



Phenological visual rhythms: Compact representations for fine-grained plant species identification[☆]



Jurandy Almeida^{a,d,*}, Jefersson A. dos Santos^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, Federal University of Minas Gerais – UFMG, 31270-010 Belo Horizonte, MG, Brazil

^c Department of Botany, Institute of Biosciences, São Paulo State University – UNESP, 13506-900 Rio Claro, SP, Brazil

^d Institute of Computing, University of Campinas – UNICAMP, 13083-852 Campinas, SP, Brazil

ARTICLE INFO

Article history:

Available online 23 December 2015

Keywords:

Remote phenology
Plant identification
Image analysis
Time series
Visual rhythm

ABSTRACT

Plant phenology, the study of recurrent life cycles events and its relationship to climate, is a key discipline in climate change research. In this context, digital cameras have been effectively used to monitor leaf flushing and senescence on vegetations across the world. A primary condition for the phenological observation refers to the correct identification of plants by taking into account time series associated with their crowns in the digital images. In this paper, we present a novel approach for representing phenological patterns of plant species. The proposed method is based on encoding time series as a visual rhythm. Here, we focus on applications of our approach for plant species identification. In this scenario, visual rhythms are characterized by image description algorithms. A comparative analysis of different descriptors is conducted and discussed. Experimental results show that our approach presents high accuracy on identifying individual plant species from its specific visual rhythm. Additionally, our representation is compact, making it suitable for long-term data series.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Plant phenology, the study of recurrent life cycles events and its relationship to climate, is a key discipline in climate change research [34]. One key component of phenology research is the leaf exchange patterns from leaf budding to senescence, due to its relevance to comprehend ecosystem processes, such as growth, water and gas exchange, and nutrient cycling [24]. The dynamics of plant growing seasons define the spatial and temporal patterns of carbon balance and water exchange, and ultimately the productivity of terrestrial ecosystems [22,35].

Recently, digital cameras have been effectively applied as multi-channel imaging sensors to estimate color changes (RGB channels) that are related to leaf flushing and senescence phenology [1,2,30]. The technique allows to increase the range of study sites and

species and the accuracy of phenological observations, and a clear perception of the start and end of the growing season [2,16].

We have monitored leaf changing patterns of a tropical cerrado-savanna vegetation by taking daily digital images [2]. We extracted image color information from the RGB (red, green, and blue) channels and correlated the changes in pixel levels over time with leaf phenology patterns [2]. The analysis was conducted after we defined regions of interest (ROI) based on the random selection of plant species crowns identified in the digital image [29]. We obtained a time series associated with each ROI, raising the need of using appropriate tools for mining patterns of interest in a given digital image [7,11,32,33].

The plant species identification is a key issue for the phenological observation of tree crowns using phenocams, especially in tropical vegetations where one single image may include a high number of species [2,6,7]. This task is time consuming since first each crown in the image has to be matched to the tree in the soil and then the tree is identified at species level. In this sense, we have developed and deployed computational methods to find similar patterns in the digital images and then we checked if they correspond to similar species or leaf functional groups [5,6].

The major challenge of designing automatic tools for addressing the plant identification task is to deal with fine-grained

[☆] This paper has been recommended for acceptance by Concetto Spampinato.

* 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: jurandy.almeida@unifesp.br (J. Almeida), jefersson@dcc.ufmg.br (J.A. dos Santos), bru.alberton@gmail.com (B. Alberton), pmorella@rc.unesp.br (L.P.C. Morellato), rtorres@ic.unicamp.br (R. da S. Torres).

recognition, where the categories are visually similar. In general, different plant species may have an analogous behavior with respect to leaf color change and, hence, the differences between their time series are quite subtle and hard to be detected, even for humans without careful training. Usually, plant identification is based on the analysis of huge volumes of sequential vegetation images, i.e., vegetation images obtained over time. In this scenario, another challenge refers to efficiency aspects, both in terms of storage requirements and processing time. Typical existing solutions for plant identification based on vegetation images [6,28] do not scale properly for handling ever-growing collections.

In this paper, we present an effective and efficient approach for capturing phenological patterns from time series generated by digital images. Our strategy consists of encoding time series as a visual rhythm [25]. This simple, yet effective, approach offers rich information regarding spatio-temporal data, which is useful in many fields of applicability. Here, we focus on applications of the proposed method to identify and distinguish the behavior of plant species. In this scenario, visual rhythms are characterized by traditional and recently proposed image description algorithms. Such methods are able to codify key image features into fixed-size representations.

The proposed method was evaluated in a dataset recorded during the main leaf flushing season composed of about 2,700 images [2]. We performed a detailed experimental comparison of several image descriptors. The results show that our approach presents high accuracy on identifying regions in the images belonging to a same plant species. In addition, our strategy provides a compact representation for time series. The improvement of the computation makes it suitable for long-term data sets.

This paper extends substantially our preliminary works presented in [3,4]. Here, we introduce several innovations. First, we present a review of the state-of-the-art approaches for processing spatio-temporal data. In addition, we discuss new strategies of generating visual rhythms from time series. Finally, after a much more thorough presentation of the proposed method, we extend the experimental evaluation of our technique, including a statistical analysis of its performance.

The remainder of this paper is organized as follows. Section 2 briefly describes related work. Section 3 discusses the methodology adopted for acquiring time series. Section 4 presents our approach and shows how to apply it to characterize time series. Section 5 presents the adopted experimental protocol, while Section 6 reports the results of our experiments and compares our technique with other methods. Finally, we offer our conclusions and directions for future work in Section 7.

2. Related work

The increasing accessibility to data with high spatio-temporal resolution has enabled a detailed analysis of vegetation properties. At the same time, it requires feature extraction techniques able to represent such properties, taking into account storage aspects.

Time series-based vegetation indices from remote sensing images (RSIs) are widely used for phenological and land cover change studies [10,13,17,31]. Rodrigues et al. [31] presented a software to extract phenological parameters (e.g., maturity and senescence) from Normalized Difference Vegetation Index (NDVI) time series. Foster et al. [13] also applied NDVI time series to detect grassland vegetation. Brooks et al. [10], in turn, proposed a Fourier-based algorithm to fit NDVI multitemporal curves and reduce missing data effects in the analysis. Hmimina et al. [17] exploited NDVI time-series to evaluate the potential use of MODerate resolution Imaging Spectroradiometer (MODIS) remote sensing data for monitoring phenological patterns in a African savanna.



Fig. 1. Sample image of the cerrado savanna recorded by the digital camera on October 5th, 2011.

In [6,8,23,28], the authors consider not only temporal but also spatial properties. For that, they extract time series from segmented regions. Petitjean et al. [28] proposed a strategy to encode spatial data over time. Their strategy consists of segmenting each image of the series in order to characterize each pixel of the data with spatial properties. The time series are computed for each pixel based on the properties extracted from the segmented regions. Ardila et al. [8], in turn, used time series based on spatial properties, from pre-defined regions to monitor urban trees. Almeida et al. [6] exploited a multiscale segmentation structure to compute time series with spatial information, which were used to detect phenological patterns in a cerrado-savanna vegetation. Ma et al. [23] analyzed spatial and temporal patterns in savanna vegetation phenology in Australia by comparing image datasets from different spatial resolutions.

In spite of all the advances, existing strategies for processing spatio-temporal data usually require a considerable amount of storage space. A traditional phenology database storing information from just one event per individual per year, for several species and observation stations may encompass a enormous amount of data. For instance, the data set of phenological observations of plant species from Central Europe, largely Germany, from about 9000 stations, covering 130 years (1880–2009), includes more than 16 millions observations [12]. This paper aims to fill such a gap. Here, we introduce a compact representation for identifying and characterizing plant species in time series obtained from phenological observations.

3. Time series acquisition

The near-remote phenological system was set up in an 18 m tower in a Cerrado *sensu stricto*, a savanna-like vegetation located at Itirapina, São Paulo State, Brazil. We set up the camera to automatically take a daily sequence of five JPEG images (at 1280×960 pixels of resolution) in the first 10 min of each hour, from 6:00 to 18:00 h (UTC3), totaling 65 images per day. The present study was based on the analysis of over 2,700 images (Fig. 1), recorded at the end of the dry season, between August 29th and October 3rd 2011, day of year (DOY) 241 to 278, during the main leaf flushing season [2].

The image analysis was conducted by defining different regions of interest (ROI), as described in [1,2,29,30]. For each ROI, a binary image with the same dimensions of the original image was created. These images are later used as masks. A mask of white pixels indicates the ROI, and the remaining area is filled by black pixels. As defined in [2], we selected six ROIs (Fig. 2) of six plant species described as follow: (1) *Aspidosperma tomentosum* (Fig. 2(a)), (2) *Caryocar brasiliensis* (Fig. 2(b)), (3) *Myrcia guianensis* (Fig. 2(c)), (4)

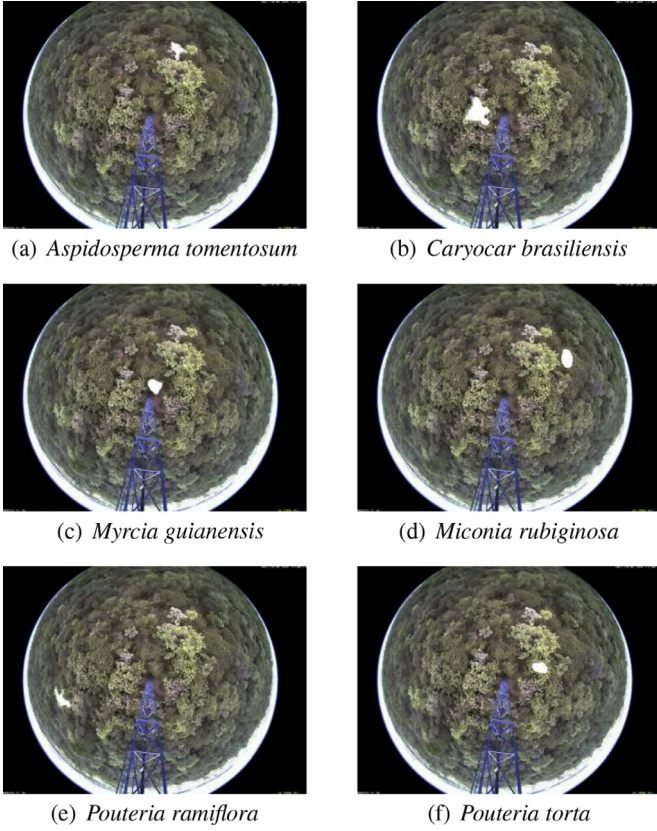


Fig. 2. Regions of interest (ROIs) defined for the analysis of cerrado-savanna digital images: (a) *Aspidosperma tomentosum*, (b) *Caryocar brasiliensis*, (c) *Myrcia guianensis*, (d) *Miconia rubiginosa*, (e) *Pouteria ramiflora*, and (f) *Pouteria torta*.

Miconia rubiginosa (Fig. 2(d)), (5) *Pouteria ramiflora* (Fig. 2(e)), and (6) *Pouteria torta* (Fig. 2(f)).

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 [2]: (i) deciduous: *Aspidosperma tomentosum* and *Caryocar brasiliensis*; (ii) evergreen: *Myrcia guianensis* and *Miconia rubiginosa*; and (iii) semideciduous: *Pouteria ramiflora* and *Pouteria torta*.

4. Visual rhythm-based description

Visual rhythms [25] are an effective way to analyze temporal properties from video data. It consists of an abstraction of a video that encodes the temporal change of pixel values along a specific sampling line [21], as illustrated in Fig. 3(a). In this example, the central column of a set of images are put together to create a single image, the visual rhythm. A clear advantage of this approach is the reduction of the storage space of the extracted features. Therefore, it also speeds up data processing.

Formally, a visual rhythm is a simplification of a video $\mathcal{V} = \{f_t, t \in [1, T]\}$, in domain $2D + t$, with T frames of dimensions $\mathcal{W}_V \times \mathcal{H}_V$, in which each frame f_t is transformed into a vertical line on an image \mathcal{R} , in domain $1D + t$, such that,

$$\mathcal{R}(t, z) = f_t(r_x \times z + a, r_y \times z + b), t \in [1, \mathcal{W}_R], z \in [1, \mathcal{H}_R],$$

where \mathcal{W}_R ($\mathcal{H}_R = T$) and \mathcal{H}_R are its width and height, respectively; r_x and r_y are the sampling rates along the horizontal and vertical directions; a and b are the horizontal and vertical offsets on each frame, respectively.

Without loss of generality, a time series comprised of images taken by digital cameras at fixed time intervals can be viewed as a

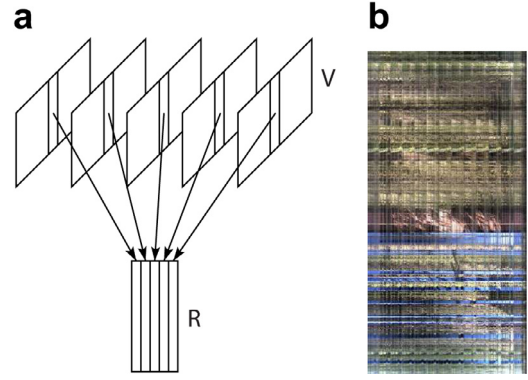


Fig. 3. Visual rhythm: (a) simplification of a video content by mapping each frame into one column of an image; (b) a real example produced by sampling the central vertical line of the digital images.

video of the vegetation. Therefore, a visual rhythm can be used to simplify a time series into a single image, as illustrated in Fig. 3(b). This example shows the visual rhythm produced by sampling the central vertical line of vegetation digital images such as the one showed in Fig. 1. The parameters r_x , r_y , a , and b used to generate the visual rhythm are 0, 1, $\mathcal{W}_V/2$, and 0, respectively. In this paper, we propose to take advantage of existing image descriptors to identify and characterize phenological changes in visual rhythm images.

The major problem with the previous definition of visual rhythms is that it has been designed for the pixel sampling of specific lines (e.g., diagonal, horizontal, and vertical). Here, we are interested in analyzing unshapely regions related to plant species that are identified by phenology experts (see Fig. 2). However, it is impossible to adjust values for the parameters r_x , r_y , a , and b so that we can transform a ROI into a vertical line of a visual rhythm.

The novelty of this paper is to generalize the notion of visual rhythms. From a generic point of view, this approach relies on taking samples of the information to be analyzed and then grouping them in an orderly manner. The key contribution of our idea is the mapping function we design to encode the temporal change of a ROI into a single image. In the following, we present different strategies of generating visual rhythms from time series obtained by digital images of vegetation data.

4.1. Pixel-based visual rhythm

Let $S = \{S_h\}$, $h \in [1, H]$ be a set of H image sequences, in which each $S_h = \{\mathcal{I}_{dh}\}$, $d \in [1, D]$ is composed of D images \mathcal{I}_{dh} , with dimensions $\mathcal{W}_S \times \mathcal{H}_S$, taken by the digital camera at the day of year d and the hour h ; and \mathcal{M} be a binary image, with the same dimensions of S , in which white pixels indicate an area of interest. Fig. 4 shows how pixel-based visual rhythm images are created.

Initially, we convert the binary image \mathcal{M} into a list of Cartesian coordinates $\mathcal{L}_{xy} = \{(x, y) \mid \mathcal{M}(x, y) = 1\}$. Next, we use this list for computing the geometric center (x_c, y_c) of the area of interest. After that, we translate the Cartesian coordinate system of the elements in the list \mathcal{L}_{xy} to have its origin at the point (x_c, y_c) and then we convert them to the polar coordinate system, creating a list of polar coordinates $\mathcal{L}_{r\theta}$. Thereafter, we create an index $\mathcal{K} = \{k \mid \forall (r, \theta) \in \mathcal{L}_{r\theta}, k = 2\pi r + \theta\}$ which assigns a unique value to each element in the list $\mathcal{L}_{r\theta}$. Finally, we sort the keys in the index \mathcal{K} in an increasing order and then we use them to arrange the elements in the list \mathcal{L}_{xy} .

Thus, we can define a visual rhythm as a mapping of an image sequence S_h into a single image \mathcal{R}_h , in which each image \mathcal{I}_{dh} is a column (i.e., vertical line) at the row d , such that

$$\mathcal{R}_h(d, z) = \mathcal{I}_{dh}(\mathcal{L}_{xy}(z)), d \in [1, \mathcal{W}_R], z \in [1, \mathcal{H}_R],$$

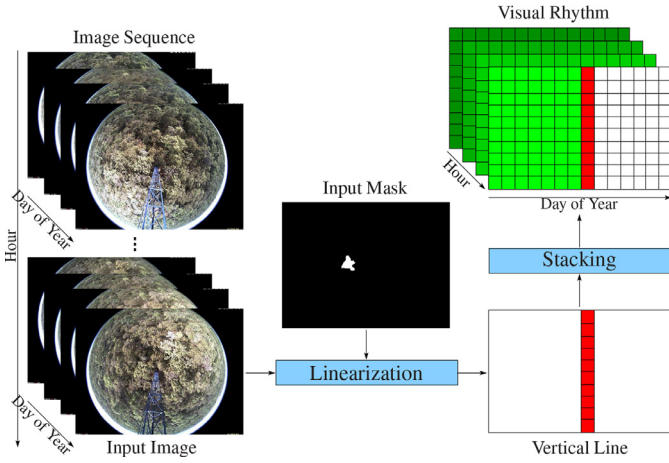


Fig. 4. Overview of the pixel-based strategy. For each hour, the pixel set of the segmented region is linearized. At the end of the process, the pixel values of the segmented region along the time is a column in a new image: the visual rhythm. One visual rhythm is computed for each hour of the day along the time.

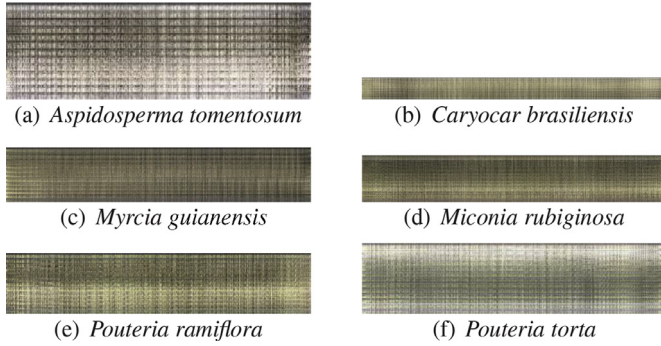


Fig. 5. Visual rhythms obtained for each ROI by using the pixel-based strategy.

where $\mathcal{W}_R = D$ and $\mathcal{H}_R = |\mathcal{L}_{xy}|$ are its width and height, respectively. Fig. 5 presents the visual rhythms produced by the pixel sampling of the digital images using each ROI from Fig. 2.

4.2. Area-based visual rhythm

Let $\mathcal{S} = \{\mathcal{I}_{dh}\}$, $d \in [1, D]$, $h \in [1, H]$ be an image sequence composed of $D \times H$ images \mathcal{I}_{dh} , with dimensions $W_S \times H_S$, taken by the digital camera at the day of year d and the hour h ; and \mathcal{M} be a binary image, with the same dimensions of \mathcal{S} , in which white pixels indicate an area of interest. Fig. 6 shows how area-based visual rhythm images are created.

Initially, we convert the binary image \mathcal{M} into a list of Cartesian coordinates $\mathcal{L}_{xy} = \{(x, y) \mid \mathcal{M}(x, y) = 1\}$. After that, we use this list to draw a sample of the pixels from an input image \mathcal{I}_{dh} . Finally, we extract a feature \mathcal{F}_{dh} that uniquely characterizes the natural distribution of all those pixels by calculating color moments of this segmented region.

Here, we use the three central moments of a color distribution [37]. The first-order moment can be interpreted as the average color intensity, and it can be calculated by using the formula:

$$\mathcal{F}_{dh} = E_{dh} = \frac{\sum_{(x,y) \in \mathcal{L}_{xy}} \mathcal{I}_{dh}(x, y)}{|\mathcal{L}_{xy}|}.$$

The second-order moment is the standard deviation, which is obtained by taking the square root of the variance of the color dis-

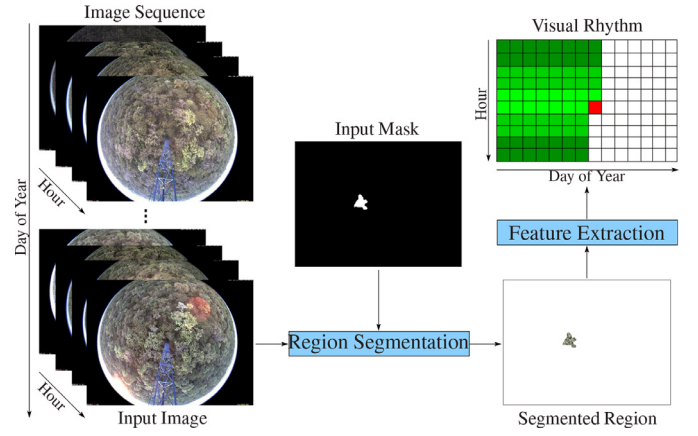


Fig. 6. Overview of the area-based strategy. This representation encodes the color intensity of each region in each hour (vertically) along the year (horizontally). The color intensity of a entire region is computed by using statistical moments.

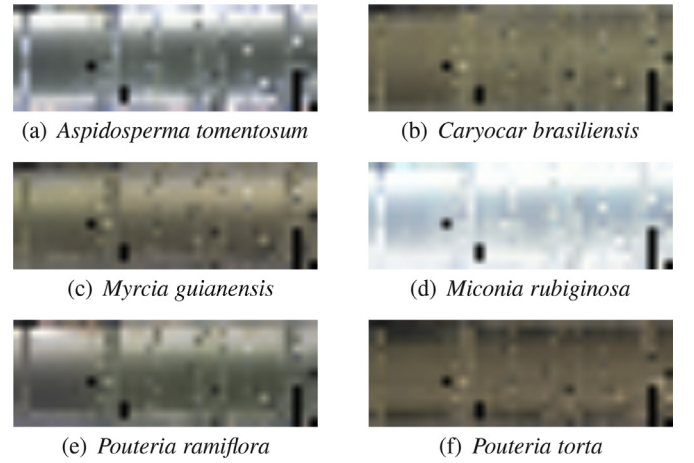


Fig. 7. Visual rhythms obtained for each ROI by using the area-based strategy. Each ROI was encoded by the first-order moment.

tribution, i.e.,

$$\mathcal{F}_{dh} = \sigma_{dh} = \sqrt{\frac{\sum_{(x,y) \in \mathcal{L}_{xy}} [\mathcal{I}_{dh}(x, y) - E_{dh}]^2}{|\mathcal{L}_{xy}|}}.$$

The third-order moment is the skewness. It measures how asymmetric the color distribution is, and thus it gives information about the shape of the color distribution. It can be computed with the following formula:

$$\mathcal{F}_{dh} = s_{dh} = \sqrt[3]{\frac{\sum_{(x,y) \in \mathcal{L}_{xy}} [\mathcal{I}_{dh}(x, y) - E_{dh}]^3}{|\mathcal{L}_{xy}|}}.$$

Thus, we can define a visual rhythm as a mapping of an image sequence \mathcal{S} into a single image \mathcal{R} , in which each feature \mathcal{F}_{dh} is a pixel at the position (d, h) , i.e.,

$$\mathcal{R}(d, h) = \mathcal{F}_{dh}, d \in [1, W_R], h \in [1, H_R],$$

where $W_R = D$ and $H_R = H$ are its width and height, respectively. Figs. 7–9 present the visual rhythms produced by the pixel sampling of the digital images using each ROI from Fig. 2, where each ROI was encoded by the first-order, second-order, and third-order moments, respectively.

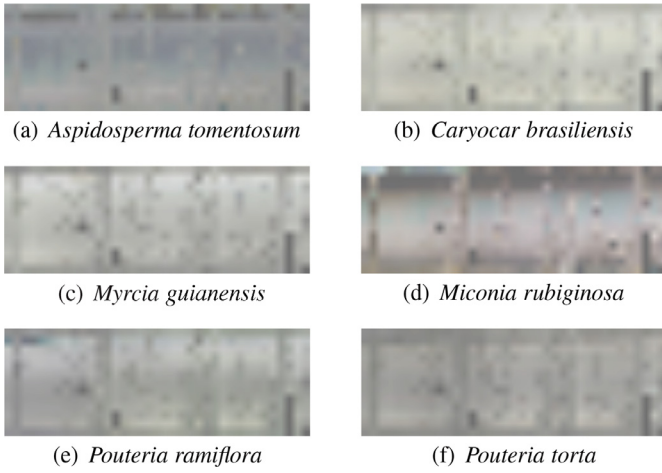


Fig. 8. Visual rhythms obtained for each ROI by using the area-based strategy. Each ROI was encoded by the second-order moment.

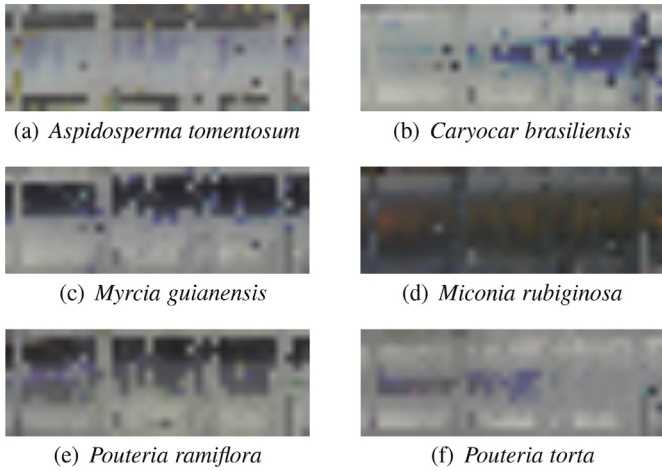


Fig. 9. Visual rhythms obtained for each ROI by using the area-based strategy. Each ROI was encoded by the third-order moment.

5. Experimental protocol

This section presents the adopted experimental protocol. First, we introduce the region-based image retrieval scenario used in our evaluation in Section 5.1. Next, we present the used evaluation metrics and the baseline considered in our study in Sections 5.2 and 5.3, respectively.

5.1. Image retrieval protocol

We carried out experiments to identify plant species in the image using the proposed visual rhythm representations. In this work, we approach the plant identification as an image retrieval problem, in contrast to some initiatives that have addressed this task in the context of image classification [5,6]. Our objective is to use the proposed representations in search services that could leverage the understanding of phenological changes over time by providing areas of plant individuals whose visual features are similar to those of a region of interest defined as input.

We adopted a content-based region retrieval approach in our evaluation protocol. This approach relies on the execution of *similarity searches* [41], according to which image regions are ranked in order of their distance from a given query region. From each image region, feature vectors are extracted by taking into account different representations (e.g., pixel-based or area-based visual rhythm)

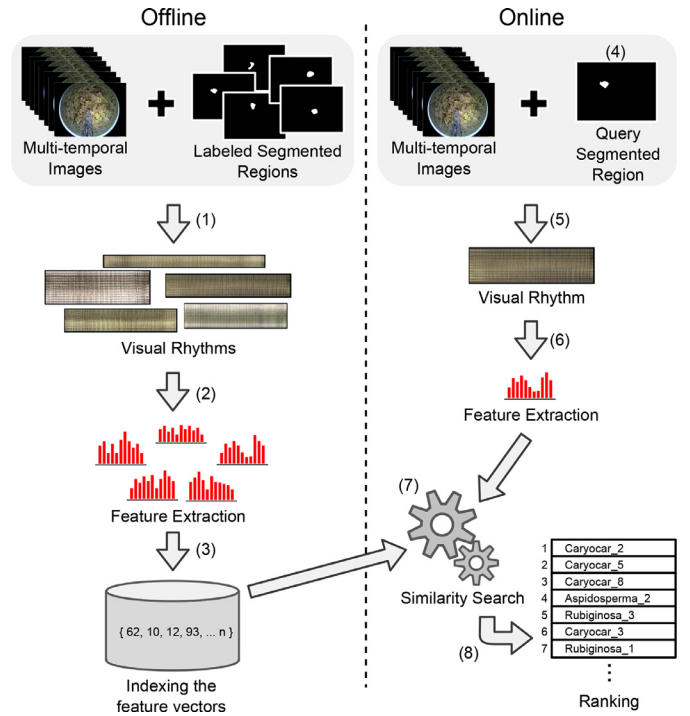


Fig. 10. Flowchart of a content-based region retrieval system used in the evaluation protocol.

and descriptors (e.g., Global Color Histogram). Two regions are considered similar to each other, if the distance of their feature vectors are small. The more effective a descriptor is, the more relevant image regions are ranked at top positions of the returned ranked list.

The flowchart of the content-based region retrieval system is illustrated in Fig. 10. The process is composed of offline and online steps. The offline steps comprise: (1) the representation of each labeled segmented regions in the multitemporal data by using visual rhythms; (2) the extraction of features from each visual rhythm through descriptors; and (3) the indexing of features in a data repository. The online steps consists of a query search composed of the following steps: (4) the selection of a query pattern, which is a segmented region along the multitemporal images; (5) the computation of the query's visual rhythm; (6) the extraction of features by using descriptors; (7) the search computation by similarity; and (8) the final similarity ranking, including all patterns learned at the offline stages.

In this system, we provide a time series extracted from an image area associated with a given species and we query for similar time series computed from other image areas that belong to the same species. For describing time series encoded into a visual rhythm, we used six traditional and recently proposed image descriptors: Auto Color Correlogram (ACC) [18], Color Coherent Vector (CCV) [26], Border/Interior pixel Classification (BIC) [36], and Global Color Histogram (GCH) [38], for encoding color information; Generic Fourier Descriptor (GFD) [42] and Haar-Wavelet Descriptor (HWD) [20], for analyzing spectral properties. The distance function used for feature comparison is the Manhattan (L_1) distance. For more details regarding those image descriptors, refer to [27].

Our strategy to evaluate image descriptors in the context of time series description is based on assessing the similarity among regions associated with individuals of a same species. For that, we used the Guigues algorithm [15] to segment the hemispheric image into small polygons, obtaining 8, 849 segmented regions (SR). Then, we associated each SR with a single ROI aiming to label it. A

Table 1

MAP scores obtained by each of the image descriptors along all the available periods of the day.

Hour	Pixel-based visual rhythm						Baseline		
	ACC	BIC	CCV	GCH	GFD	HWD	RGB	ExG	NDI
6	0.587	0.601	0.601	0.608	0.338	0.374	0.739	0.523	0.490
7	0.561	0.555	0.549	0.545	0.353	0.377	0.720	0.515	0.569
8	0.571	0.555	0.554	0.546	0.355	0.385	0.716	0.480	0.532
9	0.572	0.574	0.548	0.569	0.353	0.380	0.733	0.422	0.619
10	0.620	0.612	0.588	0.614	0.354	0.382	0.741	0.407	0.644
11	0.607	0.586	0.562	0.585	0.345	0.372	0.744	0.398	0.640
12	0.535	0.528	0.519	0.534	0.342	0.356	0.731	0.381	0.632
13	0.542	0.526	0.506	0.513	0.334	0.364	0.718	0.407	0.628
14	0.577	0.577	0.553	0.554	0.341	0.387	0.723	0.419	0.625
15	0.567	0.562	0.542	0.542	0.346	0.395	0.718	0.427	0.633
16	0.558	0.548	0.530	0.524	0.347	0.386	0.700	0.434	0.615
17	0.554	0.555	0.531	0.548	0.358	0.388	0.686	0.464	0.610
18	0.576	0.590	0.596	0.595	0.338	0.376	0.688	0.498	0.656

Table 2

P@5 scores obtained by each of the image descriptors along all the available periods of the day.

Hour	Pixel-based visual rhythm						Baseline		
	ACC	BIC	CCV	GCH	GFD	HWD	RGB	ExG	NDI
06	0.761	0.795	0.769	0.767	0.451	0.502	0.878	0.779	0.728
07	0.779	0.778	0.729	0.782	0.495	0.523	0.922	0.764	0.848
08	0.805	0.742	0.713	0.749	0.506	0.513	0.878	0.724	0.777
09	0.778	0.757	0.706	0.735	0.500	0.521	0.856	0.661	0.800
10	0.807	0.824	0.788	0.818	0.494	0.532	0.834	0.642	0.816
11	0.833	0.817	0.765	0.770	0.466	0.503	0.852	0.639	0.806
12	0.785	0.743	0.728	0.749	0.471	0.449	0.860	0.620	0.809
13	0.810	0.788	0.743	0.775	0.463	0.466	0.835	0.675	0.792
14	0.780	0.768	0.763	0.794	0.475	0.517	0.872	0.702	0.814
15	0.714	0.711	0.751	0.736	0.466	0.523	0.862	0.700	0.798
16	0.697	0.729	0.732	0.725	0.480	0.518	0.866	0.676	0.800
17	0.738	0.740	0.724	0.762	0.505	0.538	0.866	0.676	0.817
18	0.798	0.787	0.788	0.777	0.432	0.520	0.906	0.720	0.908

labeled region is created if there is at least 80% of overlapped area between a SR and a ROI. In the remainder of this paper, when we refer to regions of interest related to tree crowns of plant species identified manually in the digital image, we use the acronym ROI; and when we refer to segmented regions obtained from the segmentation algorithm, we use the acronym SR. The similarity between two SRs is computed as a function of the distance between the feature vectors extracted from their visual rhythms. An image descriptor is better than another if it ranks more SRs belonging to the same ROI of a query SR at the first positions.

For each ROI, we randomly selected twenty percent of its total number of SRs to be used as queries. Five replications were performed in order to ensure statistically sound results. Presented results consider the average performance of the evaluated image descriptors, which were computed based on the mean and standard deviation of each replication.

5.2. Evaluation metrics

We assess the effectiveness of each approach using the metrics of Precision and Recall. Precision is the ratio of the number of relevant SRs retrieved to the total number of irrelevant and relevant SRs retrieved. Recall is the ratio of the number of relevant SRs retrieved to the total number of relevant SRs in the database. Here, a given SR is considered as relevant only if it belongs to the same ROI of a query SR. However, there is a trade-off between Precision and Recall. Greater Precision decreases Recall and greater Recall leads to decreased Precision. So, we choose to report the results using unique-value measurements: Mean Average Precision (MAP), which is the mean of the precision scores obtained at the ranks of each relevant SR; and Precision at 5 (P@5), which is the average precision after 5 SRs are returned. These metrics combine Preci-

sion and Recall into a single measure, which makes the comparison easier.

5.3. Baseline

We compare the visual rhythm-based techniques against three approaches widely used by phenology experts for characterizing leaf-changing patterns of plant species from digital images. The first approach is a normalized index called RGB chromatic coordinates (RGBcc), which was developed by Gillespie et al. [14] and is considered the most efficient index to detect the color of plants in relation to their background [40]. The normalized RGBcc index undergoes a nonlinear transform, as follows: $r = R/(R + G + B)$, $g = G/(R + G + B)$, $b = B/(R + G + B)$; where R , G , and B are the average pixel intensity of the red, green, and blue bands, respectively.

The second approach is a contrast index named as Excess Green (ExG), which was introduced by Woebbecke et al. [40] and is commonly applied to separate green plants from soil and residue background. The ExG index is defined as: $ExG = 2g - r - b$. Similarly, the third approach, known as Normalized Difference Index (NDI), uses only green and red channels and is given by [39]: $NDI = (G - R)/(G + R)$.

6. Experimental results

The objective of our evaluation is to confirm that the use of the proposed visual rhythm representations yields comparable results, in terms of search result effectiveness, when compared with the traditional RGBcc index. The evaluation results are discussed in Sections 6.1 and 6.2, for the pixel-based and area-based visual rhythm representations, respectively. The correlation analysis and feature combination between different approaches is discussed in

Table 3

Differences between MAP of the different image descriptors by considering the best result of each approach.

Approach	Confidence interval (99%)	
	Min.	Max.
RGB@11h - ACC@10h	−0.236	0.401
RGB@11h - BIC@10h	−0.291	0.461
RGB@11h - CCV@06h	−0.290	0.475
RGB@11h - GCH@10h	−0.282	0.441
RGB@11h - GFD@17h	0.022	0.744
RGB@11h - HWD@15h	−0.045	0.757

Table 4

Differences between P@5 of the different image descriptors by considering the best result of each approach.

Approach	Confidence interval (99%)	
	Min.	Max.
RGB@07h - ACC@11h	0.012	0.378
RGB@07h - BIC@10h	0.085	0.298
RGB@07h - CCV@10h	−0.001	0.316
RGB@07h - GCH@10h	−0.006	0.207
RGB@07h - GFD@08h	0.201	0.750
RGB@07h - HWD@17h	0.109	0.730

Section 6.3. Finally, in **Section 6.4**, we discuss on efficiency aspects of the visual rhythm-based techniques, highlighting the main advantages of using the proposed approaches.

6.1. Pixel-based visual rhythm

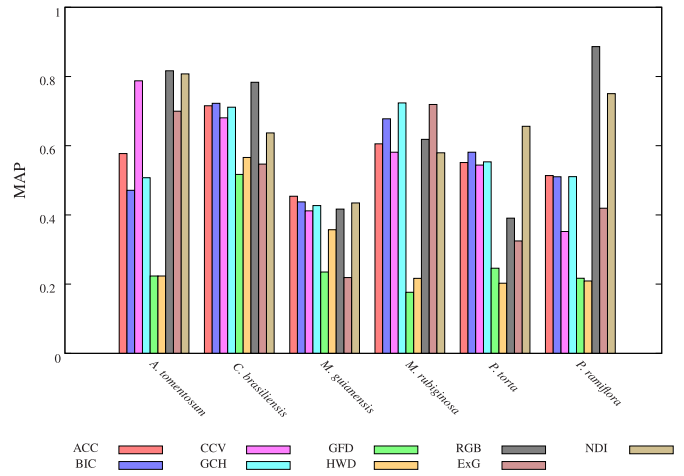
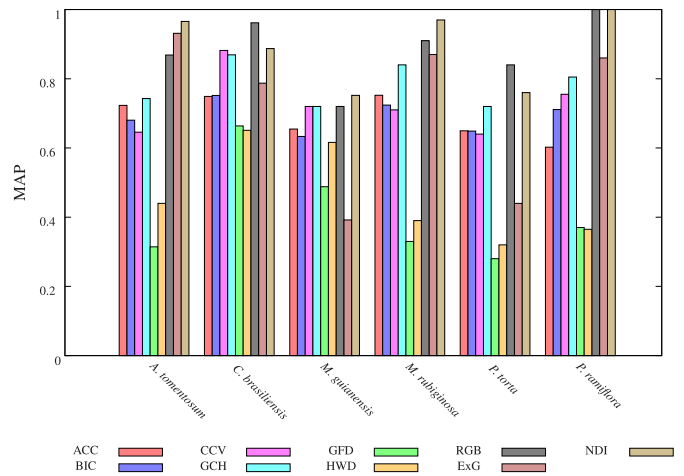
Tables 1 and 2 compare the pixel-based visual rhythm techniques and the baseline methods with respect to the MAP and P@5 measures, respectively. MAP is a good indication of the effectiveness considering all positions of obtained ranked lists. P@5, in turn, focuses on the effectiveness of the methods considering only the first positions of the ranked lists. For each approach, we highlight the hour of day that provided the best result.

Those results indicate that the performance of the different evaluated approaches is similar, with a small advantage to the RGB-based baseline. Notice that early hours (from 8h to 11h) are better to characterize the phenological changes of plant species by using color descriptors. As we can observe, the best performances were achieved using the digital images taken at ten in the morning. This finding disagrees with the general suggestion of extracting color information from midday hours (from 11h to 14h) for ecological studies [1,19,29].

Paired *t*-tests were performed to verify the statistical significance of those results. For that, the confidence intervals for the differences between paired means of each ROI were computed to compare every pair of approaches. If the confidence interval includes zero, the difference is not significant at that confidence level. If the confidence interval does not include zero, then the sign of the difference indicates which alternative is better.

Tables 3 and 4 present the 99% confidence intervals of the differences between the RGB-based baseline and the pixel-based visual rhythm techniques for the MAP and P@5 measures, respectively. For simplicity and readability purposes, we report only the results for the hour of day that provided the best result of each approach. Such analyses confirm that the pixel-based visual rhythm techniques and the RGB-based baseline exhibit similar performance. Notice that the confidence intervals include zero and, hence, the differences between those approaches are not significant at that confidence level.

Figs. 11 and 12 compare the individual scores obtained for each ROI considering the best results of the evaluated methods in terms

**Fig. 11.** MAP scores obtained for each ROI.**Fig. 12.** P@5 scores obtained for each ROI.

of the MAP and P@5 measures, respectively. It is interesting to note the differences in responsiveness of the different approaches with respect to each of the species individually. The main reason for those results is the different patterns of the leaf color change of each species. In general, different image descriptors are designed to capture different visual features.

6.2. Area-based visual rhythm

In **Fig. 13**, we compare the effectiveness of the baseline methods and the area-based visual rhythm techniques by considering different image descriptors. The graphs present the individual scores obtained for visual rhythms encoded by the first-order (first column), second-order (second column), and third-order (third column) moments. We show the results for the MAP (top row) and P@5 (bottom row) measures.

In general, those graphs demonstrate that visual rhythms encoded by lower order moments (left column) outperform the higher order ones (right column). On the other hand, for a same statistics, the performance of different image descriptors of a same type (color or texture) is similar. Unlike the results obtained for the pixel-based visual rhythm techniques, the texture descriptors (GFD and HWD) are more effective than the color ones (ACC, BIC, CCV, and GCH).

Table 5 presents the 99% confidence intervals of the differences between the RGB-based baseline and the area-based visual rhythm

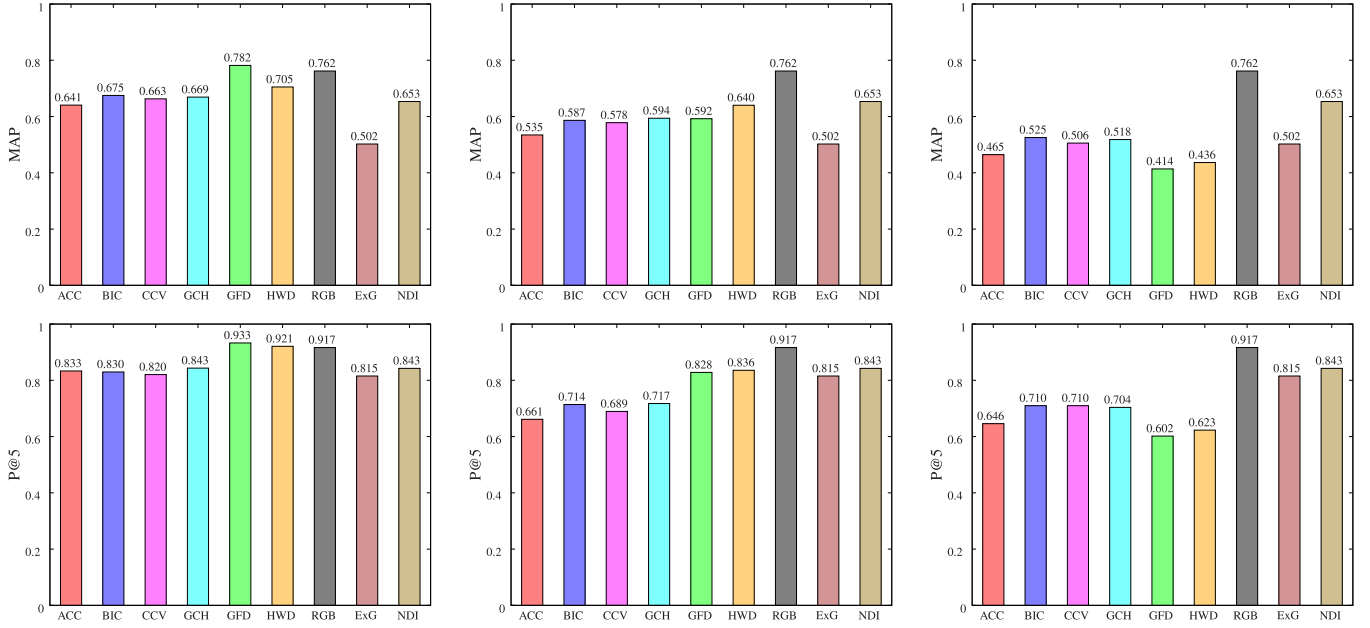


Fig. 13. Comparison between the effectiveness measures obtained by the evaluated approaches. These graphs present the results for visual rhythms encoded by the first-order (first column), second-order (second column), and third-order (third column) moments. They report the MAP (top row) and P@5 (bottom row) scores.

Table 5

Differences between MAP and P@5 scores of the different image descriptors by considering visual rhythms encoded by different color moments.

Method		MAP		P@5	
		Min.	Max.	Min.	Max.
First-Order	RGB - ACC	-0.062	0.311	-0.084	0.407
	RGB - BIC	-0.087	0.267	-0.047	0.314
	RGB - CCV	-0.048	0.276	-0.097	0.437
	RGB - GCH	-0.086	0.306	-0.093	0.360
	RGB - GFD	-0.181	0.145	-0.094	0.059
	RGB - HWD	-0.134	0.198	-0.141	0.161
Second-Order	RGB - ACC	0.010	0.475	-0.020	0.654
	RGB - BIC	-0.066	0.434	0.001	0.513
	RGB - CCV	-0.089	0.445	-0.031	0.529
	RGB - GCH	-0.105	0.457	-0.015	0.494
	RGB - GFD	-0.134	0.444	-0.066	0.265
	RGB - HWD	-0.199	0.361	-0.125	0.239
Third-Order	RGB - ACC	0.082	0.520	0.131	0.518
	RGB - BIC	0.087	0.384	0.035	0.496
	RGB - CCV	0.098	0.414	0.041	0.511
	RGB - GCH	0.083	0.386	0.071	0.471
	RGB - GFD	0.029	0.566	0.050	0.550
	RGB - HWD	-0.002	0.540	0.065	0.570

techniques for the MAP and P@5 measures, respectively. Such analyses confirm that the area-based visual rhythm techniques and the RGB-based baseline exhibit similar performance. Note that the confidence intervals include zero and, hence, the differences between those approaches are not significant at that confidence level.

In Fig. 14, we compare the individual scores obtained for each ROI in terms of the MAP (top row) and P@5 (bottom row) measures, respectively. We show the results for visual rhythms encoded by the first-order (first column), second-order (second column), and third-order (third column) moments by considering different image descriptors. Notice the differences in responsiveness of the evaluated methods with respect to each of the species individually. For instance, despite the overall performance of the first-order visual rhythms outperform the second-order ones, these latter have achieved the best results for the *Pouteria torta*. On the other hand, they have obtained the worst results for the *Myrcia guianensis*.

This behavior reflects their contrasting leaf phenology [2]: the *Myrcia guianensis* is an evergreen species and, therefore, the leaf senescence is a continuous process and color changes are more subtle over time; in contrast, the *Pouteria torta* is semideciduous, thus the color change reflects the rapid leaf senescence and the flush of new leaves.

6.3. Correlation analysis and descriptor combination

One important issue with regard the evaluation of multiple representations and descriptors concerns the investigation of their correlation. The objective is to somehow confirm if the different representations/descriptions provide complementary information regarding the image visual properties.

Figs. 15 (a) and 16(a) present the correlation among the ranked lists defined by the most effective descriptors associated with the pixel-based and area-based visual rhythm representations, respectively. The correlation score is computed using the Kendall rank correlation coefficient, defined as:

$$\tau(x, y) = \frac{(P - Q)}{\sqrt{(P + Q + T) \times (P + Q + U)}}$$

where P is the number of concordant pairs, Q the number of discordant pairs, T the number of ties only in rank x , and U the number of ties only in rank y . If a tie occurs for the same pair in both x and y , it is not added to either T or U .

The RGB-based baseline is referred at the first line and column in both figures. Notice that the proposed descriptors are not correlated to each other. More importantly, they are not correlated to the RGB-based baseline. That opens a novel possibility of investigation concerning their combination.

In this sense, we have combined the ranked lists associated with the two best descriptors for the pixel-based and area-based visual rhythm representations with the RGB-based baseline using the traditional Borda Count rank aggregation approach [9]. The Borda Count algorithm is an order-based method, according to which a score is assigned to an element x in the ranked list r_i , equal to $|r_i| - r_i(x)$, where $r_i(x)$ is the position of the element x in r_i . The final score of an element is the sum of the scores obtained in each ranked list.

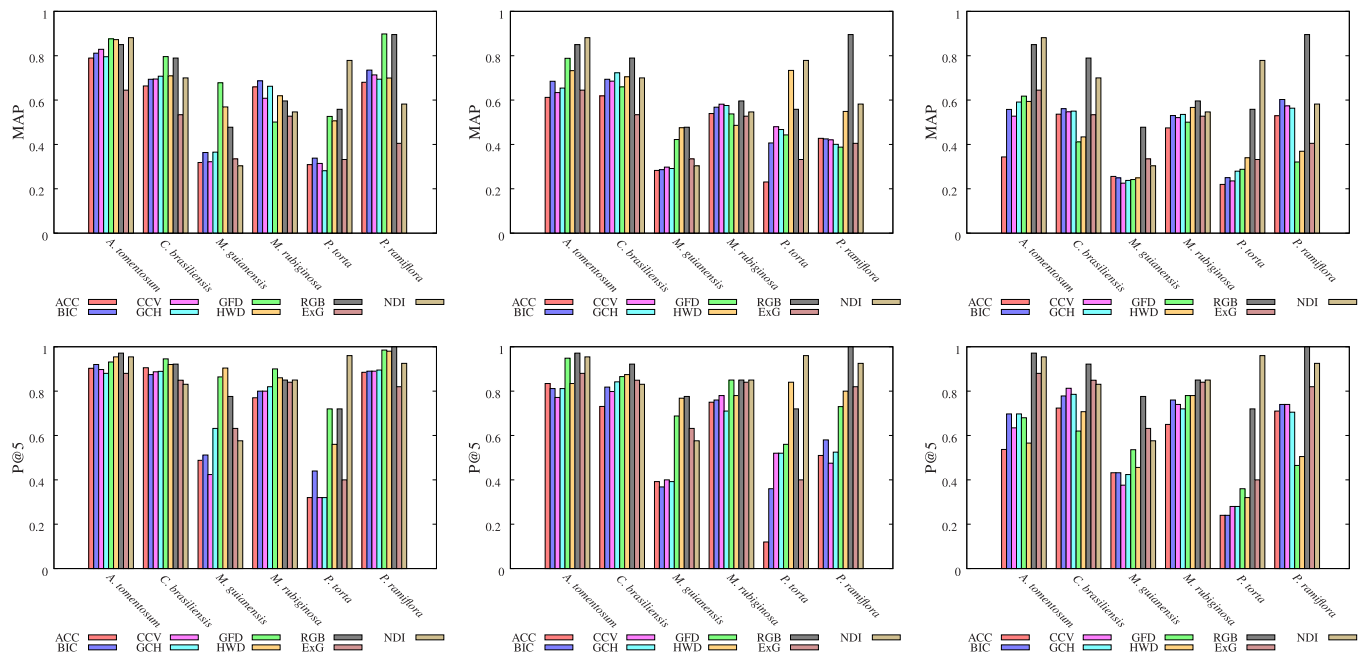


Fig. 14. Individual scores obtained for each ROI. These graphs present the results for visual rhythms encoded by the first-order (first column), second-order (second column), and third-order (third column) moments. They report the MAP (top row) and P@5 (bottom row) measures.

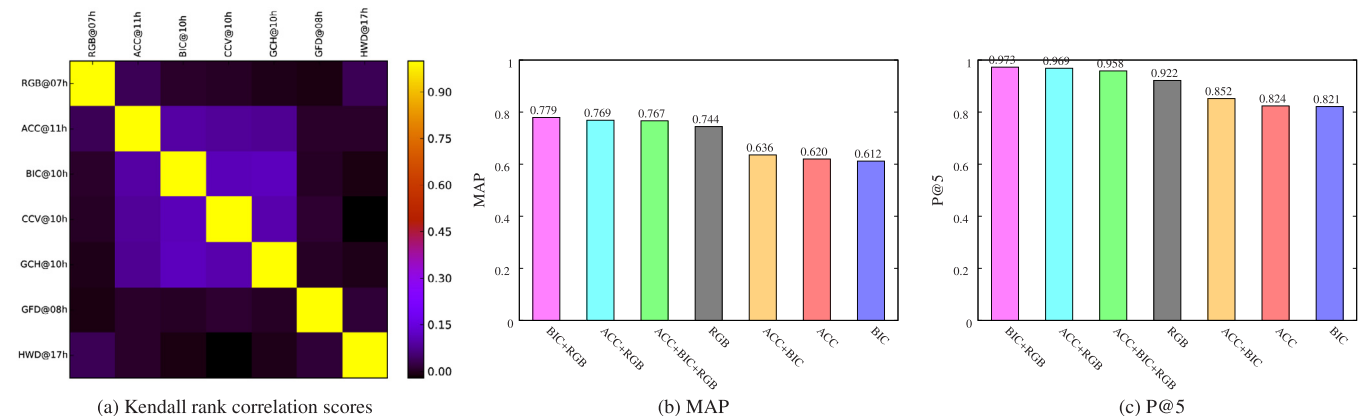


Fig. 15. (a) Correlation scores and (b and c) Combination results of the two best pixel-based descriptions with the RGB-based baseline.

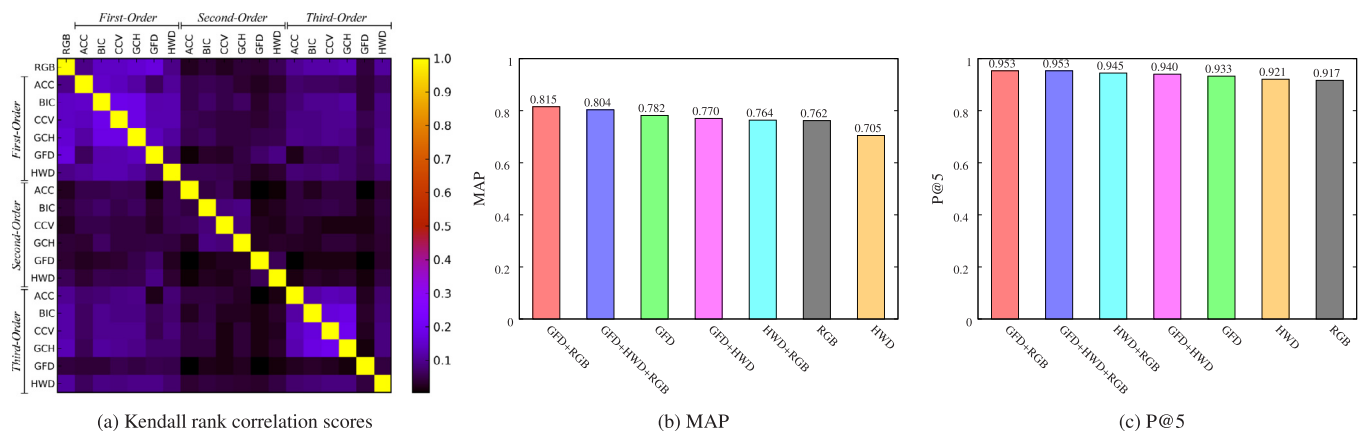


Fig. 16. (a) Correlation scores and (b and c) Combination results of the two best area-based descriptions with the RGB-based baseline.

Table 6

Computational costs and space requirements of each of the evaluated approaches.

Method		Computational cost		Space requirements
		Extraction	Matching	
Visual rhythm	VR + ACC	$O(n)$	$O(1)$	$O(1)$
	VR + BIC	$O(n)$	$O(1)$	$O(1)$
	VR + CCV	$O(n)$	$O(1)$	$O(1)$
	VR + GCH	$O(n)$	$O(1)$	$O(1)$
	VR + GFD	$O(n \log n)$	$O(1)$	$O(1)$
	VR + HWD	$O(n \log n)$	$O(1)$	$O(1)$
Baseline	RGB	$O(n)$	$\Omega(n)$	$\Omega(n)$
	ExG	$O(n)$	$\Omega(n)$	$\Omega(n)$
	NDI	$O(n)$	$\Omega(n)$	$\Omega(n)$

Figs. 15 (b and c) and 16(b and c) present the MAP and P10 scores achieved by the most promising combinations for the pixel-based and area-based visual rhythm representations, respectively. As we can observe, even considering a fairly simple approach, the combination of the proposed methods with the RGB-based baseline improved the effectiveness results both in terms of MAP and P@5. These results show the potential of the idea, opening a new world of research possibilities.

6.4. Computational efficiency

The key advantage of our technique is its computational efficiency. Table 6 presents the computational cost and the space requirements (in terms of the length n of the time series) of all the compared methods. In this way, we can investigate the relative difference of performance among different approaches.

Clearly, the visual rhythm-based techniques are much more efficient than the current solutions. This improvement makes our approach suitable for long-term collections of image data.

Note also that the larger the time series, the bigger would be the visual rhythm image generated, independently of the representation considered (either the pixel-based or area-based approach). For larger time series, the feature extraction process will probably take more time. Note, however, that the size of the final feature vector generated is not dependent on the size of the input image. It only depends on the descriptor used to characterize the visual properties of the visual rhythm images.

7. Conclusions

In this paper, we have presented a novel approach for capturing phenological patterns from time series and distinguishing the behavior of plant species. Our technique relies on encoding time series as a visual rhythm, which is characterized by image descriptors. The improvement of the computational efficiency makes our method suitable for long-term temporal data.

We have validated our technique using about 2700 images, taken from a tropical cerrado-savanna vegetation, including a high diversity of plant species. Experimental results obtained by the application of our method with several image descriptors show that it presents high accuracy and computational speed when identifying regions that belong to the same species.

Future work includes the evaluation of other visual features for image retrieval. In addition, the proposed method can be augmented to consider temporal segmentation and/or summarization methods. Finally, we also plan to consider learning-to-rank methods (e.g., genetic programming) for combining different descriptors.

Acknowledgments

This research was supported by São Paulo Research Foundation FAPESP–Microsoft Research Virtual Institute (grants #2010/52113-

5, #2013/50169-1 and #2013/50155-0), CNPq (grant 449638/2014-6), FAPESP (grants #2007/52015-0, #2007/59779-6, #2009/18438-7, #2010/51307-0), and FAPEMIG (grant APQ-00768-14). BA received a doctoral fellowship from FAPESP (grant #2014/00215-0); LPCM and RST receive a Productivity Research Fellowship from CNPq (grants 310761/2014-0 and 306587/2009-2) and CNPq PVE (grant 400717/2013-1).

References

- [1] H. Ahrends, S. Etzold, W. Kutsch, R. Stoeckli, R. Bruegger, F. Jeanneret, H. Wanner, N. Buchmann, W. Eugster, Tree phenology and carbon dioxide fluxes: Use of digital photography for process-based interpretation at the ecosystem scale, *Climate Res.* 39 (2009) 261–274.
- [2] B. Albetton, J. Almeida, R. Henneken, R. da S. Torres, A. Menzel, L.P.C. Morellato, Using phenological cameras to track the green up in a cerrado savanna and its on-the-ground validation, *Ecol. Inform.* 19 (2014) 62–70.
- [3] J. Almeida, J.A. dos Santos, B. Albetton, L.P.C. Morellato, R. da S. Torres, Plant species identification with phenological visual rhythms, in: *Proceedings of the IEEE International Conference on eScience (eScience'13)*, 2013a, pp. 148–154.
- [4] J. Almeida, J.A. dos Santos, B. Albetton, L.P.C. Morellato, R. da S. Torres, Visual rhythm-based time series analysis for phenology studies, in: *Proceedings of the IEEE International Conference Image Process. (ICIP'13)*, 2013b, pp. 4412–4416.
- [5] J. Almeida, J.A. dos Santos, B. Albetton, R. da S. Torres, L.P.C. Morellato, Remote phenology: Applying machine learning to detect phenological patterns in a cerrado savanna, in: *Proceedings of the IEEE International Conference on (eScience'12)*, 2012, pp. 1–8.
- [6] J. Almeida, J.A. dos Santos, B. Albetton, R. da S. Torres, L.P.C. Morellato, Applying machine learning based on multiscale classifiers to detect remote phenology patterns in cerrado savanna trees, *Ecol. Inform.* 23 (2014) 49–61.
- [7] J. Almeida, J.A. Santos, W.O. Miranda, B. Albetton, L.P.C. Morellato, R.S. Torres, Deriving vegetation indices for phenology analysis using genetic programming, *Ecol. Inform.* 26 (2015) 61–69.
- [8] J.P. Ardila, W. Bijker, V.A. Tolpekin, A. Stein, Multitemporal change detection of urban trees using localized region-based active contours in vhr images, *Remote Sens. Environ.* 124 (2012) 413–426.
- [9] J.C. Borda, Mémoire sur les élections au scrutin, *Histoire de l'Académie Royale des Sciences*, 1781.
- [10] E.B. Brooks, V.A. Thomas, R.H. Wynne, J.W. Coulston, Fitting the multitemporal curve: A fourier series approach to the missing data problem in remote sensing analysis, *IEEE Trans. Geosci. Remote Sens.* 50 (2012) 3340–3353.
- [11] J.C. Conti, F.A. Faria, J. Almeida, B. Albetton, L.P.C. Morellato, L. Camolesi Jr., R. da S. Torres, Evaluation of time series distance functions in the task of detecting remote phenology patterns, in: *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR'14)*, 2014, pp. 3126–3131.
- [12] J. Dierenbach, F.W. Badeck, J. Schaber, The plant phenological online database (ppodb): An online database for long-term phenological data, *Int. J. Biometeorol.* 57 (2013) 805–812.
- [13] M. Forster, T. Schmidt, C. Schuster, B. Kleinschmit, Multi-temporal detection of grassland vegetation with rapid-eye imagery and a spectral-temporal library, in: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'12)*, 2012, pp. 4930–4933.
- [14] A.R. Gillespie, A.B. Kahle, R.E. Walker, Color enhancement of highly correlated images. ii. channel ratio and “chromaticity” transformation techniques, *Remote Sens. Environ.* 22 (1987) 343–365.
- [15] L. Guigues, J. Cocquerez, H. Le Men, Scale-sets image analysis, *Intl. J. Comput. Vis.* 68 (2006) 289–317.
- [16] R. Henneken, V. Dose, C. Schleip, A. Menzel, Detecting plant seasonality from webcams using bayesian multiple change point analysis, *Agr. Forest Meteorol.* 168 (2013) 177–185.
- [17] G. Hmimina, E. Dufrêne, J.Y. Pontailler, N. Delpierre, M. Aubinet, B. Caquet, A. de Grandcourt, B. Burban, C. Flechard, A. Granier, P. Gross, B. Heinesch, B. Longdoz, C. Moureaux, J.M. Ourcival, S. Rambal, L.S. Andr, K. Soudani, Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements, *Remote Sens. Environ.* 132 (2013) 145–158.
- [18] J. Huang, R. Kumar, M. Mitra, W.J. Zhu, R. Zabih, Image indexing using color correlograms, in: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997, pp. 762–768.
- [19] R. Ide, H. Oguma, A cost-effective monitoring method using digital time-lapse cameras for detecting temporal and spatial variations of snowmelt and vegetation phenology in alpine ecosystems, *Ecol. Inform.* 16 (2013) 25–34.
- [20] C.E. Jacobs, A. Finkelstein, D. Salesin, Fast multiresolution image querying, in: *Proceedings of the International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'95)*, 1995, pp. 277–286.
- [21] J.S. Lee, T. Ebrahimi, Perceptual video compression: A survey, *IEEE J. Sel. Topics Signal Process.* 6 (2012) 684–697.
- [22] D. Loustau, A. Bosc, A. Colin, H. Davi, C. François, E. Dufrêne, M. Équé, E. Cloppet, D. Arrouays, C. Le Bas, N. Saby, G. Pignard, N. Hamza, A. Granier, N. Breda, P. Ciais, N. Viovy, J. Ogée, J. Delage, Modeling the climate change effects on the potential reduction of french plains forests at the sub regional level, *Tree Physiol* 25 (2005) 813–823.

- [23] X. Ma, A. Huete, Q. Yu, N.R. Coupe, K. Davies, M. Broich, P. Ratana, J. Beringer, L.B. Hutley, J. Cleverly, N. Boulain, D. Eamus, Spatial patterns and temporal dynamics in savanna vegetation phenology across the north australian tropical transect, *Remote Sens. Environ.* 139 (2013) 97–115.
- [24] G.C.S. Negi, Leaf and bud demography and shoot growth in evergreen and deciduous trees of central himalaya, india, *Trees* 20 (2006) 416–429.
- [25] C.W. Ngo, T.C. Pong, R.T. Chin, Detection of gradual transitions through temporal slice analysis, in: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'99)*, 1999, pp. 1036–1041.
- [26] G. Pass, R. Zabih, J. Miller, Comparing images using color coherence vectors, in: *Proceedings of the ACM International Conference on Multimedia (ACM-MM'96)*, 1996, pp. 65–73.
- [27] O.A.B. Penatti, E. Valle, R. da S. Torres, Comparative study of global color and texture descriptors for web image retrieval, *J. Vis. Commun. Image Represent.* 23 (2012) 359–380.
- [28] F. Petitjean, C. Kurtz, N. Passat, P. GanĀřarski, Spatio-temporal reasoning for the classification of satellite image time series, *Pattern Recognit. Lett.* 33 (2012) 1805–1815.
- [29] A.D. Richardson, B.H. Braswell, D.Y. Hollinger, J.P. Jenkins, S.V. Ollinger, Near-surface remote sensing of spatial and temporal variation in canopy phenology, *Ecol. Appl.* 19 (2009) 1417–1428.
- [30] A.D. Richardson, J.P. Jenkins, B.H. Braswell, D.Y. Hollinger, S.V. Ollinger, M.L. Smith, Use of digital webcam images to track spring greening in a deciduous broadleaf forest, *Oecologia* 152 (2007) 323–334.
- [31] A. Rodrigues, A.R.S. Marcal, M. Cunha, Phenology parameter extraction from time-series of satellite vegetation index data using phenosat, in: *Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing Society (IGARSS'12)*, 2012, pp. 4926–4929.
- [32] R. da S. Torres, M. Hasegawa, S. Tabbone, J. Almeida, J.A. dos Santos, B. Alberton, L.P.C. Morellato, Shape-based time series analysis for remote phenology studies, in: *Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing Society (IGARSS'13)*, 2013, pp. 3598–3601.
- [33] L.C.B. Santos, J. Almeida, J.A. dos Santos, S.J.F. Guimarães, A.A. Araújo, B. Alberton, L.P.C. Morellato, R. da S. Torres, Phenological event detection by visual rhythm dissimilarity analysis, in: *Proceedings of the IEEE International Conference on eScience (eScience'14)*, 2014, pp. 263–270.
- [34] M.D. Schwartz, *Phenology: An Integrative Environmental Science*, second, Springer, 2013.
- [35] M.D. Schwartz, B.C. Reed, M.A. White, Assessing satellite derived start-of-season measures in the coterminous, *Int. J. Climatol.* 22 (2002) 1793–1805.
- [36] R.O. Stehling, M.A. Nascimento, A.X. Falcão, A compact and efficient image retrieval approach based on border/interior pixel classification, in: *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM'02)*, 2002, pp. 102–109.
- [37] M.A. Stricker, M. Orengo, Similarity of color images, in: *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases*, 1995, pp. 381–392.
- [38] M.J. Swain, B.H. Ballard, Color indexing, *Intl J. Comput Vis* 7 (1991) 11–32.
- [39] D.M. Woebbecke, G.E. Meyer, K. Von-Bargen, A.D. Mortensen, Plant species identification, size, and enumeration using machine vision techniques on near-binary images, *SPIE Opt. Agr. For.* 1836 (1992) 208–219.
- [40] D.M. Woebbecke, G.E. Meyer, K. Von-Bargen, A.D. Mortensen, Color indices for weed identification under various soil, residue, and lighting conditions, *Trans. ASAE* 38 (1995) 259–269.
- [41] P. Zezula, G. Amato, V. Dohnal, M. Batko, *Similarity Search: The Metric Space Approach*, first, Springer Publishing Company, Incorporated, 2010.
- [42] D. Zhang, G. Lu, Shape-based image retrieval using generic fourier descriptor, *Signal Process. Image Commun.* 17 (2002) 825–848.