IDENTIFYING FEATURES FOR TEMPORAL ANALYSIS OF VEGETATION INDICES

Thales Sehn Korting, Leila Maria Garcia Fonseca, Gilberto Câmara

National Institute for Space Research – INPE
São José dos Campos, Brazil – (tkorting, leila, gilberto)@dpi.inpe.br

KEY WORDS: Vegetation Indices, Multitemporal Classification, Automatic Classification

ABSTRACT:

Multitemporal remote sensing imagery is a source of valuable information for land change analysis. Certain types of land cover can be monitored with low spatial resolution but high temporal resolution. From this data it is possible to create temporal profiles of the regions during an entire year. This scenario results in a database with huge volume of information. Therefore manual techniques are unfeasible for classifying this data. Automatic methods arise as an appealing alternative, and this work shows preliminary results of identifying land changes in a region of Mato Grosso state, Brazil. We propose the use of temporal profiles of vegetation indices and features extracted from the profiles, gathering the traditional features with linearity measures. Decision trees algorithm were used to define a classification model to distinguish agricultural sites from other land uses.

1. INTRODUCTION

Multitemporal remote sensing imagery is a source of valuable information for land change detection and classification. MODIS sensor, for example, provides a rich dataset of images with around 250m resolution every 8 days (Jiang et al. 2008). Certain types of land use or change can be monitored with this resolution, like agriculture or deforestation. Using vegetation index for a point, measured every 8 days, it is possible to recover a temporal profile of a region in an entire year, with nearly 45 measurements. Figure 1 shows an example of a temporal profile using the MODIS Enhanced Vegetation Index - EVI2 (Freitas et al. 2011).

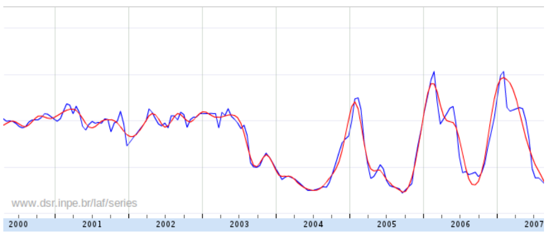

Figure 1. MODIS Enhanced Vegetation Index (EVI2) time series. Blue curve is the original and the Red one is a filtered curve (Freitas et al. 2011).

However, an entire EVI2 scene for a region must have several profiles, increasing the information to be analyzed. For example, the Brazilian state of Mato Grosso is captured in an MODIS scene with about 5000x5000 pixels of 250m spatial resolution. This results in around 1 billion values by year. As manual techniques for classifying land change are unfeasible with large databases, automatic methods arise as an appealing alternative. Since every pixel contains a temporal profile, its variation suggests different land cover classes, or even land changes (Aguiar et al. 2010).

Temporal profiles can vary according to the classes of land use detected in the ground. Figure 2 shows the profiles that characterize the transition from pasture to agriculture cover land. Therefore, classification methods can use these profiles to describe land use and land change, based on features extracted from these profiles (Korting et al. 2010).

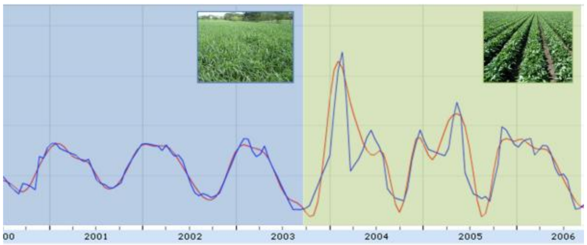

Figure 2. Land transition from pasture (years 2000 to 2003) to agriculture (2003 to 2006).

The automatic detection of land use/change using profiles has already been discussed by proposed in the literature (Hüttich et al. 2009; Boriah et al. 2008). Research relies on classifying basic features from the profiles such as amplitude, maximum and minimum values, area under the curve, and so on. In some applications these features may not precisely characterize the classes of interest. Other features, such as linearity measures, which are invariant to rotation, scaling, and translation (Stojmenovic et al., 2008), can improve the land use and cover classification.

Therefore, as a case study we propose to evaluate the distinction of annual crops from other land cover types (pasture, forest, bare soil, and sugarcane). In this work we propose to use features extracted from temporal profiles, including the linearity measures.

2. METHODOLOGY

The proposed method is composed by several steps towards certifying the most suitable features for profiles classification. To evaluate the features, we gathered a training set of EVI2 profiles and extracted several features, submitting to feature selection. The aim is to distinguish annual crops from other targets in the Brazilian state of Mato Grosso.

The features extracted from training samples were used to automatically create the classification model through data mining techniques (Quinlan, 1993). This scheme is depicted in Figure 3. The accuracy evaluation was assessed through the number of correctly classified instances and *Kappa* indices.

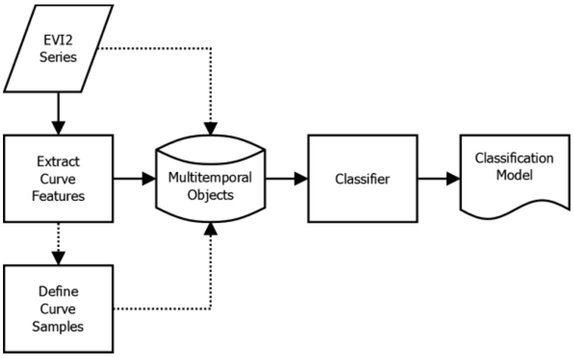

Figure 3. The outline for time series classification using EVI2.

The set of features are divided into two main groups. The traditional features, as shown in Figure 4, and the linearity features. Traditional measures include amplitude, maximum and minimum values, area under the curve, first and second slopes and their respective attributes, and so on.

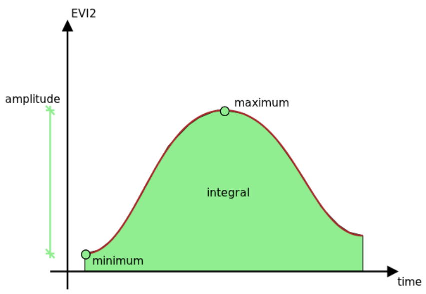

Figure 4. Traditional features from a temporal profile.

The second group comprises linearity measures, such as eccentricity, angle of orientation, and ellipse ratio. These features were adapted from a technique to identify lines in images. According to Stojmenovic et al., (2008), measuring the linearity of a finite set of points is an interesting way of identifying important components of a picture.

The features are derived from the central moment of order pq from a set of points Q. It is defined as

$$\mu_{pq} = \sum_{x,y \in Q} (x - x_c)^p (y - y_c)^q ,$$

where (x_c, y_c) is the center of mass from Q, i.e. the average position of the points in a time series. One axis is time, and the other is EVI2. The feature eccentricity is calculated using the central moments with different orders, as follows:

$$e = \frac{\sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{\mu_{20} + \mu_{02}} .$$

The angle of orientation of Q is determined by

$$a = 0.5 \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) .$$

The ellipse ratio is obtained by the following equation:

$$r = 1 - \frac{b}{a} ,$$

where

$$a^+, b^- = \sqrt{\frac{2\left[\mu_{20} + \mu_{02} \pm \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}\right]}{\mu_{00}}} .$$

The set of features plus the class information describe the multitemporal objects. These objects fed an algorithm to produce the classification model. We employed the decision tree algorithm, based on Quinlan (1993), version C4.5. As pointed by Wang and Li (2008), decision tree algorithms also select the best features, since the model will point to the most relevant features to detect our classes of interest. The classification model created finds thresholds over the feature set. According to these thresholds, new objects are classified in one class or another.

3. PRELIMINARY RESULTS

The case study presented in this work evaluates the distinction of annual crops from other land cover types (pasture, forest, bare soil, and sugarcane). Data consist in time series of EVI2 values from MODIS sensor. The series provides curves from years 2000 to 2008, with 8 days of temporal resolution, in the Brazilian state of Mato Grosso. Each curve was classified into cycles of one crop year, suiting to time series with 45 points. The entire database contains 5358 curves. Training samples contained 444 instances of annual crops pattern and 895 instances of the remaining patterns. After running the classification algorithm, we got the following decision tree:

```
eccentricity ts <= 0.25
|   max ts <= 0.67: Others
|   max ts > 0.67
|   |   std 1s <= 0.04: Others
|   |   std 1s > 0.04: Annual Crops
eccentricity ts > 0.25
|   std 1s <= 0.03
|   |   median ts <= 0.28: Annual Crops
|   |   median ts > 0.28: Others
|   std 1s > 0.03: Annual Crops
```

By analyzing the model, we infer that four features were enough to distinguish annual crops from the remaining patterns. They include eccentricity of the time series, maximum value of the time series, standard deviation of the time series' first slope, and the median of the time series. Model validation points out 94% of correctly classified instances, and *Kappa* value of 0.87.

We performed another test with the same training set, using only the traditional features. The accuracy of the results remained the same; however the decision tree became more complex, as the following:

```
std 1s <= 0.03
|   median ts <= 0.29
|   |   std1s <= 0.02: Others
|   |   std 1s > 0.02
|   |   |   max2s <= 0.02: Annual Crops
|   |   |   max2s > 0.02
|   |   |   |   median ts <= 0.25: Annual Crops
|   |   |   |   median ts > 0.25: Others
|   median ts > 0.29: Others
std 1s > 0.03
|   mode ts <= 0: Others
|   modets > 0
|   |   median ts <= 0.38: Annual Crops
|   |   median ts > 0.38
|   |   |   std 1s <= 0.05: Others
|   |   |   std 1s > 0.05: Annual Crops
```

4. CONCLUSION

In this work we presented a method to extract features from remote sensing time series. We applied our method to distinguish annual crops from other land cover types. Our method relied on data mining techniques to discover the most proper set of features to distinguish our classes of interest.

The set of features proposed in this work is based on linearity measures. Our preliminary results showed that these new features, coupled with traditional ones, distinguished correctly 94% of our dataset with a simpler set of thresholds than using only traditional features. Future works include the evaluation of all measures with different study cases.

REFERENCES

Aguiar, D., Adami, M., Silva, W., Rudorff, B., Mello, M., Santos, J. MODIS Time Series to assess Pasture Land. IEEE International Geoscience and Remote Sensing Symposium. 2123-2126, 2010.

Boriah, S., Kumar, V., Steinbach, M., Potter, C., Klooster, S. Land Cover Change Detection: A Case Study. 14th ACM SIGKDD international conference on knowledge discovery and data mining. 857-865, 2008.

Freitas, R., Arai, E., Adami, M., Souza, A., Sato, F., Shimabukuro, Y., Rosa, R., Anderson, L., Rudorff, B. Virtual laboratory of remote sensing time series: visualization of MODIS EVI2 data set over South America. Journal of Computer Interdisciplinary Sciences, Volume 2. 57-68, 2011.

Hüttich, C., Gessner, U., Herold, M. Strohbach, B. Schmidt, M., Keil M., Dech, S. On the Suitability of MODIS Time Series Metrics to Map Vegetation Types in Dry Savanna Ecosystems: A Case Study in the Kalahari of NE Namibia. Remote Sensing, Volume 1. 620-643, 2009.

Jiang, Z., Huete, A., Didan, K., Miura, T. Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment. Volume 112, 3833-3845, 2008.

Korting, T., Fonseca, L., Câmara. G. Decision Trees to Detect Changes in Remote Sensing Image Time Series. GeoChange, GeoInfo. 2010.

Quinlan, J. C4.5: Programs for Machine Learning. Morgan Kaufmann. 1993.

Stojmenovic, M., Nayak, A., Zunic, J. Measuring linearity of planar point sets. Pattern Recognition, Volume 41. 2503-2511, 2008.

Wang, Y., Li, J. Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data. International Journal of Remote Sensing, Volume 29. 2993-3010, 2008.