Déri, R. Roggema, and R. Singh, Geospatial Technologies and Climate Change, Springer, Switzerland, 2014.

[3] Q. Wang, S. Adiku, J. Tenhunen, and A. Granier, "On the relationship of ndvi with leaf area index in a deciduous forest site", Remote Sensing of Environment, v.94, n.2, pp. 244–255, 2005.

[4] A. P. Trishchenko, "Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors: Extension to AVHRR NOAA-17, 18 and METOP-A", Remote Sensing of Environment, v.113, pp. 335–341, 2009.

[5] I. Sandholt, K. Rasmussen, and J. Andersen, "A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status", Remote Sensing of Environment, v.79, pp. 213-224, 2002.

[6] J. Li, and R. M. Narayanan, "Integrated spectral and spatial information mining in remote sensing imagery", IEEE Transactions on Geoscience and Remote Sensing, v. 42, n. 3, pp. 673-685, 2004.

[7] L. A. S. Romani, R. R. V. Gonçalves, B. F. Amaral, D. Y. T. Chino, J. Zullo Jr., C. Traina Jr., E. P. M. Sousa, and A. J. M. Traina, "Clustering analysis applied to ndvi/noaa multitemporal images to improve the monitoring process of sugarcane crops", In: Proceedings of the The 7th International Workshop on the Analysis of Multitemporal Remote Sensing images (Multitemp'2011), Trento, Italy: IEEE, pp. 33-36, 2011.

[8] O. Maimon, and L. Rokach, The Data Mining and Knowledge Discovery Handbook, Springer, Heidelberg, 2005.

[9] 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, v. 10, pp. 1-16, 2010.

[10] J. Han, and M. Kamber, Data Mining - Concepts and Techniques, 1st edition, New York, NY, USA: Morgan Kaufmann Publishers, 2001.

[11] D. Berndt, and J. Clifford, "Using dynamic time warping to find patterns in time series", In AAAI Workshop on Knowledge Discovery in Databases, Seattle – Washington, pp.359-370, 1994.