Plant Species Identification with Phenological Visual Rhythms

Jurandy Almeida[1], Jefersson A. dos Santos[1], Bruna Alberton[2],
Leonor Patricia C. Morellato[2], and Ricardo da S. Torres[1]

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

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

***Abstract*—Plant phenology studies recurrent plant life cycles events and is a key component of climate change research. To increase accuracy of observations, new technologies have been applied for phenological observation, and one of the most successful are digital cameras, used as multi-channel imaging sensors to estimate color changes that are related to phenological events. We monitored leaf-changing patterns of a cerrado-savanna vegetation by taken daily digital images. We extract individual plant color information and correlated with leaf phenological changes. To do so, time series associated with plant species were obtained, raising the need of using appropriate tools for mining patterns of interest. In this paper, we present a novel approach for representing phenological patterns of plant species derived from digital images. The proposed method is based on encoding time series as a visual rhythm, which is 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 plant species.**

***Keywords*-remote phenology; digital cameras; image analysis; time series; visual rhythm**

I. Introduction

Plant phenology studies recurrent plant life cycles events and is a key component of climate change research [1]. To increase accuracy of observations, new technologies have been applied for phenological observation, and one of the most successful are digital cameras, used as multi-channel imaging sensors to estimate color changes (RGB channels) that are related to phenological events [2]–[4].

We have been monitored leaf-changing patterns of a tropical cerrado-savanna vegetation by taken daily digital images [5]. 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 regions of interest (ROI) based on the random selection of plant species crowns identified in the digital image. Time series associated with different ROI were obtained, raising the need of using appropriate tools for mining patterns of interest.

In this paper, we present a novel approach for capturing phenological patterns from time series generated from digital images and distinguishing the behavior of plant species. It relies on encoding time series as a visual rhythm [6], which is characterized by traditional and recently proposed image description algorithms. This computationally simple approach opens a new area of investigation related to the use of image descriptors to identify and characterize phenological changes.

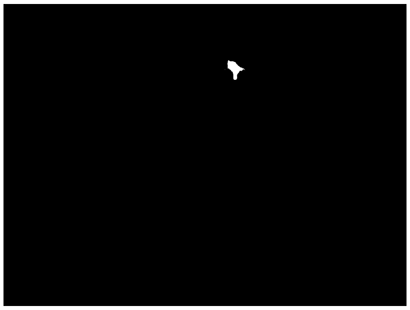
We evaluate the proposed algorithm on about 2,700 images, recorded during the main leaf flushing season [5]. Results from a detailed experimental comparison of several descriptors show that our method presents high accuracy on identifying regions in the images belonging to a same plant species.

The remainder of this paper is organized as follows. Section II discusses the methodology adopted for acquiring time series. Section III presents our approach and shows how to apply it to characterize time series. Section IV 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 V.

II. Time Series Acquisition

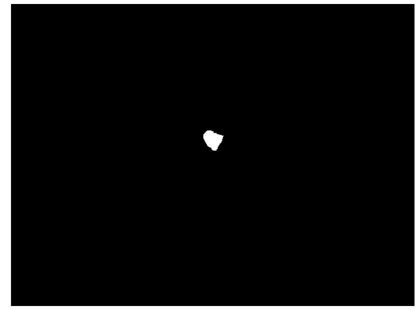
The near-remote phenological system was set up in a 18m tower in a Cerrado *sensu stricto*, a savanna-like vegetation located at Itirapina, São Paulo State, Brazil. A digital hemispherical lens camera (Mobotix Q24) was setup at the top of the phenology tower, attached in an iron arm facing northeast.

The 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, during the main leaf flushing season [5].

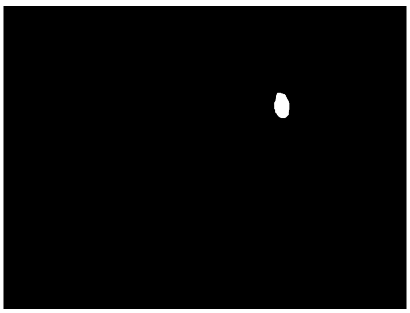


(a) *Aspidosperma tomentosum*

(b) *Caryocar brasiliensis*

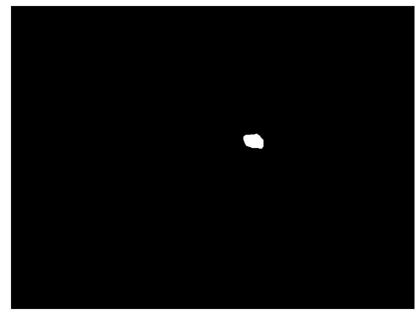


(c) *Myrcia guianesis*



(d) *Miconia rubiginosa*

(e) *Pouteria ramiflora*



(f) *Pouteria torta*

Figure 2. Regions of interest (ROIs) defined for the analysis of six plant species from the cerrado-savanna vegetation.

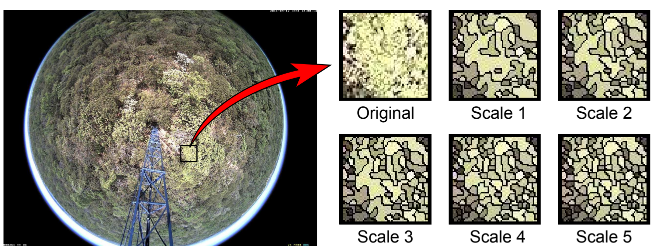


Figure 1. Sample image of the cerrado savanna recorded by the digital camera on October 15th, 2011; and the segmentation results for the selected scales in a subimage sample.

The image analysis was conducted by defining different regions of interest (ROI), as described by Richardson *et al.* [4], Richardson *et al.* [2], and Ahrends *et al.* [3]. 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 guianensis* (Figure 2(c)), (4) *Miconia rubiginosa* (Figure 2(d)), (5) *Pouteria ramiflora* (Figure 2(e)), and (6) *Pouteria torta* (Figure 2(f)).

III. Visual Rhythm-based Description

Visual rhythms [6] are an effective way to analyze temporal properties from video data. It consists in an abstraction of a video that encodes the temporal change of pixel values along a specific sampling line [7], as illustrated in Figure 3(a). A clear advantage of this approach is to reduce the storage space of the extracted features. Therefore, it also speeds up data processing.

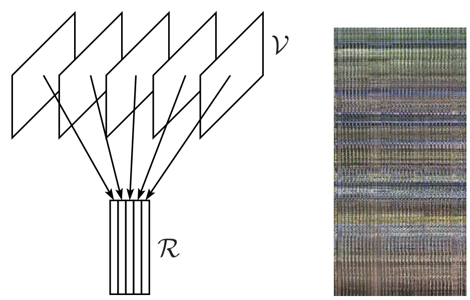


Figure 3. Visual rhythm: (a) simplification of a video content by mapping each frame into one column of an image; (b) a real example produced by sampling the central vertical line of the digital images.

Without loss of generality, a time series comprised of images taken by digital cameras at fixed time intervals can be viewed as a video of the vegetation. Thus, a visual rhythm can be used to simplify a time series into a single image, as illustrated in Figure 3(b). In this way, we can take the advantage of existing image descriptors to identify and characterize phenological changes.

The major problem with the current definition of visual rhythms is to have been designed for the pixel sampling of specific lines (e.g., diagonal, horizontal, and vertical) [8]. Here, we are interested in analyzing unshapely regions related to plant species that are identified by phenology experts (see Figure 2).

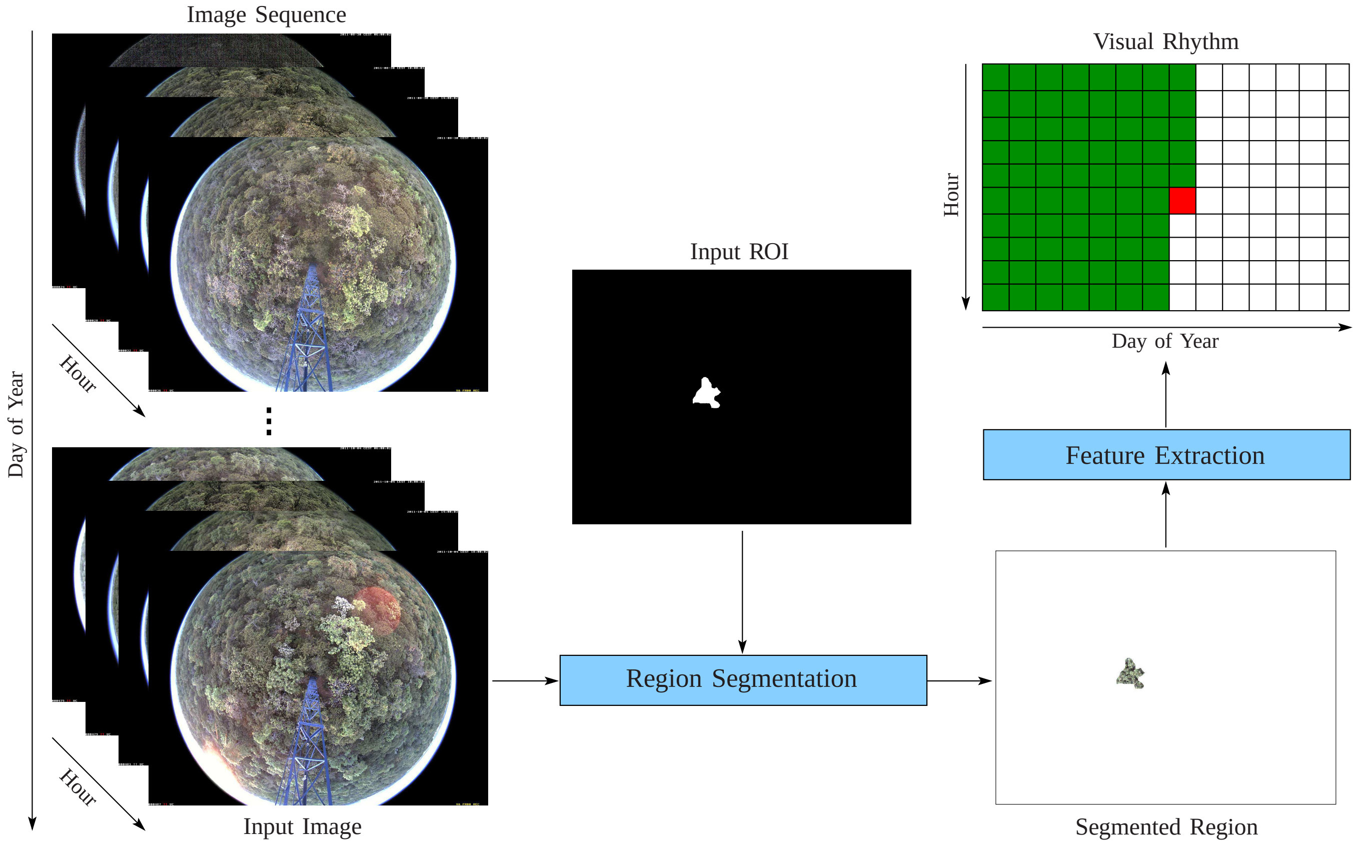


Figure 4. An overview of the proposed method.

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. For clarifying this process, look at Figure 4.

Let $\mathcal{S} = \{\mathcal{I}_{dh}\}, d \in [1, D], h \in [1, H]$ be an image sequence composed by $D \times H$ images $\mathcal{I}_{dh}$, with dimensions $\mathrm{W}_{\mathcal{S}} \times \mathrm{H}_{\mathcal{S}}$, 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 a ROI.

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 the color moments of this segmented region. Here, we adopt the first-order moment, which is the average color intensity, i.e.,

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

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.,

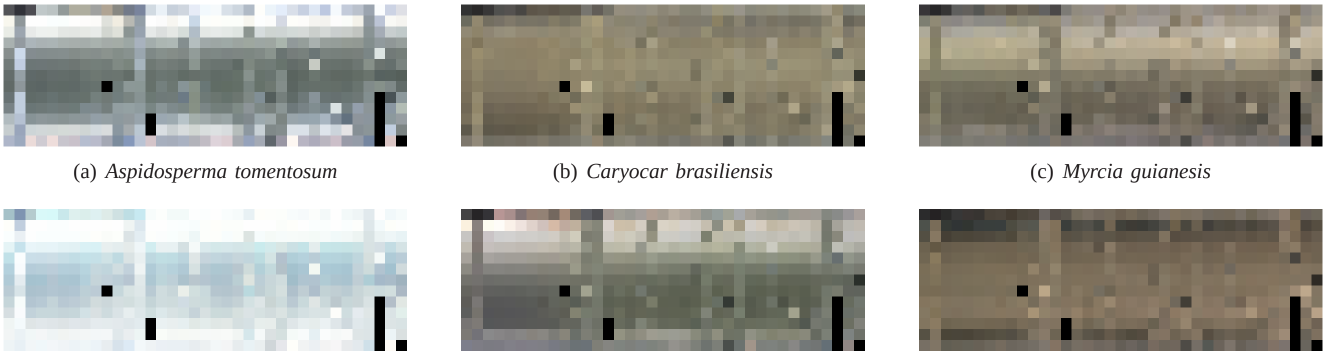
$$\mathcal{R}(d, h) = \mathcal{F}_{dh}, d \in [1, \mathrm{W}_{\mathcal{R}}], h \in [1, \mathrm{H}_{\mathcal{R}}],$$

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

IV. Experiments and Results

We carried out experiments to identify the plant species in the image. For describing time series encoded into a visual rhythm, we used six traditional and recently proposed image descriptors: ACC [9], CCV [10], BIC [11], and GCH [12], for encoding color information; GFD [13] and HWD [14], 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 [15].

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 the same species. Regions are defined by using the hierarchical segmentation based on the Guigues algorithm [16]. The image used to obtain the hierarchy of segmented regions was taken at noon on October 15th, 2011 (Figure 1). We



(a) *Aspidosperma tomentosum* (b) *Caryocar brasiliensis* (c) *Myrcia guianesis*

(d) *Miconia rubiginosa* (e) *Pouteria ramiflora* (f) *Pouteria torta*

Figure 5. Visual rhythms obtained for each ROI.

have selected 5 segmentation scales from the hierarchy to perform feature extraction, as illustrated in Figure 1. The finest scale is composed of $27,380$ regions and the coarsest scale contains $8,849$ regions.

The similarity between two regions 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 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 is overlapped by such a ROI. In our experiments, we have used only regions from the coarsest scale, as they have been shown the most effective ones to characterize plant species [17], [18].

For each ROI, we randomly selected 20 percent of its total number of regions 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.

We assess the effectiveness of each approach using the metrics of Precision and Recall. Precision is the ratio of the number of relevant regions retrieved to the total number of irrelevant and relevant regions retrieved. Recall is the ratio of the number of relevant regions retrieved to the total number of relevant regions in the database. Here, a given region is considered as relevant only if it belongs to the same ROI of a query region.

However, there is a trade-off between Precision and Recall. Greater Precision decreases Recall and greater Recall leads to decreased Precision. Therefore, 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 region; and Precision at 5 (P@5), which is the average precision after 5 regions are returned. These metrics combine both Precision and Recall into a single measure, which makes the comparison easier.

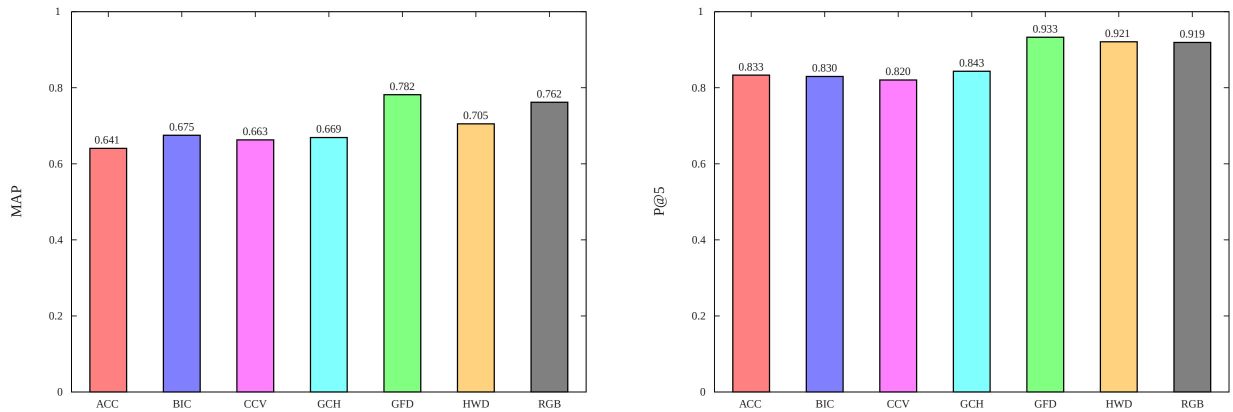
We compare the visual rhythm-based techniques with the method proposed by Richardson *et al.* [2], which is the most popular approach and widely used by the phenology community for characterizing phenological patterns of plant species. It consists in analyzing each region in terms of the variation of the relative (or normalized) brightness of the primary colors (RGB channels).

In Figure 6, we compare the visual rhythm-based techniques and the baseline method 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. Those results indicate that the performance of the different evaluated approaches is similar.

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.

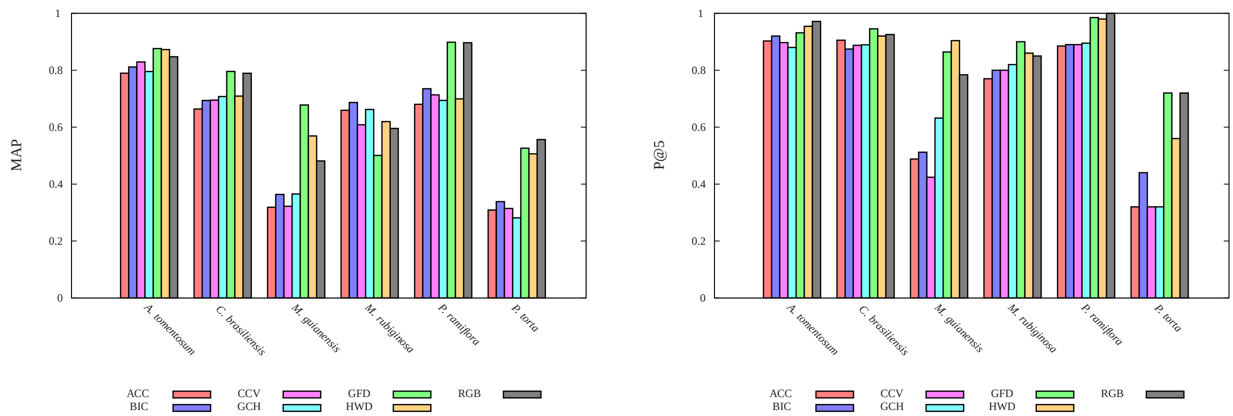
Table I presents the 99% confidence intervals of the differences between the baseline method and the visual rhythm-based techniques for the MAP and P@5 measures, respectively. Such analyses confirm that the visual rhythm-based techniques and the baseline method exhibit similar performance. Notice that the confidence intervals include zero and, hence, the differences between those approaches are not significant at that confidence level.

In Figure 7, we compare the individual scores obtained for each ROI in terms of the MAP and P@5 measures, respectively. It is interesting to note the differences in responsiveness of the different evaluated methods 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.



(a) MAP (b) P@5

Figure 6. MAP and P@5 scores obtained by each of the evaluated approaches.



(a) MAP (b) P@5

Figure 7. MAP and P@5 scores obtained for each ROI.

Table I
DIFFERENCES BETWEEN MAP AND P@5 SCORES OF THE DIFFERENT APPROACHES.

Method	**MAP**		**P@5**	
	min.	**max.**	**min.**	**max.**
RGB - ACC	-0.063	0.312	-0.084	0.410
RGB - BIC	-0.088	0.267	-0.047	0.319
RGB - CCV	-0.049	0.277	-0.097	0.441
RGB - GCH	-0.086	0.306	-0.090	0.362
RGB - GFD	-0.179	0.142	-0.088	0.056
RGB - HWD	-0.134	0.198	-0.135	0.159

Table II
COMPUTATIONAL COSTS AND SPACE REQUIREMENTS OF DIFFERENT APPROACHES.

Method	*Computational Cost*		*Space Requirements*
	Extraction	*Matching*	
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	$O(n \log n)$	$O(1)$	$O(1)$
RGB	$O(n)$	$\Omega(n)$	$\Omega(n)$

The key advantage of our technique is its computational efficiency. Table II 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 solution. This improvement makes our approach suitable for long-term collections of image data.

V. 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 2,700 images, taken from a tropical cerrado-savanna vegetation, including a high diversity of plant species. Results from the application of our method with several image descriptors show that it presents high accuracy and computational speed on identifying regions in the images belonging to a same species.

Future work includes the evaluation of other visual features for image retrieval (e.g., shape [19]). In addition, the proposed method can be augmented to consider temporal segmentation [20] and/or summarization methods [21]. Finally, we also plan to consider learning-to-rank methods (e.g., genetic programming [22]) for combining different descriptors.

Acknowledgment

This research was partially supported by Microsoft Research and Brazilian agencies FAPESP (grants 2010/52113-5, 2011/11171-5, 2012/18768-0, and 2012/16253-2), CNPq (grants 306243/2010-5, 484254/2012-0, and 306580/2012-8), and CAPES.

References

[1] M. D. Schwartz, *Phenology: An Integrative Environmental Science*. Academic Publishers, 2003.

[2] A. D. Richardson, J. P. Jenkins, B. H. Braswell, D. Y. Hollinger, S. V. Ollinger, and M. L. Smith, "Use of digital webcam images to track spring greep-up in a deciduous broadleaf forest," *Oecologia*, vol. 152, pp. 323–334, 2007.

[3] H. Ahrends, S. Etzold, W. Kutsch, R. Stoeckli, R. Bruegger, F. Jeanneret, H. Wanner, N. Buchmann, and W. Eugster, "Tree phenology and carbon dioxide fluxes: Use of digital photography for process-based interpretation at the ecosystem scale," *Climate Research*, vol. 39, pp. 261–274, 2009.

[4] A. D. Richardson, B. H. Braswell, D. Y. Hollinger, J. P. Jenkins, and S. V. Ollinger, "Near-surface remote sensing of spatial and temporal variation in canopy phenology," *Ecological Applications*, vol. 19, pp. 1417–1428, 2009.

[5] B. Alberton, J. Almeida, R. Henneken, R. da S. Torres, A. Menzel, and L. P. C. Morellato, "Near remote phenology: Applying digital images to monitor leaf phenology in a brazilian cerrado savanna," in *International Conference on Phenology (Phenology'12)*, 2012, p. 2.

[6] C. W. Ngo, T. C. Pong, and 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.

[7] J.-S. Lee and T. Ebrahimi, "Perceptual video compression: A survey," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 684–697, 2012.

[8] J. Almeida, J. A. dos Santos, B. Alberton, L. P. C. Morellato, and R. da S. Torres, "Visual rhythm-based time series analysis for phenology studies," in *IEEE International Conference on Image Processing (ICIP'13)*, 2013, pp. 1–5.

[9] J. Huang, R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997, pp. 762–768.

[10] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *ACM International Conference on Multimedia (ACM-MM'96)*, 1996, pp. 65–73.

[11] R. O. Stehling, M. A. Nascimento, and A. X. Falcão, "A compact and efficient image retrieval approach based on border/interior pixel classification," in *ACM International Conference on Information and Knowledge Management (CIKM'02)*, 2002, pp. 102–109.

[12] M. J. Swain and B. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[13] D. Zhang and G. Lu, "Shape-based image retrieval using generic fourier descriptor," *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825–848, 2002.

[14] C. E. Jacobs, A. Finkelstein, and D. Salesin, "Fast multiresolution image querying," in *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'95)*, 1995, pp. 277–286.

[15] O. A. B. Penatti, E. Valle, and R. da S. Torres, "Comparative study of global color and texture descriptors for web image retrieval," *Journal of Visual Communication and Image Representation*, vol. 23, no. 2, pp. 359–380, 2012.

[16] L. Guigues, J. Cocquerez, and H. Le Men, "Scale-sets image analysis," *International Journal of Computer Vision*, vol. 68, pp. 289–317, 2006.

[17] J. Almeida, J. A. dos Santos, B. Alberton, R. da S. Torres, and L. P. C. Morellato, "Remote phenology: Applying machine learning to detect phenological patterns in a cerrado savanna," in *IEEE International Conference on eScience (eScience'12)*, 2012, pp. 1–8.

[18] J. Almeida, J. A. dos Santos, B. Alberton, R. da S. Torres, and L. P. C. Morellato, "Applying machine learning to detect remote phenology patterns in cerrado savanna trees," *Ecological Informatics*, 2013. DOI: 10.1016/j.ecoinf.2013.06.011

[19] R. da S. Torres, M. Hasegawa, S. Tabbone, J. Almeida, J. A. dos Santos, B. Alberton, and L. P. C. Morellato, "Shape-based time series analysis for remote phenology studies," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS'13)*, 2013, pp. 1–4.

[20] J. Almeida, N. J. Leite, and R. da S. Torres, "Rapid cut detection on compressed video," in *Iberoamerican Congress on Pattern Recognition (CIARP'11)*, 2011, pp. 71–78.

[21] J. Almeida, N. J. Leite, and R. da S. Torres, "Online video summarization on compressed domain," *Journal of Visual Communication and Image Representation*, vol. 24, no. 6, pp. 729–738, 2013.

[22] F. S. P. Andrade, J. Almeida, H. Pedrini, and R. da S. Torres, "Fusion of local and global descriptors for content-based image and video retrieval," in *Iberoamerican Congress on Pattern Recognition (CIARP'12)*, 2012, pp. 845–853.