SHAPE-BASED TIME SERIES ANALYSIS FOR REMOTE PHENOLOGY STUDIES

Ricardo da S. Torres1, Makoto Hasegawa2, Salvatore Tabbone2, Jurandy Almeida1, Jefersson A. dos Santos1, Bruna Alberton3, and Leonor Patricia C. Morellato3

1 RECOD Lab, Institute of Computing, University of Campinas – UNICAMP
13083-852, Campinas, SP – Brazil
Email: {jurandy.almeida, jsantos, rtorres}@ic.unicamp.br

2 LORIA, UMR 7503, Université de Lorraine
54506, Vandoeuvre-lès-Nancy – France
Email: hasegawa.makoto@m.ieice.org, tabbone@loria.fr

3 Phenology Lab, Dept. of Botany, São Paulo State University – UNESP
13506-900, Rio Claro, SP – Brazil
Email: bru.alberton@gmail.com, pmorella@rc.unesp.br

1. INTRODUCTION

Plant phenology has gained importance as one of the most reliable indicators of species responses to global climate change, stimulating the development of new technologies for phenological observation. In this context, digital cameras have been successfully used as multi-channel imaging sensors, providing measures to estimate changes on phenological events, such as leaf flushing and senescence [1, 2, 3, 4, 5, 6, 7]. We have been monitoring leaf-changing patterns of a cerrado-savanna vegetation by taken daily digital images [8]. For that, we extracted leaf color information from the RGB channels and correlated the changes in pixel levels over time with leaf phenology patterns. The image analysis was conducted by defining six ROIs based on the random selection of six plant species identified in the digital image. Time series associated with different regions in the images have been obtained, raising the need of using appropriate tools for mining patterns of interest.

In this paper, we aim to identify appropriate shape descriptors for characterizing time series. In our study, time series will be seen as open contours that will be characterized by traditional and recently proposed shape description algorithms. Our first goals are: i) to determine which color channel is better for extracting shape descriptors and ii) to analyze the impact of the sunshine on the performance of shape descriptors. Our study opens a new area of investigation related to the use of shape descriptors to identify and characterize phenological changes.

2. TIME SERIES ACQUISITION

The near-remote phenological system was set up in a 18m tower in a Cerrado sensu stricto, a savanna-like vegetation located at Itirapina (22°10’49.18″S / 47°52’16.54″O), São Paulo State, Brazil. A digital hemispherical lens camera (Mobotix Q24) was setup at the top of the phenology tower, attached in an iron arm facing northeast. The camera activity is controlled by a timer and the energy source is a 12V battery charged by a solar panel.

The first data collection from the digital camera started on 18th August 2011. We set up the camera to automatically take a daily sequence of five JPEG images (at 1280 × 960 pixels of resolution) per hour, from 6:00 to 18:00 h (UTC-3). The present study was based on the analysis of over 2,700 images (Figure 1), recorded at the end of the dry season, between August 29th and October 3rd 2011, day of year 241 to 278 (DOY), during the main leaf flushing season [8].

Our strategy to evaluate shape descriptors in the context of time series description is based on assessing the similarity among regions associated with individuals of the same species. Regions are defined by using the hierarchical segmentation based on the Guigues algorithm [9]. The image used to obtain the hierarchy of segmented regions was taken at noon on October 15th, 2011. We have selected 5 segmentation scales from the hierarchy to perform feature extraction. The finest scale is composed by 27,380 regions and the coarsest scale contains 8,849 regions. Figure 1 illustrates the segmented scales in a subimage sample.

We analyze each region in terms of the contribution of the primary colors (Red, Green, and Blue), as proposed by Richardson et al. [1]. Initially, we analyze each color channel and compute the average value of the pixel intensity. After that, we compute the relative (or normal-
Fig. 1. Sample image of the cerrado savanna recorded by the digital camera on October 15th, 2011; and the segmentation results for the selected scales in a subimage sample.

The average (or total) brightness of each color channel, as:

\[ \text{Total}_{\text{avg}} = \text{Red}_{\text{avg}} + \text{Green}_{\text{avg}} + \text{Blue}_{\text{avg}}. \]  

\[ \% \text{ of Red} = \frac{\text{Red}_{\text{avg}}}{\text{Total}_{\text{avg}}}; \quad \% \text{ of Green} = \frac{\text{Green}_{\text{avg}}}{\text{Total}_{\text{avg}}}; \quad \% \text{ of Blue} = \frac{\text{Blue}_{\text{avg}}}{\text{Total}_{\text{avg}}}. \]

Figure 3(a) shows a time series related to the values extracted from the R channel for one of regions at scale 5, considering only the digital images taken at the midday.

We defined six ROIs (Figure 2) based on the random selection of six plant species identified in the hemispheric image: (1) *Aspidosperma tomentosum* (Figure 2(a)), (2) *Caryocar brasiliensis* (Figure 2(b)), (3) *Myrcia guianesis* (Figure 2(c)), (4) *Miconia rubiginosa* (Figure 2(d)), (5) *Pouteria ramiflora* (Figure 2(e)), and (6) *Pouteria torta* (Figure 2(f)).

Fig. 2. ROIs defined for the analysis of six plant species from the cerrado-savanna vegetation.

### 3. SHAPE-BASED TIME SERIES DESCRIPTION

Shape description and analysis play important roles in several applications, such as document analysis and content-based image retrieval. Proposed approaches to characterize the shape complexity of objects can be divided into contour or region methods. This classification takes into account whether shape features are extracted from the contour only or from the whole shape region. Descriptors belonging to both classes are considered in our study.

Each time series is seen as an open contour and is modeled under two different perspectives: as either an 1D or 2D signal. A time series is seen as an 1D signal, when used shape descriptors characterize the time series complexity by taking into account the change of normalized values over time, as illustrated in Figure 3(b). In the case of the 2D-signal modeling, we have transformed each time series into a binary object, as illustrated in Figure 3(c). Next, the shape complexity of each object is characterized by shape descriptors.

### 4. EXPERIMENTS AND RESULTS

Our study includes traditional and recently proposed shape descriptors. The descriptors used to characterize time series encoded in an 1D signal are: Centroid Distance (CD), Area with Regard to Centroid (ARC), Triangle Area Representation (TAR) [10], 1D Fourier [11], Derivative Curve, and Curvature. The descriptors used to describe time series encoded in a 2D signal are: Beam Angle Statistics (BAS) [12], Multiscale Fractal Dimension (MFD) [13], Fourier Descriptors (FD) [11], and Moment Invariants (MI) [14].

We have used a protocol based on the similarity among regions. The similarity between two regions is computed as a function of the distance between the feature vectors extracted from their time series. A shape descriptor is better than another if it ranks more regions belonging to the same ROI of an input region at the first positions. We consider a given region as belonging to a ROI if at least 80% of its
size (in pixels) is overlapped by such a ROI. In our experiments, we have used only regions from the finest scale, as they have been shown the most effective ones to characterize plant species [7].

A total of 398 regions associated with the six ROIs (c.f., Section 2) were randomly selected as queries. Presented results consider the average performance of evaluated shape descriptors. We assess the effectiveness of each approach using \( p@5 \) (precision at 5). This measure refers to the number of relevant regions (i.e., regions belonging to the same ROI of a query) ranked at the first 5 positions.

Figure 4 presents the \( p@5 \) effectiveness measure observed for each of the 1D shape descriptors along all the available periods of the day. Those graphs indicate that the performance of the different evaluated methods is similar, with a small advantage to the Fourier shape descriptor. Observe that early and late hours are better to characterize the phenological changes of plant species using 1D shape descriptors. This finding disagrees with the general suggestion of extracting color information from midday hours for ecological studies [1, 2, 3, 4]. Notice also the differences between the behavior of each of the evaluated descriptors with respect to the color channels. As we can observe, the best performances were achieved using the R color channel (Figure 4(a)).

The \( p@5 \) effectiveness measure observed for each of the 2D shape descriptors is presented in Figure 5. Those graphs show the behavior of those approaches along the day hours. Clearly, the FD shape descriptor outperforms all the other methods for the majority of the available periods of a day. Unlike the results obtained for the 1D shape descriptors, it is difficult to identify a pattern for the influence of the sunshine in the performance of the 2D shape descriptors. Nevertheless, it is important to realize that the worst results were achieved for late afternoon hours. On the other hand, different from 1D shape descriptors, the best performances were achieved using the B color channel (Figure 4(c)). Despite all those differences, the 1D shape descriptors perform better than the 2D ones, specially for extreme hours (morning and afternoon).

5. CONCLUSIONS

This paper has discussed the impact of applying shape descriptors to characterize time series related to phenological changes in the high diversity of the tropical cerrado savanna vegetation.

By considering time series as open contours that are characterized by shape description algorithms, we were able to define the best hours of the day for characterizing the leaf-changing patterns of plant species using shape descriptors. Different from the suggestion of using midday hours reported in ecological studies, the extreme hours (morning and afternoon) have shown the best results for the time series description using shape descriptors. Moreover, we have also shown the behavior of different shape descriptors with respect to the color channels.

Future work includes the evaluation of other shape descriptors. We also plan to consider learning-to-rank methods (e.g., genetic programming [15]) for combining different descriptors. Finally, we want to investigate the effects of using shape descriptors for characterizing time series from other application domains.
Fig. 5. \( p@5 \) effectiveness measure of evaluated 2D shape descriptors for different timestamps.

6. REFERENCES


