Automatic Classification of Lobed Simple and Unlobed Simple Leaves for Plant Identification

Juan Carlos Flores-Bastida, Asdrúbal López-Chau, Rafael Rojas-Hernández, Valentin Trujillo-Mora

Universidad Autónoma del Estado de México, Zumpango, Estado de México, México

juancarlosfloresbastida@gmail.com

Abstract. Commonly, classic plant identification methods use dichotomous or multi-access keys that compare characteristics of the leaves, asking if they are lobed, unlobed, simple or compound, among others leaf features. However, in the literature very little attention has been paid to make an automatic distinction of leaves using such features. In this paper, we contribute to fill this gap. We propose a novel method to differentiate between types of leaves. The proposal is invariant to rotation and also to scaling. In order to show the effectiveness of the proposal, we tested it with more than 1,900 images of leaves which are publicly available on the Internet, achieving correct identification rates greater than 86%.

Keywords: compound leaf, leaf feature, lobed simple leaf, unlobed simple leaf.

1 Introduction

Most of plant identification methods use leaves. There are some good reasons for this, for example, plants have leaves almost all year [6], the number of them is usually high [1]; depth can be ignored in images of leaves, unlike flowers or other organs of plants; and leaves are different between plants [10].

One of the first phases in automatic plant identification is to extract a set of features from leaves (after some pre-processing on images). In spite of many descriptors based on the shape, color, texture and veins of leaves [18,3,4,5,11,12,16] have been proposed in last decade, little attention has been paid to develop new methods to make an automatic distinction between the different types of leaves, such as unlobed, lobed, simple and compound. This categorization of leaves is important, and it is usually used in dichotomous keys for classic plant identification, see for example [14,2,9,17,15].

We propose a new method to differentiate between simple lobed and simple unlobed leaves. Our method counts the number of changes of color of lines that are traced over the image of leaves. The proposal is invariant to rotation and also to scaling of images.

Juan Carlos Flores-Bastida, Asdrúbal López-Chau, Rafael Rojas-Hernández, Valentin Trujillo-Mora

The rest of the paper is organized as follows. Subsection 2.1 explains the types of leaves, and also explains the difference between lobed simple leaves and unlobed simple ones, 2.1 describes some basic types of plant leaves. Section 2.2 shows those materials used in this paper. We present our proposals in Section 3, then Section 4 shows experiments and results. Finally, last section of this paper presents conclusions and future works.

2 Preliminaries

2.1 Types of Plant Leaves

Categorizing leaves (and plants) is not a trivial task. Leaves can be classified according to their blade (simple or compound), edge (smooth, dentate, etc.), petiole (petiolated or sessile), shape of blade, etc. Among these categories, simple, compound, unlobed and lobed are very common in dichotomous keys. For the former category, the leaf blade is a single, continuous unit. For the second category, the blade is divided into two or more leaflets arising from the petiole. Simple leaves can be unlobed or lobed. For unlobed leaves, the blade is completely undivided. Lobed leaves have projections off the midrib with individual inside veins.

In some cases, such as the leaves shown in Figure 1, the definitions above can be directly applied to categorize a leaf. However, for other leaves, such as the ones shown in Figure 2, it could be a bit more complicated to categorize them.

Fig. 1. Example of a simple leave (left) and a compound leaf (right).

Fig. 2. Some types of lobed simple leaves.

Currently, there is not a single leaf descriptor that allows to identify all types of leaves perfectly. Instead, each leaf descriptor focuses on extracting one characteristic. The methods proposed in this paper identify to which of the following two groups belongs a plant leaf: a) unlobed simple leaves with smooth margins, and b) other types (compound, lobed, palmate, etc.). This information can be encoded as a binary leaf feature in plant identification.

2.2 Materials

Flavia data set is one of most widely used data sets for testing plant identification systems. It is publicly available at `http://flavia.sourceforge.net`. Flavia set contains 1,907 color images of 32 different species of plants. These images have a dimension of $1,600 \times 1,200$ pixels.

Figures 3 and 4 show the scientific name of plants, the class, and an example of a leaf for each one of the 32 species of plants in Flavia set.

Name	Class	Example	Name	Class	Example	Name	Class	Example
Phyllostachys edulis (Carr.) Houz. **Attribute: FALSE**	C1		Kalopanax septemlobus (Thunb. ex A.Murr.) Koidz. **Attribute: TRUE**	C8		Viburnum awabuki K.Koch **Attribute: FALSE**	C16	
Aesculus chinensis **Attribute: FALSE**	C2		Cinnamomum japonicum Sieb. **FALSE**	C9		Osmanthus fragrans Lour. **Attribute:FALSE**	C17	
Berberis anhweiensis Ahrendt **Attribute:FALSE**	C3		Koelreuteria paniculata Laxm. **Attribute:FALSE**	C10		Cedrus deodara (Roxb.) G. Don **Attribute:FALSE**	C18	
Cercis chinensis **Attribute:FALSE**	C4		Ilex macrocarpa Oliv. **Attribute:FALSE**	C11		Ginkgo biloba L. **Attribute:FALSE**	C19	
Indigofera tinctoria L. **Attribute:FALSE**	C5		Pittosporum tobira (Thunb.) Ait. f. **Attribute:FALSE**	C12		Lagerstroemia indica (L.) Pers. **Attribute:FALSE**	C20	
Acer Palmatum **Attribute:TRUE**	C6		Chimonanthus praecox L. **Attribute:FALSE**	C14		Nerium oleander L. **Attribute:FALSE**	C21	
Phoebe nanmu (Oliv.) Gamble **Attribute:FALSE**	C7		Cinnamomum camphora (L.) J. Presl **Attribute:FALSE**	C15		Podocarpus macrophyllus (Thunb.) Sweet **Attribute:FALSE**	C22	

Fig. 3. Species of plants in Flavia set, first part.

Juan Carlos Flores-Bastida, Asdrúbal López-Chau, Rafael Rojas-Hernández, Valentin Trujillo-Mora

Name	Class	Example	Name	Class	Example
Prunus serrulata Lindl. var. lannesiana auct. **Attribute:FALSE**	C23		Magnolia grandiflora L. **Attribute:FALSE**	C30	
Ligustrum lucidum Ait. f. **Attribute:FALSE**	C24		Populus ×canadensis Moench **Attribute:FALSE**	C31	
Tonna sinensis M. Roem. **Attribute:FALSE**	C25		Liriodendron chinense (Hemsl.) Sarg. **Attribute:TRUE**	C32	
Prunus persica (L.) Batsch **Attribute:FALSE**	C26		Citrus reticulata Blanco **Attribute:FALSE**	C33	
Manglietia fordiana Oliv. **Attribute:FALSE**	C27				
Acer buergerianum Miq. **Attribute:TRUE**	C28				
Mahonia bealei (Fortune) Carr. **Attribute:FALSE**	C29				

Fig. 4. Species of plants in Flavia set, second part.

3 Proposed Method to Distinguish Lobed Simple from Unlobed Simple Leaves

In this subsection, we introduce a novel method to detect whether a leaf is lobed from simple.

The first step is to align a binary image of a leaf. Instead of rotating L manually as in other works, we obtain automatically two new reference axes for L. The first reference axis is the line that joins the two most separate pixels in the leaf (diameter). The second reference axis is an orthogonal line to the first axis.

Algorithm 1 shows the procedure to compute the references axes for a leaf L. Figure 5 shows two examples of the references axes computed with Algorithm 1.

The second step in our method, is to drawn a set of equidistant lines over the body of L. A number of these lines are parallel to $\mathcal{L}_1$, and the same number of lines are parallel to $\mathcal{L}_2$.

Changes (from white to black or vice-versa) along each line are counted and stored in an array whose length is equal to the number of lines drawn. Then, a threshold is used to determine if the leaf is lobed. Algorithm 2 summarizes our method.

Figure 6 shows two examples of the lines obtained with Algorithm 2

Algorithm 1: Reference axes for a leaf.

   Input: A binary image of leaf L
   Output: New reference axes ($\mathcal{L}_1$ and $\mathcal{L}_2$) for L

1 Get de contour of the binary image
2 Compute $\underset{p_i,p_j}{\operatorname{argmax}}\, d(p_i, p_j)$ such that p_i, p_j belong to contour.
3 Let be $\mathcal{L}_1$ a line which satisfies equation (1):

$$y = \frac{\Delta_y}{\Delta_x}x + \left(p_{iy} - \frac{\Delta_y}{\Delta_x}p_{ix} \right) \tag{1}$$

$$\Delta_x = p_{ix} - p_{jx}$$

$$\Delta_y = p_{iy} - p_{jy}$$

   Compute $\underset{p_k,\mathcal{L}_1}{\operatorname{argmax}}\, d(p_k, \mathcal{L}_1)$ such that p_k belong to contour.
4 Let be $\mathcal{L}_2$ a line which satisfies equation (2):

$$y = \frac{\Delta_x}{\Delta_y}\left(x - p_{kx} \right) + p_{ky} \tag{2}$$

5 **return** $\mathcal{L}_1$ and $\mathcal{L}_2$ as the new reference axes.

Fig. 5. Reference axes computed for a simple leaf (left), and a lobed leaf (right).
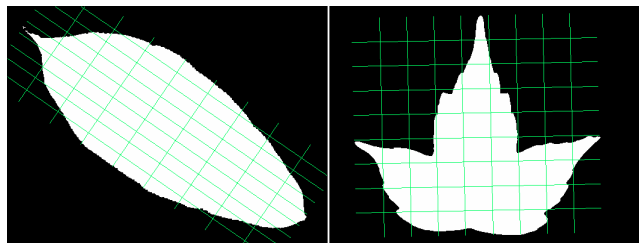
Fig. 6. Equidistant lines.

Algorithm 2: Method one to identify lobed from simple leaf.

Input: A binary image of leaf L, N: Number of lines, T: Threshold
Output: *0* if leaf is lobed, *1* if leaf is simple, *2* if the algorithm can not distinghish

1 Compute $\mathcal{L}_1$ and $\mathcal{L}_2$ using Algorithm 1
2 Draw N equidistant lines parallel to $\mathcal{L}_1$
3 Draw N equidistant lines parallel to $\mathcal{L}_2$
4 Count the number of changes along each line. Store it in an array A_1 and A_2 respectively;
5 Find the greatest number in array A_1 and delete it;
6 Find the greatest number in array A_2 and delete it;
7 Count the number of elements in A_1 and A_2 which are greater than T, call it W_1 and W_2
8 **if** $W_1 > T$ *and* $W_2 > T$ **then**
9 $\quad$ **return** 0
10 **else**
11 $\quad$ **if** $W_1 = 0$ *and* $W_2 = 0$ **then**
12 $\quad\quad$ **return** 2
13 $\quad$ **else**
14 $\quad\quad$ **return** 1

Lines traced on the body of the leaf are based on reference axes. These axes are computed regardless the orientation of the image. It is important to say that our method does not vary when orientation changes. Likewise, changes of color along lines do not vary when scale changes. Our method is also invariant to scaling.

4 Experiments and Results

In this section, we present the results of experiments. We measured the capabilities of our proposals to identify lobed simple leaves. Our method was tested with the images in Flavia data set. Because in the literature there are not features specifically designed to identify lobed leaves, we do not compare the obtained results with others methods. Instead, we measure accuracy, specificity and sensitivity of the two introduced methods.

4.1 Detection of Lobed Simple and Unlobed Simple Leaves

Henceforth, our method will be referred as M_L. In order to measure the performance of M_L, we use Flavia data set [20]. In all our experiments we did not rotate or scale any image.

We manually identified the type of leaf and added a label (binary attribute) called Lobed to each leaf. The value of this attribute was set to true for the images of leaves of classes C6, C8, C28 and C32 (lobed simple leaves with smooth margins). For the rest of the leaves the value of the attribute was set to false (unlobed simple leaves). This identification is used to test the performance of our method.

The confusion matrix obtained for M_L is presented in Table 1. The positive cases correspond to lobed simple leaves, whereas the negative cases are the unlobed simple ones.

Table 1. Confusion matrix for M_L.

Prediction		
Unlobed	Lobed	**Type of leaf**
1,492	201	Unlobed
58	156	Lobed

Based on these last table, the following measures can be obtained:

− **Accuracy**: the proportion of the total number of predictions (positive and negative) that were correct.
− **Sensitivity or Recall**: the proportion of actual lobed leaves which are correctly identified.
− **Specificity**: the proportion of actual simple leaves which are correctly identified.

Table 2 shows the performances of our proposal.

Table 2. Performance of proposed method.

	Accuracy	Sensivity or recall	Specificity
M_L	86.42%	0.7290	0.8813

To measure the effect of our method in the performance of classification methods, we use 10-fold cross validation. Table 3 summarizes the classification accuracy achieved by each classification method. We observed that performances of classifiers are lower than those reported in the literature. However, in our experiments we only took into account six basic leaf features. This number is lesser and simpler than the used in many other works [7], [8,13,19]. Our goal is to compare basic leaf features with our proposal, as we consider it a basic leaf feature too.

The method which obtains the best performance is Multiclass classifier. This method decompose the multiclass problem into simpler ones, which are solved

Juan Carlos Flores-Bastida, Asdrúbal López-Chau, Rafael Rojas-Hernández, Valentin Trujillo-Mora

Table 3. Effect of type of leaf (binary attribute) on seven classification methods.

Method	Classification accurary (%)	
	Six features	*With our method*
1 C4.5	58.78	**60.15**
KNN (K=1)	64.13	**64.24**
Random Forrest	65.44	**66.60**
Multi Class Classifier (Weka)	68.38	**71.95**
NN	**72.94**	72.93
Naive Bayes	55.06	**56.58**
Random Tree	56.63	**57.63**

with logistic regression. The method with second best performance is Random forest. This method uses 100 trees trained with a subset of attributes, and then uses a mechanism of votes to make predictions.

5   Conclusions

Many classic plant identification methods use dichotomous keys that take into account specific features of leaves, such as aspect ratio, leaf area, area convexity, diameter, among others. Motivated by this, we designed a new method to discriminate automatically between unlobed simple and lobed simple leaves. Our method detects changes between background and leaf (and vice versa) in binary images, previously obtained from color images. The unlobed simple and lobed simple leaves feature is an auxiliary characteristic to classification methods, since is independent of rest of leaf features, which when combined achieve better classification results.

The principal characteristic of our method is that it is invariant to rotation and scale of images, because we find a new axes in the leaf. With this axes all the operations can be defined as in linear algebra, line equation, parallel and orthogonal lines equations.

Currently, we are working on designing new compound-leaf features to detect the number of leaflets, also, we are improving basic leaf features to apply in real-world scenarios with challenging conditions.

References

1. Aakif, A., Khan, M.F.: Automatic classification of plants based on their leaves. Biosystems Engineering 139, 66 − 75 (2015), `http://www.sciencedirect.com/science/article/pii/S1537511015001373`

2. Baker, R., Dengler, H.: Leaf Key to Common Trees in Maryland. Extension bulletin, Cooperative Extension Service, University of Maryland (1970), `https://books.google.com.mx/books?id=DbosAQAAMAAJ`

3. Bama, B.S., Valli, S.M., Raju, S., Kumar, V.A.: Content based leaf image retrieval (cblir) using shape, color and texture features. Indian Journal of Computer Science and Engineering 2(2), 202–211 (2011)

4. Berretti, S., Del Bimbo, A., Pala, P.: Retrieval by shape similarity with perceptual distance and effective indexing. Multimedia, IEEE Transactions on 2(4), 225–239 (2000)

5. Cerutti, G., Tougne, L., Coquin, D., Vacavant, A.: Leaf margins as sequences: A structural approach to leaf identification. Pattern Recognition Letters 49, 177 – 184 (2014), `http://www.sciencedirect.com/science/article/pii/S0167865514002335`

6. Cerutti, G., Tougne, L., Mille, J., Vacavant, A., Coquin, D.: Understanding leaves in natural images–a model-based approach for tree species identification. Computer Vision and Image Understanding 117(10), 1482–1501 (2013)

7. Chaki, J., Parekh, R., Bhattacharya, S.: Plant leaf recognition using texture and shape features with neural classifiers. Pattern Recognition Letters 58, 61 – 68 (2015), `http://www.sciencedirect.com/science/article/pii/S0167865515000586`

8. Di Ruberto, C., Putzu, L.: A fast leaf recognition algorithm based on svm classifier and high dimensional feature vector. In: Computer Vision Theory and Applications (VISAPP), 2014 International Conference on. vol. 1, pp. 601–609. IEEE (2014)

9. Dozier, H., Mills, R.: Leaf key to common trees in Louisiana. http://www.lsuagcenter.com/NR/rdonlyres/BA8FFA18-B7CD-4D98-88FF-AF234D5F9ACD/18437/pub1669LeafKey.pdf (Dec 2016), LSU AGCenter

10. Ehsanirad, A.: Plant classification based on leaf recognition. International Journal of Computer Science and Information Security 8(4), 78–81 (2010)

11. Gwo, C.Y., Wei, C.H., Li, Y.: Rotary matching of edge features for leaf recognition. Computers and Electronics in Agriculture 91, 124 – 134 (2013), `http://www.sciencedirect.com/science/article/pii/S0168169912002906`

12. Harish, B., Hedge, A., Venkatesh, O., Spoorthy, D., Sushma, D.: Classification of plant leaves using morphological features and zernike moments. In: Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on. pp. 1827–1831 (Aug 2013)

13. Kalyoncu, C., Toygar, O.: Geometric leaf classification. Computer Vision and Image Understanding 133, 102 – 109 (2015), `http://www.sciencedirect.com/science/article/pii/S1077314214002124`

14. Newcomb, L.: Newcomb's Wildflower Guide. Little, Brown (1989), `https://books.google.com.mx/books?id=7DBvQgAACAAJ`

15. Oregon State, U.: Dichotomous key. http://oregonstate.edu/trees/dichotomous_key.html (2016), consulted 20-04-2016

16. Qi, H.N., Yang, J.G.: Sawtooth feature extraction of leaf edge based on support vector machine. In: Machine Learning and Cybernetics, 2003 International Conference on. vol. 5, pp. 3039–3044 Vol.5 (Nov 2003)

17. Randall, D.J.: Trees of iowa: An interactive key. http://www.extension.iastate.edu/forestry/iowa_trees/key/key.html (December 2016), Iowa State University

Juan Carlos Flores-Bastida, Asdrúbal López-Chau, Rafael Rojas-Hernández, Valentin Trujillo-Mora

18. Vijayalakshmi, B.: A new shape feature extraction method for leaf image retrieval. In: Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012). pp. 235–245. Springer (2013)
19. Wang, B., Brown, D., Gao, Y., Salle, J.L.: March: Multiscale-arch-height description for mobile retrieval of leaf images. Information Sciences 302, 132 – 148 (2015), `http://www.sciencedirect.com/science/article/pii/S0020025514007282`
20. Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y., Chang, Y.F., Xiang, Q.L.: A leaf recognition algorithm for plant classification using probabilistic neural network. CoRR abs/0707.4289 (2007), `http://dblp.uni-trier.de/db/journals/corr/corr0707.html#abs-0707-4289`