Ecological Informatics 26 (2015) 61–69

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Deriving vegetation indices for phenology analysis using genetic programming

Jurandy Almeida [a,d,*], Jefersson A. dos Santos [b], Waner O. Miranda [b], Bruna Alberton [c], Leonor Patricia C. Morellato [c], Ricardo da S. Torres [d]

[a] Institute of Science and Technology, Federal University of São Paulo — UNIFESP, 12247-014 São José dos Campos, SP, Brazil
[b] Department of Computer Science, Universidade Federal de Minas Gerais — UFMG, 31270-010 Belo Horizonte, MG, Brazil
[c] Phenology Lab, Dept. of Botany, São Paulo State University — UNESP, 13506-900 Rio Claro, SP, Brazil
[d] RECOD Lab, Institute of Computing, University of Campinas — UNICAMP, 13083-852 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 26 September 2014
Received in revised form 19 January 2015
Accepted 19 January 2015
Available online 30 January 2015

Keywords:
Remote phenology
Digital cameras
Image analysis
Vegetation indices
Genetic programming

ABSTRACT

Plant phenology studies recurrent plant life cycle events and is a key component for understanding the impact of climate change. To increase accuracy of observations, new technologies have been applied for phenological observation, and one of the most successful strategies relies on the use of digital cameras, which are used as multi-channel imaging sensors to estimate color changes that are related to phenological events. We monitor leaf-changing patterns of a cerrado-savanna vegetation by taking daily digital images. We extract individual plant color information and correlate with leaf phenological changes. For that, several vegetation indices associated with plant species are exploited for both pattern analysis and knowledge extraction. In this paper, we present a novel approach for deriving appropriate vegetation indices from vegetation digital images. The proposed method is based on learning phenological patterns from plant species through a genetic programming framework. A comparative analysis of different vegetation indices is conducted and discussed. Experimental results show that our approach presents higher accuracy on characterizing plant species phenology.

1. Introduction

Phenology is the study of recurrent natural phenomena and its relationship to climate (Schwartz, 2013). Traditional phenology studies rely on the direct observation of plants, which is a time consuming and error-prone task. Recently, digital cameras have been applied as tools to monitor leaf changes on plants automatically (Alberton et al., 2014; Morellato et al., 2013; Schwartz, 2013).

The digital cameras can increase the accuracy of phenological observation and widen the study area, but their successful application as multi-channel imaging sensors to capture vegetation changes is reliant to the extraction of color change information out of the images (Alberton et al., 2014; Sonnentag et al., 2012). Generally, leaf color information is extracted from the red, blue, and green (RGB) color channels and the green channel is the most utilized to describe leaf changes, in combination to the red color (Richardson et al., 2009, 2007). The normalized RGB chromatic coordinates are currently considered one of the most reliable indices to describe phenological changes based on image analyses (Sonnentag et al., 2012).

Considering the actual relevance of phenology as a tool for monitoring plant responses to climatic change and the need to reveal the environmental triggers of tropical phenology, here we propose a novel approach for deriving appropriate vegetation indices from vegetation digital images. The proposed method is based on learning phenological patterns from plant species through a genetic programming (GP) framework (da S. Torres et al., 2009). According to this framework, complex combinations of vegetation indices are modeled as individuals of a population. These individuals are then evolved by means of genetic operators (e.g., crossover, mutation, and reproduction) along generations. The objective is to obtain better performing complex vegetation indices that can be used to characterize the behavior of plant species or functional groups.

We performed a rigorous comparative analysis of different vegetation indices and discussed our proposed index accuracy in relation to the most utilized in the literature, the chromatic coordinates index, for near surface remote phenology studies (e.g., Ahrends et al. (2009); Alberton et al. (2014); Nagai et al. (2011); Richardson et al. (2009, 2007); Sonnentag et al. (2012)). Our experimental results show that vegetation indices may complement each other and improve indices' accuracy on characterizing plant species.

* Corresponding author at: Institute of Science and Technology, Federal University of São Paulo — UNIFESP, 12247-014 São José dos Campos, SP, Brazil. Tel.: +55 12 3309 9590; fax: +55 12 3921 8857.
E-mail addresses: jurandy.almeida@unifesp.br (J. Almeida), jefersson@dcc.ufmg.br (J.A. dos Santos), wanermiranda@gmail.com (W.O. Miranda), bru.alberton@gmail.com (B. Alberton), pmorella@rc.unesp.br (L.P.C. Morellato), rtorres@ic.unicamp.br (R. da S. Torres).

http://dx.doi.org/10.1016/j.ecoinf.2015.01.003
1574-9541/

The remainder of this paper is organized as follows. Section 2 describes the methodology adopted for acquiring time series. Section 3 briefly discusses vegetation indices utilized in the phenology analysis. Section 4 presents the GP framework and shows how to apply it to model complex vegetation indices. Section 5 describes the experimental protocol adopted for evaluating vegetation indices. Section 6 reports the results of our experiments and compares our proposed index with other ones. Finally, we offer our conclusions and directions for future work in Section 7.

2. Time series acquisition

A digital hemispherical lens camera (Mobotix Q24) was set up in an 18 m tower in a Cerrado *sensu stricto*, a savanna vegetation located at Itirapina, São Paulo State, Brazil. We set up the camera to 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 2700 images, recorded at the end of the dry season, between August 29th and October 3rd 2011, day of year 241 to 278, during the main leaf flushing season (Alberton et al., 2014; Almeida et al., 2014).

The image analysis was conducted by defining different regions of interest (ROI), as described in Richardson et al. (2009, 2007) and defined by Alberton et al. (2014) for our target species (Fig. 1). We analyzed 22 ROIs (Fig. 1) of six plant species randomly selected in the hemispheric image: (i) three regions associated with *Aspidosperma tomentosum* (red areas), (ii) four regions for *Caryocar brasiliensis* (green areas), (iii) two regions for *Myrcia guianesis* (blue areas), (iv) seven regions for *Miconia rubiginosa* (orange areas), (v) two regions for *Pouteria ramiflora* (magenta areas), and (vi) four regions for *Pouteria torta* (cyan areas).

According to the leaf exchange data from the on-the-ground field observations on leaf fall and leaf flush at our study site, those species were classified into three functional groups (Alberton et al., 2014; Morellato et al., 1989): (i) deciduous, *A. tomentosum* and *C. brasiliensis*; (ii) evergreen, *M. guianensis* and *M. rubiginosa*; and (iii) semideciduous, *P. ramiflora* and *P. torta*.

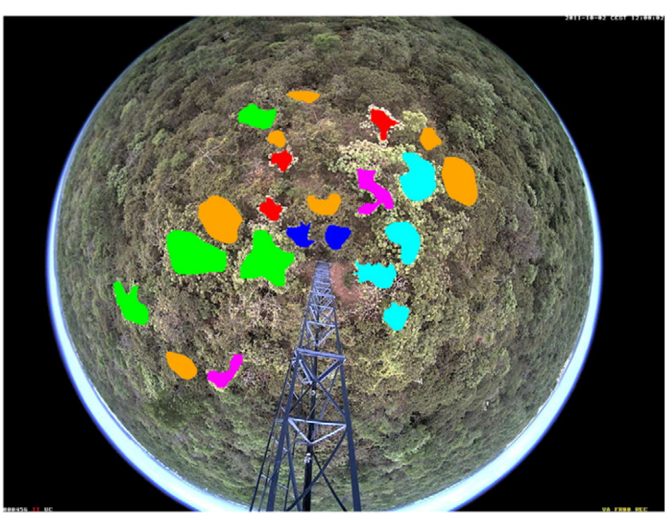

Fig. 1. Regions of interest (ROI) defined for the analysis of cerrado-savanna digital images, following Alberton et al. (2014). Each color represents a species: red = *Aspidosperma tomentosum*; green = *Caryocar brasiliensis*; blue = *Myrcia guianensis*; orange = *Miconia rubiginosa*; magenta = *Pouteria ramiflora*; cyan = *Pouteria torta*.

3. Vegetation index-based phenology analysis

Digital images allow the detection of phenological events according to the changes of red, green, and blue (RGB) color channels along time (Richardson et al., 2007). By quantifying the RGB color channels it is possible to estimate, for instance, leaf flushing and senescence, using the green and red channels, respectively (Ahrends et al., 2009; Henneken et al., 2013; Morisette et al., 2009; Richardson et al., 2009). The quantification of RGB is performed by applying indices of color channels to detect the leaf color changes in temporal time series of digital images (Nagai et al., 2011; Richardson et al., 2007; Sonnentag et al., 2012; Zhao et al., 2012; Zhou et al., 2013).

Non-normalized RGB coordinates correspond to the individual red (R), green (G), and blue (B) components of a digital image of 24 bit color (8 bits for each color). These values can be highly sensitive to the intensity and variation of the illuminating source and its angle with the background (Gonzalez andWoods, 2007). Non-normalized RGB coordinates tend to present considerable variation between images, specially related to light variation in shaded and not shaded surfaces (Woebbecke et al., 1995). Zhou et al. (2013) tested non-normalized values for the RGB color channels (RGBDN) and the indices fail to capture any seasonal change in the development of the canopy of winter wheat cultivated area.

The contrast indices can detect optical contrast between plant and non-plant surfaces (background) in an image (Woebbecke et al., 1995). The Excess green ($2G–R–B$) is a commonly applied contrast index to highlight the green information and complement phenological interpretation in several studies (e.g., Kurc and Benton (2010); Migliavacca et al. (2011); Nagai et al. (2011); Sonnentag et al. (2012)). The G/R, calculated on the basis of the difference of absorptive/reflective bands, corresponding to the vegetation canopy and soil surface, provides effective vegetative information for leaf color changes in (Zhou et al., 2013). Other examples of contrast indices are $R–G$, $G–B$, and $(G–B)/(R–G)$ (Woebbecke et al., 1995).

The perceptive attributes of color are classified as hue, saturation, and intensity (HSI model) (Woebbecke et al., 1995). Hue (or spectral shape) is measured as the wavelength of maximal reflection ($\lambda(R_{max})$). Hue modeling describes “redness”, “greenness”, and “blueness” of an object and it is calculated by the non-normalized RGB color channels (Woebbecke et al., 1995). Methods calculating Hue index were useful to detect leaf-color and leaf-fall patterns and their timings among species in deciduous broad-leaved trees (Nagai et al., 2011).

The normalized index called RGB chromatic coordinates (RGBcc) was developed by Gillespie et al. (1987) and it is considered the most efficient index to detect the color of plants in relation to their background (Woebbecke et al., 1995). Shaded and unshaded surfaces presented less variation in relation to other non-normalized indices (Woebbecke et al., 1995). Normalized chromatic coordinate is the most efficient index to distinguish leaves between monocots and dicots and to suppress light environment variation (see Alberton et al. (2014); Sonnentag et al. (2012); Woebbecke et al. (1995)). The normalized RGBcc is calculated according to Gillespie et al. (1987) and Woebbecke et al. (1995), as follow:

$$r_{cc} = \frac{R}{(R+G+B)} \tag{1}$$

$$g_{cc} = \frac{G}{(R+G+B)} \tag{2}$$

$$b_{cc} = \frac{B}{(R+G+B)}. \tag{3}$$