ARTICLE IN PRESS
JID: PATREC [m5G;March 31, 2016;17:12]
Pattern Recognition Letters 000 (2016) 1–10
Contents lists available at ScienceDirect
Pattern Recognition Letters
journal homepage: www.elsevier.com/locate/patrec

Fusion of time series representations for plant recognition in phenology studies[☆]

Fabio A. Faria[a,b,*], Jurandy Almeida[a,b], Bruna Alberton[c], Leonor Patricia C. Morellato[c], Ricardo da S. Torres[b]

[a] Institute of Science and Technology, Federal University of São Paulo – UNIFESP, São José dos Campos 12247-014, SP, Brazil
[b] Institute of Computing, University of Campinas – UNICAMP, Campinas 13083-852, SP, Brazil
[c] Dept. of Botany, Sao Paulo State University – UNESP Rio Claro, 13506-900, SP, Brazil

A R T I C L E I N F O

Article history:
Available online xxx

Keywords:
Plant species identification
Classifier fusion
Diversity measures

A B S T R A C T

Nowadays, global warming and its resulting environmental changes is a hot topic in different biology research area. Phenology is one effective way of tracking such environmental changes through the study of plant's periodic events and their relationship to climate. One promising research direction in this area relies on the use of vegetation images to track phenology changes over time. In this scenario, the creation of effective image-based plant identification systems is of paramount importance. In this paper, we propose the use of a new representation of time series to improve plants recognition rates. This representation, called recurrence plot (RP), is a technique for nonlinear data analysis, which represents repeated events on time series into a two-dimensional representation (an image). Therefore, image descriptors can be used to characterize visual properties from this RP images so that these features can be used as input of a classifier. To the best of our knowledge, this is the first work that uses recurrence plot for plant recognition task. Performed experiments show that RP can be a good solution to describe time series. In addition, in a comparison with visual rhythms (VR), another technique used for time series representation, RP shows a better performance to describe texture properties than VR. On the other hand, a correlation analysis and the adoption of a well successful classifier fusion framework show that both representations provide complementary information that is useful for improving classification accuracies.

1. Introduction

Global warming and its resulting environmental changes have raised important research topics of different disciplines. Among those is phenology that studies recurrent life cycles events and its relationship to climate [35]. To increase the range of study sites and species and the effectiveness of phenological observations, technological devices (e.g., multi-channel imaging sensors) have been successfully applied to provide metrics for estimating changes on phenological events, such as leaf development and senescence [3,4,33].

Plant species recognition in the digital images is not a trivial task, especially in tropical vegetations, where one single image may include a huge number of species [4]. This task is very time consuming since it has to be done in the field, first by matching each crown in the image to the tree in the soil and then by identifying the tree at species level.

Our goal in this work is to support automatic plant species recognition tasks based on phenological pattern information. Different patterns correspond to different species, as well as similar patterns can be grouped in one species type or in a leaf functional group that encompasses several species. This may not just save time for phenologists but also complement the phenological interpretation of the data collected.

Almeida et al. [7] have proposed the use of machine-learning methods to find similar patterns in the digital images and then check if those patterns correspond to similar species or functional groups. Their work was focused on the intraspecies analysis, i.e., on detecting different individuals of a same species. However, different species from a same functional group may exhibit a similar phenological pattern [4], confounding the classifiers. Hence, many questions arise when considering interspecies interactions, i.e., the recognition of individuals from different species but belonging to the same functional group [4].

☆ This paper has been recommended for acceptance by Jenny Qian Du.

* 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 9500; fax: +55 12 3921 8857.
E-mail addresses: ffaria@unifesp.br, juruna18@gmail.com (F.A. Faria), jurandy.almeida@unifesp.br (J. Almeida), 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.patrec.2016.03.005

Please cite this article as: F.A. Faria et al., Fusion of time series representations for plant recognition in phenology studies, Pattern Recognition Letters (2016), http://dx.doi.org/10.1016/j.patrec.2016.03.005

In this work, we adopt the strategy of characterizing phenological patterns from time series and distinguishing species from the same leaf functional group in plant species recognition tasks. In fact, several time series representations have been proposed in the literature. Some successful approaches include data-adaptive (e.g., SAX [26] and APCA [23]) and non-data adaptive representations (e.g., wavelets [31]). A good survey upon this subject can be found in [40].

In this work, we present a novel approach for time series representation, which is based on a technique for nonlinear data analysis called recurrence plots (RP). Different from other time series representations, RP provides a visual mechanism for pattern identification, being suitable for combining with state-of-the-art computer vision description approaches. This work has also been motivated by the results of [24] and [37]. Both studies indicate that it is possible to perform classification tasks through the use of recurrence plots and texture feature extraction approaches. RP technique has been used successfully in different application domain, such as action recognition [24], identification of diabetes analysis of epilepsy [1], and detection of financial crisis [2].

In our experiments, we performed four rigorous comparative analysis to show the robustness of RP-based representations for plant recognition tasks. We begin with an effectiveness study evaluating the performance of RP-based classifier associated with time series of different hours of day. Then, we compared the proposed approach with another time series representation proposed by [6], called Visual Rhythm (VR). Next, we performed a correlation analysis to find out agreement/disagreement between all classifiers involved between RP and VR-based representations. Finally, we adopt a successful classifier fusion framework [16] to combine the most suitable classifiers. Experimental results show that the combination of RP- and VR-based representations yields better results that their use in isolation.

In summary, the main contributions of this work are: (i) a new representation of time series based on recurrence plots technique for plant recognition; (ii) effectiveness study of the recurrence plots approach in different hours of day; (iii) effectiveness comparative study between recurrence plots and visual rhythm approaches; (iv) correlation analysis between recurrence plots and visual rhythm approaches; (v) use of a classifier fusion framework to combine the most suitable classifiers using both approaches.

The remainder of this paper is organized as follows. Section 2 presents the phenological data acquisition process considered in our study. Section 3 presents the recurrence plot approach and how to use it for phenological time series representation. Section 4 describes the experimental protocol adopted to validate the proposed approach. Section 5 reports the results of our experiments and compares the proposed approach with another time series representation. Finally, we offer our conclusions and directions for future work in Section 6.

2. Phenological data acquisition

A digital hemispherical lens camera (Mobotix Q24) was set up in an 18 m-high tower in a Cerrado *sensu stricto*, a neotropical savanna vegetation located at Itirapina, São Paulo State, Brazil [4,32]. Fig. 1 shows all steps of the time series acquisition process used in our work.

Firstly, 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 (Fig. 1a), 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 [4].

Next, the image analysis has been conducted by defining different regions of interest (ROI), as described in [33] and defined by [4] for our target species. Then, we analyzed 22 ROIs (Fig. 1b) obtained from a random selection of six plant species identified manually by phenology experts in the hemispheric image [4]:

(i) Three regions associated with *Aspidosperma tomentosum* (green areas).
(ii) Four regions for *Caryocar brasiliensis* (blue areas).
(iii) Two regions for *Myrcia guianesis* (orange areas).
(iv) Six regions for *Miconia rubiginosa* (magenta areas).
(v) Two regions for *Pouteria ramiflora* (cyan areas), and
(vi) Four regions for *Pouteria torta* (red areas).

We analyzed each ROI in terms of the contribution of the primary colors (R, G, and B), as proposed by [34] and described in [4]. Initially, we analyze each color channel and compute the average value of the pixel intensity (Fig. 1c). After that, we compute the normalized brightness of each color channel (RGB Chromatic coordinates) (Fig. 1d). The normalization of those values reduces the influence of the incident light, decreasing the color variability due to changes on illumination conditions [4,11,41]. Finally, by computing those values along the whole period (August 28th to October 3rd, 2011), we obtained time series to use as input data for our proposed framework (Fig. 1e).

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 on three functional groups [4,27]: (i) deciduous, *A. tomentosum* and *C. brasiliensis*; (ii) evergreen, *M. guianensis* and *M. rubiginosa*; and (iii) semideciduous, *P. ramiflora* and *P. torta*.

3. Recurrence plots for plant species recognition

This section introduces our approach for phenological time series representation using recurrence plots. Section 3.1 describes how to compute recurrence plots (RP) from time series, while Section 3.2 presents how we use RP for representing phenology data.

3.1. Recurrence plots (RP)

Recurrence plots (RP), proposed by [15] in dynamical systems, is an advanced technique of nonlinear data analysis. RP technique has been used to visualize repeated events (the recurrence of states) of higher dimensional phase spaces through projection into the two or three dimensional sub-spaces. This technique is able to investigate recurrent behavior (periodicity) at time series (m-dimensional phase space) through a two-dimensional representation, such as a distance square matrix.

Recurrence Plot might be defined by:

$$R_{i,j} = \Theta(\epsilon_i - \parallel x_i - x_j \parallel), x_i \in \mathcal{R}^m, i, j = 1 \ldots N \tag{1}$$

where N is the number of considered states (dots at the time series) x_i, ϵ_i threshold distance, $\parallel \cdot \parallel$ a norm between the states (e.g., *Euclidean* norm), m is the embedding dimension, and $\Theta(\cdot)$ the Heaviside function. This discontinuous function has value 0 for negative argument and 1 for positive argument.

Eq. (1) provides an $N \times N$ image, which shows us whether there are recurrent states or not, along the target trajectory. This created image might be binary or grayscale, depending on the choice of a threshold or not. Fig. 2 shows a real time series from the dataset and two examples of recurrence plot considering those real time series, unthresholded and thresholded.

The choice of an generalizable threshold to perform matching between two RP is a non-trivial task, but it can be possible with few heuristics [36]. As in this work we aim to make use of color and texture information, we have adopted the unthresholded approach using the *Manhattan* norm and $m = 1$. This unthresholded approach is defined in Eq. (2). However, we can not rule out that