On the Significance of Leaf Sides in Automatic Leaf-based Plant Species Identification

Carranza-Rojas, Jose (jcarranza@itcr.ac.cr); Mata-Montero, Erick (emata@itcr.ac.cr)
PARMA Group, School of Computing
Costa Rica Institute of Technology
Cartago, Costa Rica

***Abstract*—Because the front side of a leaf and the underside are functionally very different – the former captures sunlight to produce photosynthesis and the latter absorbs carbon dioxide and releases oxygen and vapor – they typically have different visual features. In this paper we study the significance of leaf sides in visual recognition systems for automatic plant species identification. We measure the accuracy of species identifications with a dataset of 63 species of trees from Costa Rica that includes pictures of both, front sides and undersides of tree leaves. The dataset is used as a global dataset and is also partitioned as two datasets: one of front side pictures and one of underside pictures. Training and testing of different algorithms is performed and their accuracies computed for the group of species and for each individual species. For the tested dataset, leaf side is a significant factor for automatic plant species identification. On the average, and for most cases, underside pictures lead to more accurate identifications.**

***Index Terms*—Biodiversity Informatics, Computer Vision, Image Processing, Plant Identification**

I. Introduction

Automatic identification of organisms has not only been a dream among systematists for centuries [1], but also a current need to understand, sustainably use, and save biodiversity. Several automatic and semi-automatic approaches have been used in the past. For example, dichotomous keys, multi-access keys, morphometrics, DNA barcoding, and image-based identification, among others [2]. A number of computer vision and machine learning techniques use leaf images to identify plant species [3], [4], [5], [6]. It is usually assumed that a user takes a picture P of the front side of a specimen's leaf, which is then used by an algorithm or model M to establish a ranking of the best k candidate species for P, for some "small" value of k (*e.g.*, $1 \leq k \leq 5$). Supervised training techniques are typically used to train model M with leaf image datasets that often include pictures of the front side and the underside of leaves of specimens that have been previously identified. Because research that aims to identify plants based on leaf images alone tries to get the best out of the leaf visible features, it is important to consider as many leaf discriminant factors as possible. Nevertheless, to our knowledge, previous research on automatic visual plant species identification based on leaf images use front side and underside pictures indiscriminately.

From a functional point of view, the front side and the underside of a leaf are in charge of two different critical tasks. The front side surface gathers energy from sunlight while apertures (stomata) on the the cooler shady underside bring in carbon dioxide and release oxygen and vapor. As a result, the front side and the underside of a leaf tend to have a different appearance. The front side tends to be glossy and has more vivid colors while the underside may have more trichomes (hairs) to keep the surface cool, could be duller, and veins could be more visible.

In this paper we study the significance of leaf sides in automatic visual plant species identification based on leaf images. Our hypothesis is that an automated leaf image-based plant identification system benefits from having the training dataset split into two subsets: one that comprises front side pictures only and one that consists of back side pictures only, which leads to two different plant identification models that we call $Model_F$ and $Model_B$, respectively. We postulate that $Model_F$ and $Model_B$ would be more accurate than M when the image P corresponds to the front and back side of a leaf, respectively. As a pragmatic consequence, when a user provides a picture P for an automatic identification, they should indicate the leaf side so that either $Model_F$ or $Model_B$ is used. However, even if the hypothesis does not hold true, it may still be significant if P is a front side or an underside picture when a general model M is used. Therefore, our experiments also address this issue.

Because of the rich diversity of plant and even tree species in Costa Rica, we realize that the results of this research are affected by the subset of species used. Some species may have front sides of leaves that are very distinctive while others may have undersides that are more discriminating. The accuracy achieved globally for the dataset used in this research may not reflect the importance of leaf sides in automatic plant species identification for individual species. Thus, our experiments also assess, for each of the 63 species in the dataset, the accuracy of models $Model_F$, $Model_B$, and M when picture P corresponds to either the front side or the underside of a leaf.

The rest of this manuscript is organized as follows: Section II summarizes relevant related work. Section III and Section IV cover methodological aspects and experiment design, respectively. Section V describes the results obtained. Section VI presents the conclusions and, finally, Section VII summarizes future work.

II. Related Work

Previous research on leaf image-based identification of plant species has been reported in [3], [7], [4], and [6]. LeafSnap [3] uses a curvature model and similarity search using k-Nearest Neighbors (kNN) with an image dataset of North American trees that comprises 184 species. Herdiyeni et al. [7] use Local Binary Patterns (LBP) features to classify medicinal and house plants from Indonesia based also on leaf images, for a total of 30 species. Nguyen et al. [4] use Speeded Up Robust Features (SURF) to develop an Android application for mobile plant species recognition based on leaf images of 32 species. Finally, [6] extends work in [3] along two lines. First, LeafSnap's underlying algorithms are applied to a set of 66 tree species from Costa Rica. Secondly, texture is used as an additional criterion to measure the level of improvement achieved in the automatic identification of Costa Rica tree species. None of these studies address the issue of significance of leaf sides in automatic leaf-based plant species identification.

III. Methodology

We used the same approach as [3] and [6] to classify leaves into species of plants. The dataset of images is a subset of the one used in [6]. However, a first step was to add metadata indicating the leaf side of each image. Then, leaf segmentation was carried out by using Expectation-Maximization (EM). After that, two leaf features were extracted, namely (visual) texture and curvature. Then, classification was done using kNN with $3 \leq k \leq 5$ and using histogram intersection as distance metric. Finally, the accuracy achieved by the classifier was calculated.

Th following subsections provide more details about the image data used, the segmentation approach, and the algorithms used for feature extraction.

A. Image Data

The dataset created by Carranza et al. [6] was used almost in its entirety; it includes images of 63 species of ranomly picked trees from Costa Rica's central plateau region. Labels were added to logically separate leaf front side from leaf underside images, which allowed us to experiment with each image dataset separately or in combined form. Following the notation presented in Section I, $Train_F$ is the subset that comprises all 998 front side leaf images, $Train_B$ is the subset that contains all 991 back side leaf images, and finally $Train_C$ is the complete dataset with all 1989 images combined.

B. Segmentation

For segmentation we used the Hue Saturation Value (HSV) color space to cluster pixels into two clusters using EM. However, we discarded the Hue channel since it often contains too much noise. One cluster corresponds to the *lamina* (leaf blade) and the other one to the background.

C. Features

We extracted two different feature sets, one for texture and one for margin or curvature. The following subsections explain briefly both algorithms.

1) Local Binary Patterns Uniform (LBPU): As mentioned in Section I, the front and back side of a leaf typically display different textures. LBPU is a feature extraction algorithm that is rotation invariant and has been proved to be excellent for texture pattern matching [8]. The following three different variations of LBPU are used:

- Radius of 1 pixel, 8 pixels of sampling. We call it R1P8.
- Radius of 3 pixels, 16 pixels of sampling. We call it R3P16.
- The concatenation of the previous 2 into a single histogram. We call it R1P8P3P16.

A LBPU descriptor is applied to each pixel c in the image and its circular neighborhood $Neighborhood(c)$ that has radius R and P pixels. For each pixel p in $Neighborhood(c)$, p has a gray level $gray(p)$. A boolean threshold function is applied to the difference of gray value between each pixel p from the neighborhood of c and the central pixel c, to form a binary number of length P. To achieve rotation invariance, right shifts are applied to the binary number and then the minimum number is selected.

2) Histogram of Curvature over Scale (HCoS): This descriptor was developed by the authors in [3]. First, a disk of radius $1 \leq r \leq 25$ is defined at every contour pixel of the leaf. Then, two different histograms are created by measuring the pixel area of the intersection of the disk with the leaf and the length of the arc defined by the intersection of the circumference of the disk and the leaf. This is calculated for all 25 values of radius r and concatenated together into a single histogram called HCoS.

This curvature descriptor, as well as the LBPU variants described, are levels of the factor named *Algorithm*, as explained in Section IV, which describes the experiments. Even though the curvature of the front side and the back side of a leaf are mirror images of each other, this feature was included in the analysis just to determine if it is relevant or should be discarded in future analysis.

D. Trained Models and Classification

Classification was carried out by using kNN with $3 \leq k \leq 5$, which, from a user point of view, is a reasonable range of "small" values of k. To calculate the distance between histograms, we used *histogram intersection* as described in [3].

Three algorithms or trained models were defined. $Model_F$ is the model trained with only front side images, $Model_B$ is the model trained with back side images, and $Model_C$ is the model trained with with the complete image dataset.

We calculated the accuracy of the different models. Let E be an identification experiment that consists of a model M, a set S that contains n images of leaves of n (not necessarily different) unknown tree species to be identified, and an integer value k, $k \geq 1$. We define $hit(M,k,x)$ as a boolean function that indicates if model M generates a ranking in which one of the top k candidate species is a correct identification of sample x. Equation 1 formally defines $Accuracy(M,S,k)$.

$$Accuracy(M,S,k) = \sum\nolimits_{x \in S} \frac{hit(M,k,x)}{n} \qquad (1)$$

Table I: Levels for *Training+Model* factor

Level	Description
$Test_B + Model_B$	Model tested with back side images and trained with back side images
$Test_B + Model_C$	Model tested with back side images and trained with complete dataset
$Test_C + Model_C$	Model tested with complete dataset and trained with complete dataset
$Test_F + Model_C$	Model tested with front side images and trained with complete dataset
$Test_F + Model_F$	Model tested with front side images and trained with front side images
$Test_B + Model_F$	Model tested with back side images and trained with front side images
$Test_F + Model_B$	Model tested with front side images and trained with back side images

IV. Experiments

We ran the classifier over the two datasets $Train_F$ and $Train_B$ to get the accuracy related to front side and back side leaf images. We also ran it for the complete, combined dataset $Train_C$. Additionally, we used a General Linear Model (GLM) to test if the leaf side was actually a significant factor during classification, with a confidence level of 95%. The three factors used in the GLM are: *Algorithm*, *k*, and *Training+Model*. Factor *Training+Model* represents the combination of a particular trained model, and the dataset used for testing. Table I shows the seven levels related to this factor. We used $3 \leq k \leq 5$ only, since those values would be suitable for a species ranking for a mobile app or similar.

After finding if the *Training+Model* factor was significant, a Tukey test was run to assess if the difference between levels for the *Training+Model* was statistically significative, with a confidence level of 95%. This would tell us how relevant leaf side are across the tests.

We ran this globally for all species, but we also ran the GLM for each species separately. This would tell us the role of the leaf side for each species.

V. Results

A. Global significance of leaf side

Table II summarizes the obtained P-Values for each of the three factors and their interactions. All datasets and all feature extraction algorithms (texture and curvature) were used, for $3 \leq k \leq 5$. The most important factor to our experiments is *Training+Model* which obtained a p-Value of 0%, suggesting leaf side significance on both training and testing. Notice also that *Training+Model* is significant together with *Algorithm*, which means that some feature extraction algorithms

Table II: Global GLM results at a 95% confidence. R-sq = 99.96%

Source	P-Value
k	0.000
Algorithm	0.000
Training+Model	0.000
*k*Algorithm*	0.000
*k*Training+Model*	0.022
*Algorithm*Training+Model*	0.000

Table III: Tukey Pairwise Comparisons at a 95% confidence, for factor *Training+Model*

Training+Model	Accuracy Mean	Grouping
$Test_B + Model_B$	0.80	A
$Test_B + Model_C$	0.79	B
$Test_C + Model_C$	0.76	C
$Test_F + Model_C$	0.74	D
$Test_F + Model_F$	0.73	E
$Test_B + Model_F$	0.37	F
$Test_F + Model_B$	0.31	G

may work better or worse depending on the leaf side images used for training and testing.

Table III summarizes the results of running the Tukey test for *Training+Model*. The mean is computed over the accuracy obtained for all feature extraction algorithms and $3 \leq k \leq 5$, for each *Training+Model* level. Group A, which uses $Test_B$ (tested with back images) and $Model_B$ (trained with back images), achieves the best average accuracy. Group B, which is closely related to Group A, but slightly inferior, also uses $Test_B$ for testing, but the combined $Model_C$ for training. This suggests that globally, the identification of back side images P is better than when P is a front side image (except if the model used is $Model_F$).

It is interesting to note that when P is a front side image, the combined $Model_C$ is slightly better than using a more specialized $Model_F$.

Additionally, it is worth noting that, consistent with intuition, testing with $Test_B$ but training with $Model_F$, and vice-versa, is not a good idea.

For the sake of completeness, we also ran tests for the curvature algorithm alone. Not surprisingly, the worst cases are also $Test_B - Model_F$ and $Test_F - Model_B$, but with a higher accuracy of 67% in both cases. Compared to the Tukey test that contains both curvature and texture in Table III, which was as low as 37%, this 67% is much better. This shows that internal texture patterns differ between leaf sides for classification and that curvature does not suffer as much when one side or the other of the leaf is used.

B. *Significance of leaf side per species*

Table IV shows the results of the GLM applied to each of the 63 species. For 39 species (61.9%) the best accuracy is obtained when back side images P are used. For 16 species (23.8%) the highest accuracy is obtained when P is a front side image. Finally, for 9 species (14.2%) there is no clear winner. This means that a large group of species are better classified when P is a back side image, but there is also another group of species that have better results when P is a front side image. In the context of a software tool, this individualized analysis is important for use cases in which the user is trying to determine if image P corresponds to a given species. For example, if we want to determine if P is an image of *Brosimum alicastrum*, we may get better accuracy in the automatic identification if P is a back side image and the model was trained with a dataset of back side images (although a general $Model_C$ would not be too bad). However, if we want to determine if P is an image of *Quercus insignis*, we may get better accuracy if P is a front side image and the model was trained with a dataset of front side images (although a general $Model_C$ would not be too bad either).

A visual example of the difference between leaf sides is shown in Figure 1 for species *Brosimum alicastrum*. Visually, both images show how images of the leaf side of a single individual differ. For this particular species, the accuracy ranges from 91% for the back side subset, down to 74% for the front side subset, according to Table IV. For the combined or complete dataset the obtained accuracy is 80%.

VI. Conclusions

For the tested dataset, leaf side is a significant factor for automatic plant species identification. On the average, and for most cases, underside pictures lead to more accurate identifications. For most species (61.9%), classification is better if the sample P to be identified is a back side leaf image; in a smaller number of cases (23.8%) a front side image P gives better results.

In agreement with intuition, the worst accuracy is obtained when the model is trained with back side images and tested with front side images and vice-versa.

However, it should be noticed that the above conclusions are due to the differences in texture displayed in the back and front sides of leaves. Because the curvature of the front side and the back side of a

Table IV: Accuracy mean per species for the *Training+Model* factor. Highlighted values belong to the most significant group according to the Tukey tests

Species	$Test_B + Model_B$	$Test_B + Model_C$	$Test_F + Model_F$	$Test_F + Model_C$	$Test_C + Model_C$
Acnistus arborescens	0.75	***0.80***	0.69	0.77	0.79
Aegiphila valerioi	0.63	***0.64***	0.47	0.55	0.6
Annona mucosa	***0.64***	***0.60***	0.47	0.5	0.55
Ardisia revoluta	***0.75***	***0.75***	0.55	0.6	0.68
Blakea maurofernandeziana	0.94	***0.98***	0.81	0.81	0.9
Brosimum alicastrum	***0.91***	***0.90***	0.74	0.69	0.8
Calophyllum brasiliense	***0.87***	***0.85***	0.68	0.67	0.76
Calycophyllum candidissimum	***0.83***	***0.83***	0.71	0.68	0.75
Cestrum tomentosum	***0.80***	***0.80***	0.75	0.77	0.78
Citharexylum donnell-smithii	0.77	***0.82***	0.68	0.67	0.75
Clethra costaricensis	***0.75***	0.65	0.64	0.7	0.67
Clusia croatii	0***.95***	0.89	0.81	0.76	0.83
Coccoloba floribunda	***0.64***	***0.63***	0.49	0.52	0.58
Cordia eriostigma	***0.68***	***0.67***	0.55	0.55	0.61
Croton draco	***0.98***	***0.87***	0.74	0.77	0.82
Croton niveus	***0.85***	***0.84***	0.79	0.79	0.81
Dalbergia retusa	***0.85***	***0.80***	0.65	0.65	0.73
Ficus cotinifolia	***0.91***	0.87	0.85	0.87	0.87
Guazuma ulmifolia	***0.93***	***0.93***	0.85	0.83	0.88
Hyeronima alchorneoides	***0.8***	***0.81***	0.71	0.71	0.76
Manilkara chicle	***0.86***	***0.85***	0.75	0.78	0.81
Ocotea sinuata	0.84	***0.86***	0.83	0.82	0.84
Persea americana	***0.78***	***0.76***	0.64	0.62	0.69
Pimenta dioica	***0.9***	***0.9***	0.58	0.58	0.74
Platymiscium pinnatum	***0.70***	***0.71***	0.6	0.57	0.64
Posoqueria latifolia	***0.72***	***0.66***	0.48	0.5	0.58
Quercus corrugata	***0.97***	0.95	0.85	0.88	0.91
Robinsonella lindeniana var. divergens	***1***	***1***	0.93	0.94	0.97
Sapium glandulosum	***0.82***	***0.80***	0.73	0.73	0.76
Sideroxylon capiri	***0.80***	0.76	0.54	0.52	0.65
Simarouba glauca	***0.97***	***0.95***	0.65	0.63	0.79
Swietenia macrophylla	***0.73***	***0.71***	0.67	0.65	0.68
Tabebuia ochracea	***0.79***	***0.82***	0.67	0.65	0.73
Tabebuia rosea	***0.83***	***0.83***	0.5	0.5	0.66
Tabernaemontana litoralis	***0.77***	***0.77***	0.67	0.67	0.72
Terminalia amazonia	0.86	***0.89***	0.84	0.83	0.86
Terminalia oblonga	***0.81***	***0.81***	0.58	0.64	0.72
Trichilia havanensis	***0.77***	0.67	0.68	0.7	0.68
Vernonia patens	***0.94***	0.93	0.87	0.84	0.89
Anacardium excelsum	0.72	0.75	0.75	***0.80***	0.78
Bauhinia purpurea	0.83	0.84	0.85	***0.87***	0.85
Colubrina spinosa	0.69	0.7	***0.89***	***0.88***	0.79
Dendropanax arboreus	0.53	0.47	***0.56***	***0.57***	0.52
Dipteryx panamensis	0.69	0.68	0.73	***0.76***	0.72
Eugenia hiraeifolia	0.83	0.79	0.87	***0.95***	0.87
Genipa americana	0.56	0.59	0.69	***0.80***	0.7
Heliocarpus appendiculatus	0.84	0.86	0.89	***0.97***	0.92
Hymenaea courbaril	0.61	0.62	***0.82***	***0.82***	0.72
Pachira quinata	0.79	0.76	***0.8***	0.78	0.77
Platymiscium parviflorum	0.65	0.61	0.62	***0.7***	0.65
Quercus insignis	0.8	0.82	***0.94***	***0.93***	0.87
Samanea saman	0.8	0.78	***0.9***	0.86	0.82
Stemmadenia donnell-smithii	0.35	0.38	***0.75***	***0.73***	0.57
Urera caracasana	0.94	0.94	***0.97***	0.85	0.94
Astronium graveolens	***0.84***	***0.85***	0.77	***0.83***	***0.84***
Erythrina poeppigiana	0.65	***0.7***	0.68	***0.71***	***0.7***
Hura crepitans	***0.83***	***0.84***	0.79	***0.84***	***0.84***
Psidium guajava	***0.75***	0.66	***0.73***	0.65	0.66
Solanum rovirosanum	0.68	***0.70***	0.65	***0.69***	***0.69***
Tabebuia impetiginosa	***0.82***	0.8	0.84	***0.83***	0.82
Bauhinia ungulata	0.79	0.79	0.81	0.79	0.79
Cedrela odorata	0.89	0.87	0.87	0.91	0.89
Muntingia calabura	0.95	94	0.93	0.93	0.94

(a) Front side image of a *Brosimum alicastrum* sample.

(b) Back side image of a *Brosimum alicastrum* sample.

Figure 1: Difference between sides of the same leaf specimen of *Brosimum alicastrum*.

leaf are mirror images of each other, this feature is not sensibly affected by which side of the leaves are used. Thus, tools based on curvature analysis alone such as LeafSnap [3] would not be affected by the indiscriminate use of leaf front and back side images.

VII. Future Work

Other feature extraction algorithms such as point of interest should undergo a similar type of analysis. Additionally, it is important to understand if different leaf regions such as the apex, base, or petiole have significant features. Understanding this could also help in classifying species even when the leaf is partially damaged or only a portion of it is available. Because the results of this type of research are affected by the subset of species used, it is very important to create a national level or global level leaf image dataset with as many species as possible. As more leaf image data becomes available, analysis by geographic regions, higher level taxa, special interest taxa (*e.g.*, endangered species and species of economic interest), and other groups would be extremely useful for biodiversity conservation. Also, as more leaf data are gathered and made available, approaches such as ConvNets [9] would be more feasible for identification even with complex backgrounds.

Acknowledgement

Thanks to the National Museum of Costa Rica and to Jose David Sánchez for their help in setting up the $Train_F$ and $Train_B$ image datasets. Additional thanks to Nelson Zamora and the National Biodiversity Institute of Costa Rica (INBio) for collecting and identifying the leaf samples that are included in the complete dataset $Train_C$. This image dataset is available under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license upon request to the authors.

References

[1] N. MacLeod, *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*. Systematics Association Special Volumes, Taylor & Francis, 2007.

[2] E. Mata-Montero and J. Carranza-Rojas, "Automated plant species identification: Challenges and opportunities," in *IFIP World Information Technology Forum*, pp. 26–36, Springer, 2016.

[3] N. Kumar, P. Belhumeur, A. Biswas, D. Jacobs, W. Kress, I. Lopez, and J. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *Computer Vision – ECCV 2012* (A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, eds.), Lecture Notes in Computer Science, pp. 502–516, Springer Berlin Heidelberg, 2012.

[4] Q. Nguyen, T. Le, and N. Pham, "Leaf based plant identification system for android using surf features in combination with bag of words model and supervised learning," in *International Conference on Advanced Technologies for Communications (ATC)*, October 2013.

[5] S. Wu, F. Bao, E. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pp. 11–16, Dec 2007.

[6] J. Carranza-Rojas and E. Mata-Montero, "Combining Leaf Shape and Texture for Costa Rican Plant Species Identification," *CLEI Electronic Journal*, vol. 19, pp. 7 – 7, 04 2016.

[7] Y. Herdiyeni and I. Kusmana, "Fusion of local binary patterns features for tropical medicinal plants identification," in *Advanced Computer Science and Information Systems (ICACSIS), 2013 International Conference on*, pp. 353–357, Sept 2013.

[8] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 971–987, Jul 2002.

[9] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*, ICDAR '03, (Washington, DC, USA), pp. 958–, IEEE Computer Society, 2003.