2011-01

Fractal analysis of leaf-texture properties as a tool for taxonomic and identification purposes: a case study with species from Neotropical Melastomataceae (Miconieae tribe)

Plant Systematics and Evolution, Wien : Springer Wien, v. 291, n. 1/2, p. 103-116, Jan. 2011
http://www.producao.usp.br/handle/BDPI/49624

Plant Syst Evol (2011) 291:103–116
DOI 10.1007/s00606-010-0366-2

ORIGINAL ARTICLE

Fractal analysis of leaf-texture properties as a tool for taxonomic and identification purposes: a case study with species from Neotropical Melastomataceae (Miconieae tribe)

Davi Rodrigo Rossatto · Dalcimar Casanova · Rosana Marta Kolb · Odemir Martinez Bruno

Received: 14 January 2010 / Accepted: 21 April 2010 / Published online: 29 October 2010

Abstract Melastomataceae is a common and dominant family in Neotropical vegetation, with high species diversity which leads to a large variation in some morphological structures. Despite this, some species of Melastomataceae are very similar in their external leaf morphology, leading to difficulties in their identification without the presence of reproductive organs. Here we have proposed and tested a computer-aided texture-based approach used to correctly identify and distinguish leaves of some species of Melastomataceae that occur in a region of Neotropical savanna in Southeastern Brazil, also comparing it with other previously proposed approaches. The results demonstrated that our approach may clearly separate the studied species, analyzing the patterns of leaf texture (both adaxial and abaxial surfaces), and achieving better accuracy (100%) than other methods. Our work has suggested that leaf texture properties can be used as a new characteristic for identification, and as an additional source of information in taxonomic and systematic studies. As the method may be supervised by experts, it is also suitable for discrimination of species with high morphological plasticity, improving the automated discrimination task. This approach can be very useful for identification of species in the absence of reproductive material, and is a rapid and powerful tool for plant identification.

Keywords Leaf texture · Fractal analysis · Volumetric fractal · Melastomataceae · Plant identification · Computer vision

D. R. Rossatto
Laboratório de Fisiologia Vegetal,
Departamento de Botânica, Universidade de Brasília-UnB,
Caixa postal 04457, 70904-970 Brasília, DF, Brazil
e-mail: drrossatto@gmail.com

D. Casanova · O. M. Bruno (✉)
Instituto de Física de São Carlos,
Universidade de São Paulo-USP, Av. Trabalhador
São Carlense, 400, 13560-970 São Carlos, SP, Brazil
e-mail: bruno@ifsc.usp.br

D. Casanova
e-mail: dalcimar@gmail.com

R. M. Kolb
Faculdade de Ciências e Letras,
Departamento de Ciências Biológicas,
Universidade Estadual Paulista-UNESP, Av. Dom Antônio,
2100, 19806-900 Assis, SP, Brazil
e-mail: rosanakolb@hotmail.com

Introduction

The process of taxonomy and plant identification has become a systematic activity; however, mankind has searched for more efficient plant-classification systems, and also for mechanisms that facilitate plant identification (Judd et al. 2008). Botanical classification is basically done by using herbaria (Bridson and Forman 1998) in which morphological and anatomical characteristics of vegetative and reproductive organs from different species can be examined and studied by observation (Leenhouts 1968), being identified by comparison with voucher specimens (DeWolf 1968) and by using diagnostic keys (Dallwitz 1974). The accuracy or success of plant identification relies on the user or specialist's experience and interpretation. The traditional discrimination method is based on morphological studies and it sometimes depends on subjective visual assessment leading, thus, to failures in the detection of very small and specific differences, for example those occurring among very similar plant species. Moreover, this procedure is very dependent on the user's knowledge of specialized terminology.

For cataloging activity, which is essential in any ecological, botanical or taxonomical study, it is necessary to collect and dry a branch of the plant, to process the voucher, comparing it with other vouchers in the herbaria, using diagnostic keys, and sending the material to specialists for confirmation (Pankhurst 1978; Holgren and Holgren 1992). One of the major problems in this process is that flowers and fruits, the main sources used for diagnostic of characteristics, are not available for studies throughout the year, but only at certain times, what causes difficulties for plant identification/classification (Ash et al. 1999). Besides, the dehydration process may cause loss of important and very informative properties for plant identification, for example color, texture, and brightness (Blanco et al. 2006).

Many attempts have been made with the purpose of performing automated plant identification, making this process faster, more precise and useful for botanists and also for plant specialists, thus improving taxonomic work (Wang et al. 2008). Some of these studies have focused on the analysis of external leaf attributes, for example the geometry and contour line of leaves (Mugnai et al. 2008; Backes et al. 2009a; Wang et al. 2003, 2005; Plotze et al. 2005; Lee and Chen 2006), whereas others have used such characteristics as the color of leaf surfaces (Castro-Esau et al. 2004; Slaughter et al. 2008). Another category in computer-aided plant identification is analysis of leaf texture (Mancuso 2002; Dean and Ashton 2008), which sometimes relies on non-computerized extraction of texture features (Dean and Ashton 2008). Finally, the work of Mugnai et al. (2008) used hybrid methodology (texture and shape-based features) to classify *Camellia japonica* L. genotypes.

Leaves can be easily found and collected practically everywhere in all seasons, and the analysis of their morphological and anatomical characteristics is easier, providing important information for identification and taxonomy purposes (Bailey 1951; Dickison 2000; Endress 2003). Another advantage of using leaves is that they are very complex in their form, shape, and color, providing greater scope for differentiation using geometric aspects (Persson and Gustavsson 2001), and also in texture, a characteristic that can bring new sources of information for plant identification (Backes et al. 2009b). The analysis of external leaf morphology without a computer-analysis approach could lead to doubts in the identification of plant species with very similar morphology, which occurs in plant species belonging to the same genus or from closely related genera, or even inside the same family.

Among Melastomataceae tribes, Miconieae is one of the most common and dominant in the Neotropics, containing approximately 2,000 species among 30 genera (Michelangeli et al. 2004), occurring practically in all plant formations of South America (Martin et al. 2008). In Savanna formations of Brazil, *Miconia* and *Leandra* are the most diversified genera, with a high diversity of species (Mendonça et al. 2008) with similar leaf morphology. Many taxonomic and phylogenetic studies have been conducted with *Miconia* and *Leandra* genera in several plant formations of Brazil (Martin et al. 2008; Goldenberg et al. 2008; Goldenberg and Martin 2008). In the same family, tribe, or genus, the within-leaf dissimilarity and within-species similarity are higher, making classification and identification difficult using only external morphological characteristics.

The main objective of this study was to investigate the taxonomic and identification value of quantitative and qualitative leaf surface characteristics across a number of closely related species from the Miconieae tribe, introducing a new and powerful tool to improve taxonomic and identification work. These species share similar leaf morphology, which causes some difficulties in their identification, even with reproductive material. Here we have analyzed surface texture from different Miconieae species by fractal analysis, using multivariate statistical techniques. We have compared our approach with other methods commonly used for analyzing leaf contour, texture, or texture and shape (hybrid), employing the same dataset as in our work, to provide evidence for the strength of the proposed model.

Also, we have tested our model using different hypotheses of plant species delimitation, to see how strong our approach is. For this purpose, we have chosen three specimens that may be viewed as either three different species or three variations of the same species.

Materials and methods

Plant material

Plant leaves were collected at Assis Ecological Station, Assis, São Paulo State, Brazil, situated between 22° 33′65″–22′36′68″S and 50′22′29″–50′23′00″W. The vegetation is a Neotropical savanna locally called “cerrado”. According to Köppen’s classification, the climate in this ecological station is transitional between Cwa and Cfa, with concentrated rainfall in summer, with annual precipitation under 1,400 mm and temperatures around 21.8°C. Dry season occurs between June and September, and wet season between October and May. We selected some species of the Melastomataceae family, especifically from the Miconieae tribe (Table 1), comprising ten species from the *Miconia* genus and two to five species from the *Leandra* genus, which had similar patterns of leaf morphology. We consulted a plant specialist and the species presented here as 2, 3, and 4 can be either variations of the same species

Table 1 Miconieae species selected for this study at the Estação Ecológica de Assis, São Paulo State, Brazil

Species	No. on Fig. 1	No. of leaves	Location of occurrence	Voucher number
Leandra melastomoides Raddi	1	19	Forest	HRCB 44091
Leandra sp1	2	20	Savanna	HRCB 44075
Leandra sp2	3	10	Forest	HRCB 44079
Leandra sp3	4	20	Savanna/forest	HRCB 44083
Miconia albicans (Sw.) Triana	5	21	Savanna	SPSF 35912
Miconia chamissois Naudin	6	20	Marsh camp	SPSF 36994
Miconia cinerascens Miq.	7	21	Forest	SPSF 34926
Miconia fallax DC.	8	20	Savanna	SPSF 35833
Miconia langsdorfii Cogn.	9	20	Forest	SPSF 35815
Miconia ligustroides (DC.) Naudin	10	20	Forest	SPSF 35823
Miconia pusilliflora (DC.) Naudin	11	20	Forest	HRCB 45614
Miconia sellowiana (DC.) Naudin	12	20	Forest	HRCB 44109
Miconia stenostachya DC.	13	20	Savanna	SPSF 37026
Miconia theaezans (Bonpl.) Cogn.	14	20	Marsh camp	SPSF 35844
Miconieae 1	15	14	Forest	HRCB 44086

(*Leandra aurea* (Cham.) Cogn., see Souza and Baumgratz 2009) or different species, impossible to ascertain without the presence of flowers. The species numbered 15 can be from *Leandra* or *Ossaea* genera, being also impossible to determine without the flowers. For this reason, we have preferred to refer to it as simply belonging to the Miconieae tribe. Species were identified through specific literature (Durigan et al. 2004; Goldenberg 2004; Camargo 2008; Ramos et al. 2008). Vegetative materials from *Leandra* species were sent to a specialist for identification. In addition, collected materials were identified by comparison with previous plant materials deposited in the University of Brasilia Herbarium (UB). Some vouchers were deposited at "Herbário Don Bento Pickel", São Paulo, SP (SPSF) and others at "Herbário Rioclarense", Rio Claro, SP (HRCB).

Fully expanded leaves were collected in July 2009, during the morning, from different individuals (4–5 per species). These were immediately placed in sealable plastic bags and carried to the laboratory for scanner digitization.

Image acquisition and determination of phyllometric characteristics

Each leaf was digitized with a traditional scanner (HP Scanjet 3800), with a resolution of 1,200 dpi (dots per inch) and saved in PNG format. The high resolution was used to enable access to details in image texture. To avoid rotation problems, all leaves were oriented in same direction while assembled on the scanner. Both surfaces (abaxial and adaxial) of the leaf epidermis were digitized. The adaxial surface of all species can be seen in Fig. 1. Ten 128×128 texture samples were randomly extracted from both sides of each leaf, without overlapping (Fig. 2). This procedure is necessary to minimize texture variation in a single leaf and to avoid texture samples containing noise caused by fungi, disease, or lesions. In addition, the extracted textures were converted to grayscale for subsequent analysis.

Computer description of leaf texture

The key issue in leaf classification is the texture features observed in the images. Texture is recognized as one of the most important sources of information in human visual perception, although there is no formal definition of it. In general terms, natural textures, for example leaf surfaces, have random but persistent patterns, and do not contain any detectable quasi-periodic structure (Kaplan 1999). The same author has suggested that fractal theory is better than statistical, spectral, and structural approaches for describing natural textures. In fact, the volumetric fractal dimension has been very useful for identification and classification of leaf textures (Backes et al. 2009b). So, we decided to describe and study texture using the volumetric fractal dimension (Backes et al. 2009b).

Considering an image $f(x, y)$ in gray scale, we can state that its three-dimensional representation can be given by $S(x, y, z)$, where (x, y) are the spatial coordinates of the image and the third coordinate z, the intensity of the gray color. By dilation of surface S until a radius r, estimated at each step, the value of $V(r)$ is given by Eq. 1:

$$V(r) = \{p \in R^3 | \exists p' \in S : |p - p'| \leq r\} \quad (1)$$

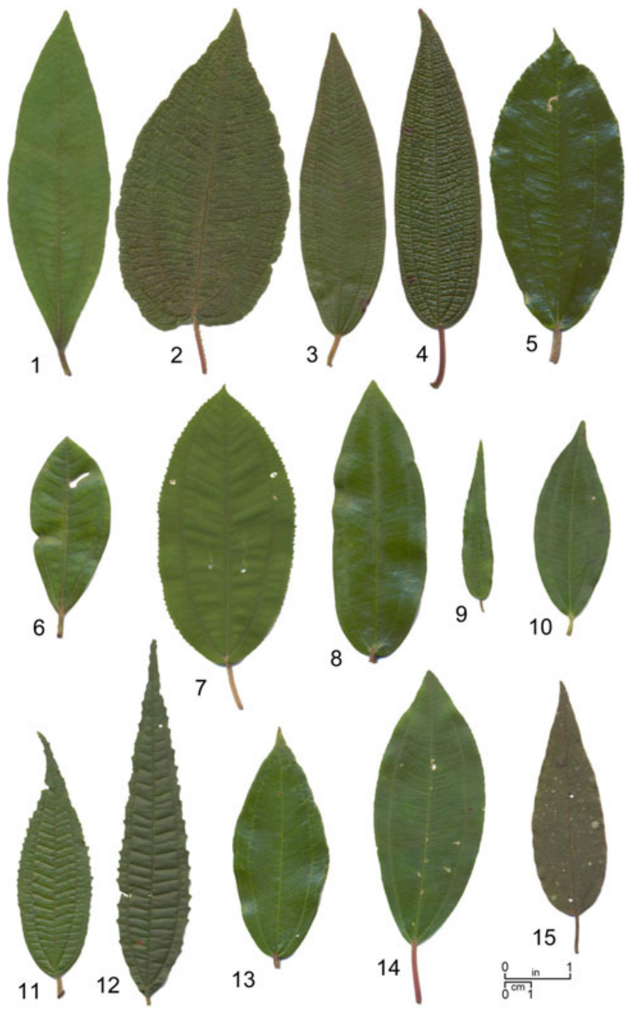

Fig. 1 Leaf adaxial surfaces of the Melastomataceae species studied

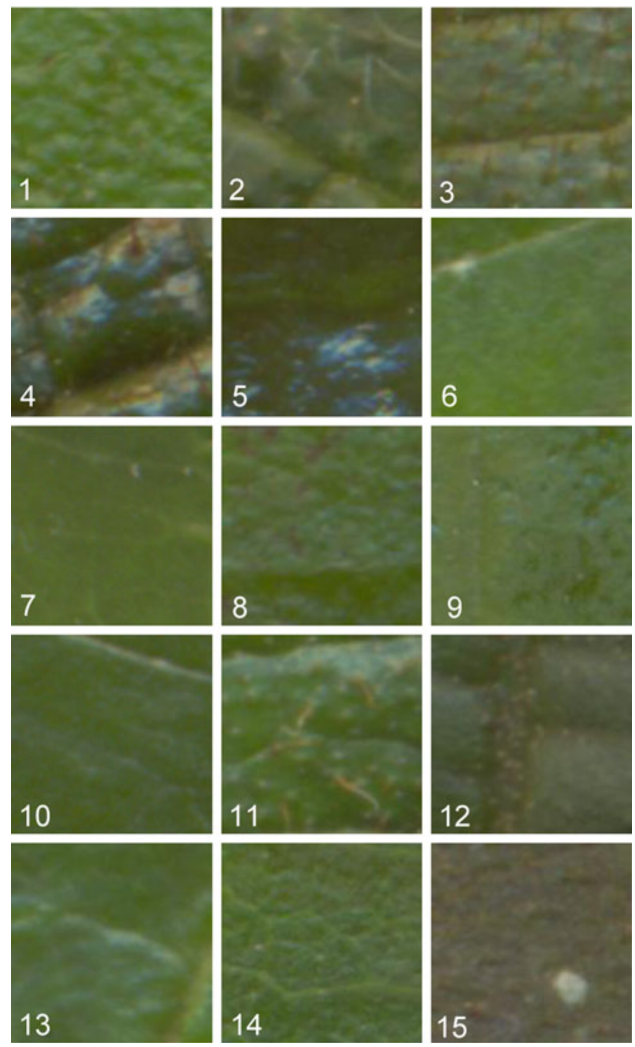

Fig. 2 Examples of leaf textures obtained from adaxial surfaces showing high within-species similarity

where $p' = (x', y', z')$ is a point in R^3 whose distance from $p = (x, y, z)$ is smaller than or equal to r, and $V(r)$ is the influence volume obtained by dilation of each point p of S, using a sphere of radius r.

In this method, the arrangement of points alters the process of expansion. While the value of r grows, the spheres produced by different points on the surface begin to interact. This interaction causes effects on the amount of $V(r)$, i.e., it affects the volume calculated for a certain radius. This feature enables the perception of changes in the texture pattern, because different textures differ in the organization of their pixels. Figure 3 exemplifies this process by the dilation of surface S to different values of r.

As a result of this procedure, each texture produces a characteristic value of $V(r)$ for each stage of the dilation process. This makes possible the use of the values of $V(r)$ as descriptors of texture, because $V(r)$ describes, indirectly, the organization of the pixels. As shown by Backes et al. (2009b), all radii between 9 and 16 yield good results, thus, in this study we chose to use a radius of 15. Thus, the feature vector is composed of 189 descriptors arising from the logarithm of the volume of influence $V(r)$, calculated for all values of $r \in E$, where E is the set of Euclidean distances to a radius $r(\max) = 15$:

$$E = \left\{1, \sqrt{2}, \sqrt{3}, \ldots, r_{\max}\right\} \quad (2)$$

$$\mathbf{x} = [\log V(1), \log V(\sqrt{2}), \ldots, \log V(r_{\max})]. \quad (3)$$

Classification strategy

With the observed features it is possible to identify different plant species. However, direct classification was not used in our strategy. As shown before, ten texture samples were randomly extracted from each leaf and, therefore, we had ten different feature vectors for each leaf. To minimize problems with noise presence and texture variation we

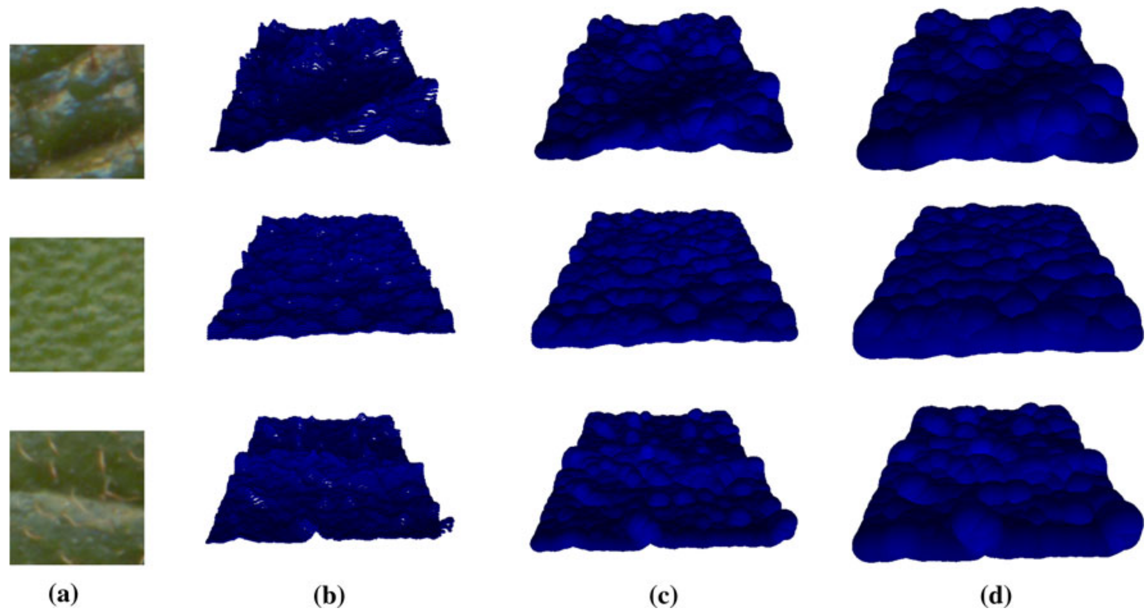

Fig. 3 Effect of volume characteristics for specific texture pattern: **a** original leaf texture; **b** $r = 2$; **c** $r = 5$; **d** $r = 10$ (adapted from Backes et al. 2009b)

calculated the mean feature vector. With a single feature vector for each leaf we chose to use the naive Bayes classifier (Mitchell 1997). In addition, we used a tenfold cross validation. The naive Bayes classifier is a simple probabilistic classifier based on Bayes's theorem. This classifier assumes the conditional independence hypothesis among attributes. The classification rule can be seen as the attribution of the object to the group with the highest conditional probability. For g groups, the Bayes rule assigns an object to the group i when:

$$P(i|\mathbf{x}) > P(j|\mathbf{x}), \quad \text{for} \quad \forall j \neq i. \tag{4}$$

In this case, assuming the hypotheses of independence, we have for the aleatory variables:

$$P(i|x) = \frac{P(i)\prod_{k=1}^{n} P(x_k|i)}{\prod_{k=1}^{n} P(x_k)} \tag{5}$$

where:

$$P(x_k|i) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} \mathrm{e}^{\frac{(x_i-\mu_{ik})^2}{2\sigma_{ik}^2}} \tag{6}$$

being $P(\mathbf{x}|i)$ is the probability of obtaining a particular set of features $\mathbf{x}$, given that the object belongs to the group i and $P(i)$ is the probability a priori, that is, the probability of choosing the group i without any feature of the known object.

The naive Bayes is a statistical classifier assuming the conditional independence hypothesis among features. Although naive Bayes is very competitive even when this assumption is violated, this methodology is not appropriate when you have characteristics that are naturally dependent and highly correlated. To solve this problem we have used canonical discriminant analysis. This method removes the correlations among features and optimizes the separation between classes. Given the original features, we can obtain the so called canonical variables; these variables are not correlated and can be used in naive Bayes classifier, respecting the hypothesis of independence among attributes. Canonical discriminant analysis (CDA) is a multivariate statistical technique with the objective of maximizing the separation between classes, and is similar to principal-components analysis (PCA) and to canonical correlations (McLachlan 1992). Also known as multiple discriminant analysis (MDA), CDA seeks linear combinations of original variables into so-called canonical variables.

Given the matrix S, indicating the total dispersion among the feature vectors, defined as:

$$S = \sum_{i=1}^{N} (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)' \tag{7}$$

and the matrix S_i indicating the dispersion of objects of C_i:

$$S_i = \sum_{i \in C_i} (\mathbf{x}_i - \mu_i)(\mathbf{x}_i - \mu_i)' \tag{8}$$

we can define the intra-class variability S_{intra} (indicating the combined dispersion within each class) and the interclass variability S_{inter} (indicating the dispersion of the classes in terms of their centroids) as:

$$S_{\text{intra}} = \sum_{i=1}^{K} S_i \tag{9}$$

$$S_{\text{inter}} = \sum_{i=1}^{K} N_i(\mu_i - \mu)(\mu_i - \mu)' \quad (10)$$

where K is the number of classes, N, the number of samples, N_i, the number of objects in class i, C_i, the set of samples of class i, μ, the global average, and μ_i, the average of objects in class i. For these measures of dispersion we have necessarily:

$$S = S_{\text{intra}} + S_{\text{inter}} \quad (11)$$

Thus, the ith canonical discriminant function is given by:

$$Z_i = a_{i1}\mathbf{X}_1 + a_{i2}\mathbf{X}_2 + \cdots + a_{ip}\mathbf{X}_p \quad (12)$$

where p is the number of features of the model and a_ij are the elements of the eigenvector $a_i = (a_{i1}, a_{i2}, ..., a_{ip})$ of matrix C given by:

$$C = S_{\text{inter}} \times S_{\text{intra}}^{-1} \quad (13)$$

This formulation leads to a condition where there is no correlation between Z_i and $Z_1, Z_2, \ldots,$ within the classes. From p-original variables the p-principal components can be obtained. However, in general, a reduction in the number of variables to be assessed is desired, i.e., that the information contained in the p-original variables be replaced by the information contained in k ($k < p$) uncorrelated principal components. Thus, the system of random variability of the original vector with p-original variables is approximated by the variability of the random vector containing the k-principal components.

Our experiments were designed to classify each test image into a single class (species) using the tenfold cross-validation strategy mentioned above. As the true species for all collected leaves are known, the classification accuracy can be defined as the ratio of the number of images correctly classified to the total number of test images. An overview of all process can be seen in Fig. 4.

Nam and Hwang (2005) and Nam et al. (2008) have discussed how images with similar color or texture, as has been presently found in some leaves, could be more effectively separated by shape-based image retrieval than by use of color or texture. To verify this, we compared our approach with a contour-based (Backes et al. 2009a) method, a texture-based method (Mancuso 2002), and an hybrid method (Mugnai et al. 2008), that uses texture and shape-based features together for leaf identification; in all analysis we used the same species and images used in our own texture method. Backes et al. (2009a) showed that classification using the complex network method is better than CSS (Abbasi et al. 1997), Fourier (Ferson et al. 1985; Neto et al. 2006), Zernike moments (Zhenjiang 2000), and fractal dimension (Plotze et al. 2005; Bruno et al. 2008) for leaves and other complex shapes. In our experiments we used the degree descriptor with $T_0 = 0.025$, $T_{\text{inc}} = 0.075$, and $T_Q = 0.775$. In Mancuso (2002) the texture features were extracted with the fractal dimensions of the color channels from 12 genotypes of *Vitis vinifera* L. Mugnai et al. (2008) discriminated 25 *C. japonica* genotypes with phyllometric and fractal data together. In our experiments we used the exact features proposed by these authors. In addition, for all comparisons, the same classification strategy (CDA + naive Bayes) was used.

Working hypotheses for the classification of images

Because it was not possible to determine if *Leandra* sp1, *Leandra* sp2, and *Leandra* sp3 are different species or phenotypic variations of a same species (because of the absence of reproductive material), we worked with both hypotheses and tested whether our classification strategy could have the same accuracy in both cases. The classifier used here (Bayes + CDA) is a type of supervised classifier. In this model, the number of classes (i.e., the number and name of species) is previously input to the system. In this way, the system needs to be adjusted with the number of classes (species) and also with the mathematical model of each class (i.e., the user trains the system with examples of each class). In this approach, someone can configure the system with a variable number of classes. The key to obtaining good classification performance is to have a good set of descriptors. With that, the model can for example, match the knowledge of a plant specialist. If the training data have 15 species (in the case of L. sp1, L. sp2, and L. sp3 being different species), the classifier will create a model to distinguish among the 15 species. But if the training data have 13 species (L. sp1, L. sp2, and L. sp3 being phenotypic variations of a same species), the classifier will create a different model to distinguish among these 13 species.

Results and discussion

This proposed classification method was 100% effective using texture features of adaxial leaf surfaces, considering 15 species (Table 2). For that, only ten canonical variables are used in the naive Bayes classifier. These ten main components represent 99.99% of total variance, demonstrating that the volumetric fractal dimension has very high correlated features. In Fig. 5 we can see the high correlation among all 189 original features, and in Fig. 6, after canonical transformation, the low correlation among the ten first canonical variables. The small number of canonical variables used enables creation of efficient and accurate leaf image retrieval systems.

In canonical eigenspace (Fig. 7), leaf samples from the same species are clustered more tightly than with samples

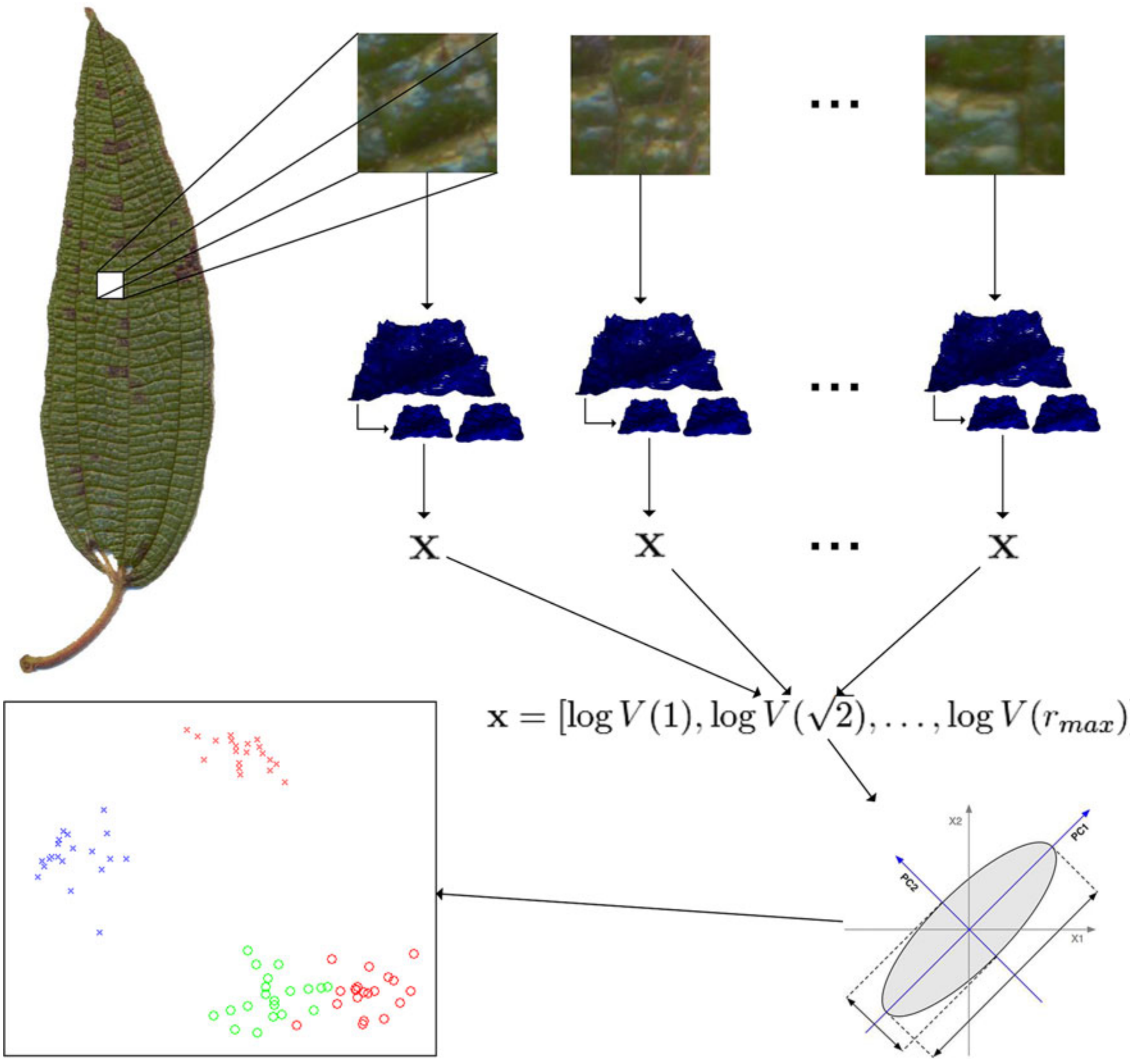

Fig. 4 Overview of the approach used in this work for plant species identification

Table 2 Leaves classification summary for texture, contour, and hybrid-based approaches, considering L. sp1, L. sp2, and L. sp3 as different species

Method	No. of features	No. of correctly identified leaves	Accuracy (%)
Adaxial texture	10	285	100.0
Abaxial texture	10	285	100.0
Backes et al. (2009a)	12	209	73.33
Mancuso (2002)	12	224	78.59
Mugnai et al. (2008)	10	253	88.77

from other species, with few exceptions, where there is no substantial overlap. It is important to emphasize here that the real separation of all species occurs when the Bayes classifier uses all ten canonical variables together. The two canonical variables plotted in the figure, and the clusters formed, give only an idea about the quality of the separation of groups. Therefore, the variability of texture features within-leaf seems to be fairly low, whereas the variability within-species is high, which makes evaluation of the proposed characteristics a good and promising method for automatic plant taxonomy and identification.

In order to verify the effect of leaf face we repeated the experiment with the abaxial surfaces. The same results were achieved in this analysis. For both adaxial and abaxial faces within-species and within-genus separation is good (Fig. 8), and only ten canonical variables are necessary to achieve accuracy of 100% (Table 2). The work of Ramos and Fernández (2009), with microphotographs, suggests that the abaxial epidermis side is more discriminative than the adaxial one, because of the presence of stomata, trichomes, and other morphological structures related to phylogenetic aspects. In our experiments, with digitized images, both sides furnished 100% accuracy. So, further experiments should be conducted to confirm the adaxial

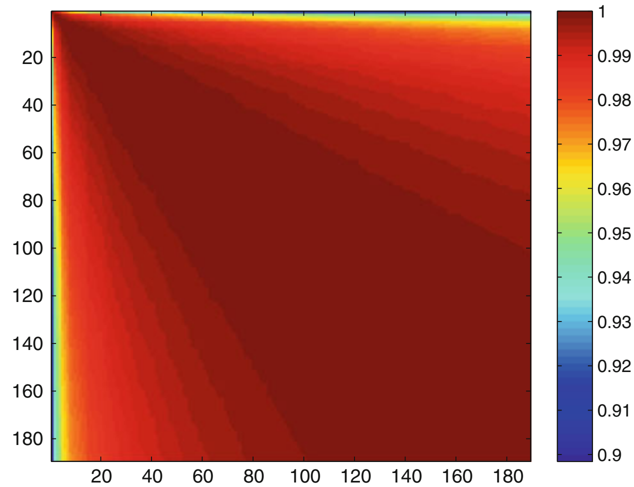

Fig. 5 Correlation coefficients for all 189 Minkowski 3D features, obtained from leaves' adaxial surfaces. All values are highly correlated

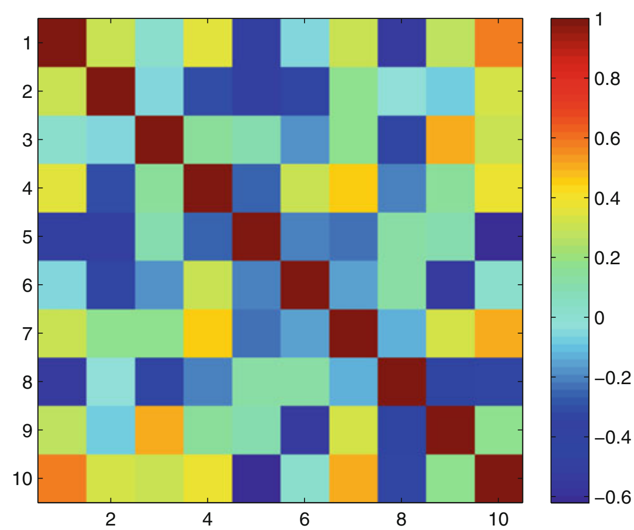

Fig. 6 Correlation coefficient after canonical transformation (adaxial surface). The ten canonical features are poorly correlated

surface texture as a good indicator for clustering plant species. On the other hand, both adaxial and abaxial features can be used together to improve the quality of discrimination. In this case it was not necessary because both resulted in 100% accuracy in plant identification.

Different leaf-surface patterns could be caused by the presence of stomata and trichomes on the abaxial surface (Dean and Ashton 2008), with the density and length of these structures being under phylogenetic and environmental control. Despite the variation of stomatal patterns that occurs in different species growing under different environmental conditions (Hlwatika and Bhat 2002; Gratani et al. 2006; Pearce et al. 2006) or even in different leaves of the same individuals (Rossatto and Kolb 2009), Beaulieu et al. (2008) have demonstrated that some stomatal patterns are intrinsic to determined species, and, sometimes, can be used as a significant source of information for identification and classification. These structures were generally assessed by the observer, by measuring their density and length (Hetherington and Woodward 2003); the analysis proposed here can, however, extract much more information hidden in the patterns of the leaf surface. On the other hand, differences on adaxial surfaces can be linked to patterns of cuticle/epicuticular wax deposition (Salatino et al. 1986), which in turn are also affected by phylogenetic and environmental factors. Furthermore, differences in adaxial surface texture can also be related to leaf-venation patterns (Fig. 2), which strongly contribute to differences in texture among species. Leaf venation patterns have been used for taxonomic and identification purposes in several plant families (Klucking 1987; Ash et al. 1999), and with the proposed method they could be assessed as a source of different texture patterns.

Comparison with the other texture, contour, and hybrid methods has shown that those methods are not as good at separating among Miconieae species (Table 2). These results have shown that the contour method used by Nam and Hwang (2005) and Nam et al. (2008) may be applicable to some specific situations, but can fail to separate very similar species with very similar shape and contour of their leaves. The texture approach proposed here, at least for some species in the Miconieae tribe, is better than the contour-based approach, for which accuracy of species identification was only 73.33%. In addition, the accuracy of the texture approach of Mancuso (2002) and the hybrid approach of Mugnai et al. (2008) was not good (78.59 and 88.77%, respectively). Tables 3, 4 and 5 show the confusion matrix for classification of the plant species using contour, texture, and hybrid-based methods. We can see that classification and separation of leaves was 100% accurate for few species only; these results show that these methods are not robust for the studied case, leading to a large confusion among most species.

So, our approach has correctly discriminated differences and similarities, by extracting texture features from foliar surfaces that are morphologically distinct from each other. However, as mentioned in "Materials and methods", there is a possibility that species 2, 3, and 4 may be variations of the same species (*L. aurea*). The work of Souza and Baumgratz (2009) describes the last species, which has high morphological plasticity, for example differences in trichomes and in leaf surface undulation. This makes its distinction in the vegetative phase difficult if leaves with different textures can belong to this taxon. Considering this working hypothesis, we have redefined the classifier method with only 13 species. The results obtained are very similar to those previously described, reaching 100% accuracy using texture features of both leaf surfaces (Table 6). This is not a surprise, because in adaxial and

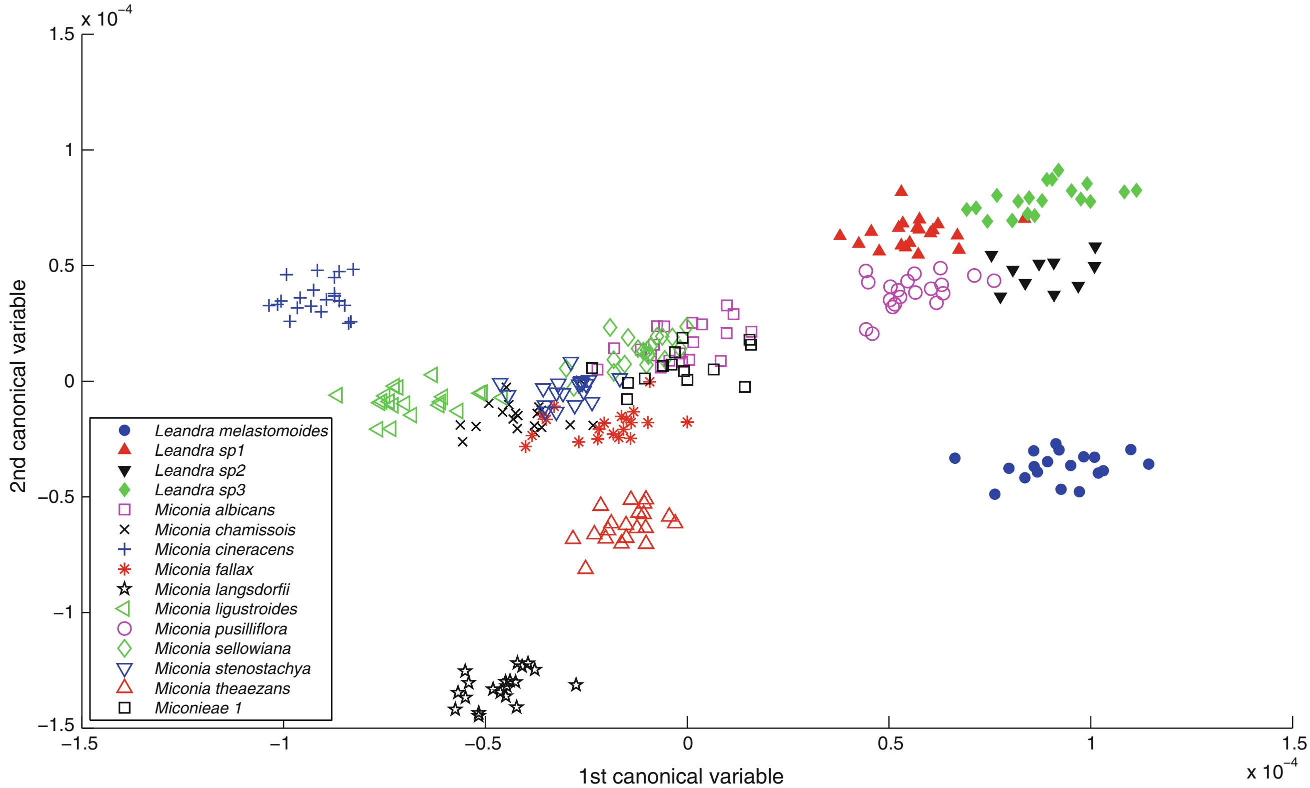

Fig. 7 Minkowski 3D descriptor: first and second canonical variables of leaves' adaxial surfaces—groups formed by the different species

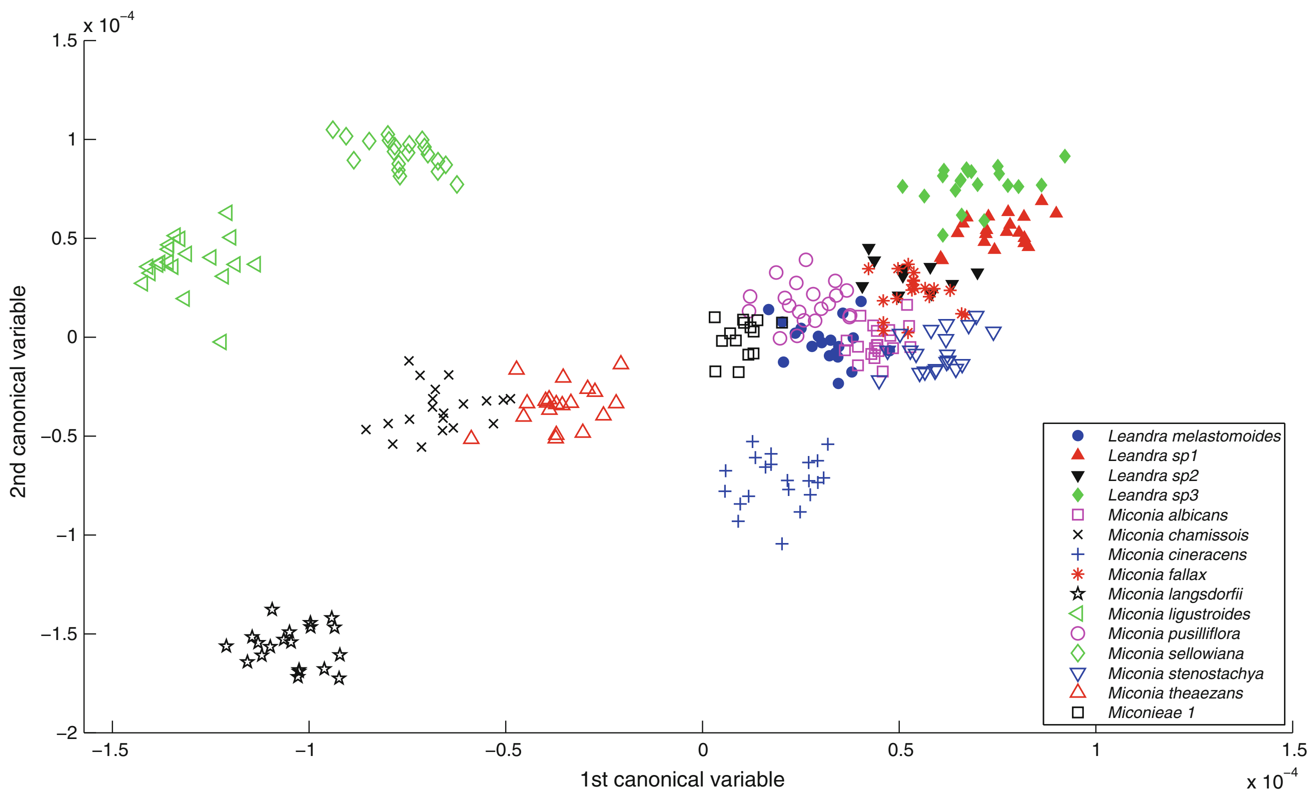

Fig. 8 Minkowski 3D descriptor: first and second canonical variables of leaves' abaxial surfaces—groups formed by different species

Table 3 Confusion matrix for classification of *Leandra* and *Miconia* species using contour descriptors (Backes et al. 2009a)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	73.7		20.0	10.0											
2		85.0					14.3						5.0		7.1
3	15.8		70.0	25.0										5.0	14.3
4				50.0				5.0							7.1
5		10.0			66.7	5.0	9.5						25.0		
6		5.0		15.0	14.3	30.0	9.5						10.0		7.1
7						15.0	52.4								
8						10.0		90.0					5.0	5.0	
9									100						
10						25.0	4.8			95.0				5.0	
11	5.3										90.0				
12												100			
13					19.0	10.0	9.5	5.0					55.0		
14										5.0	10.0			80.0	7.1
15	5.3		10.0			5.0								5.0	57.1

Table 4 Confusion matrix for classification of *Leandra* and *Miconia* species using texture descriptors (Mancuso 2002)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	68.4						9.5				5.0			10.0	
2		75.0	30.0	10.0											7.1
3	5.3	10.0	60.0			5.0					5.0	10.0		5.0	
4		15.0		85.0	4.8					5.0					
5					90.5			5.0					5.0		
6						90.0				10.0				10.0	
7	5.3						85.7			5.0				5.0	
8								85.0					10.0		
9									95.0	5.0				5.0	
10	10.5									70.0					
11				5.0							55.0	10.0			
12											30.0	80.0		5.0	7.1
13					4.8			10.0					85.0		
14	10.5					5.0	4.8		5.0	5.0	5.0			60.0	
15			10.0												85.7

abaxial canonical eigenspaces (Figs. 7, 8), we can see the variations of *L. aurea* forming groups that are close to each other. So, because of the cluster proximity in the canonical eigenspaces, it is computationally possible to create a new classifier that considers the variations of *L. aurea* as a single group. This implies that, when we have a good descriptors' set, our texture method can work perfectly well even if some species show great variability in their leaf surfaces.

The confusion matrixes for contour, texture, and hybrid-based methods are presented in Tables 7, 8, and 9 (number 2 corresponding to L. sp1, L. sp2, and L. sp3 as a single group). Here again, 100% accurate discrimination is achieved for a few species only.

Conclusion

Leaves' texture has been shown to be a good discriminative character and very useful for computer-aided plant classification. Both abaxial and adaxial epidermic surfaces lead to 100% accuracy in species identification, a result that is far superior to that from methods based on contour, texture, or both together. The discriminative quality of our

Table 5 Confusion matrix for classification of *Leandra* and *Miconia* species using hybrid descriptors (Mugnai et al. 2008)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	89.5		20.0												
2		100													
3	10.5		70.0								10.0				
4			10.0	70.0				5.0							
5					100								5.0		
6						80.0							10.0	20.0	
7							95.2								
8				5.0				95.0							
9									100						
10						5.0				80.0				5.0	
11				20.0							85.0				
12				5.0								100			
13													85.0		
14						15.0				15.0				75.0	
15							4.8			5.0	5.0				100

Table 6 Leaves classification summary for texture, contour, and hybrid-based approaches, considering species 2, 3, and 4 as variations of the same species

Method	No. of features	No. of correctly identified leaves	Accuracy (%)
Adaxial texture	10	285	100.0
Abaxial texture	10	285	100.0
Backes et al. (2009a)	12	215	75.43
Mancuso (2002)	12	223	78.24
Mugnai et al. (2008)	10	251	88.07

Table 7 Confusion matrix considering species 2, 3, and 4 as variations of *Leandra aurea*

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	84.2	2.0											
2	5.3	76.0						5.0	10.0				14.3
3		4.0	71.4	5.0	14.3						20.0		
4				15.0	14.3				5.0		15.0		
5		4.0		20.0	66.7						10.0		
6				15.0		100.0					5.0		
7							100.0						
8				20.0				95.0				5.0	
9	10.5	6.0							85.0				7.1
10										100.0			
11		4.0	23.8	15.0	4.8						50.0	5.0	
12			4.8	5.0								75.0	21.4
13		4.0		5.0								15.0	57.1

Using Backes et al. (2009a) descriptors (contour)

Table 8 Confusion matrix considering species 2, 3, and 4 as variations of *Leandra aurea*

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	73.7			5.0	14.3				5.0			15.0	
2		86.0						5.0	10.0	5.0		5.0	7.1
3			95.2								5.0		
4	5.3			80.0				10.0				15.0	
5					81.0								
6						75.0					20.0		
7							90.0					10.0	7.1
8	5.3			5.0			10.0	85.0					
9		10.0		5.0					60.0	15.0		5.0	
10		2.0							25.0	80.0			7.1
11			4.8			25.0					75.0		
12	15.8			5.0								50.0	7.1
13		2.0			4.8								71.4

Using Mancuso (2002) descriptors (texture)

Table 9 Confusion matrix considering species 2, 3, and 4 as variations of *Leandra aurea*

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	100	2.0											
2		82.0							20.0				
3			100								5.0		
4				85.0							5.0	20.0	
5					95.2								
6				5.0		95.0					5.0		
7							100						
8								75.0	5.0			10.0	
9		16.0						10.0	70.0				
10										100			
11						5.0					85.0		
12				10.0				15.0	5.0			70.0	
13					4.8								100

Using Mugnai et al. (2008) descriptors (hybrid)

approach is very good, leading to low within-leaf and high within-species variability. We recognize that further work is required to clarify the physiological and structural mechanisms behind the differences between the species and genera surfaces.

These results suggest that computer-aided plant classification can provide new useful tools for experimental taxonomists and plant morphologists, improving their work, bringing also new forms to assess informative characters that can be useful in systematic and phylogenetic studies. The method proposed here may be a new, important tool for non-taxonomic botanists or ecologists, working with plant species with very similar morphology, because it requires readily available equipment, for example a conventional computer and an optical scanner. Finally, the proposed computational analysis has been shown to be an important tool because it can be calibrated according to the specialist's expertise, thus improving the identification in confused groups with large leaf morphological variation.

Acknowledgments The authors acknowledge Assis Ecological Station and Instituto Florestal for permission to collect the leaves of the studied species, and Dr Renato Goldenberg for helping with plant identification. Odemir M. Bruno gratefully acknowledges the financial support of CNPq (National Council for Scientific and Technological Development, Brazil) (Grant #306628/2007-4 and #484474/2007-3). Dalcimar Casanova acknowledges support from FAPESP (São Paulo Research Foundation, Brazil) (2008/57313-2) for his PhD grant.

References

Abbasi SM, Mokhtarian F, Kittler J (1997) Reliable classification of chrysanthemum leaves through curvature scale space. In: Haar RB, Florack L, Koenderink J, Viergever M (eds) Scale-space theory in computer vision. Lecture notes in computer science, vol 1252. Springer, Berlin, pp 284–295

Ash A, Ellis B, Hickey LJ, Johnson K, Wilf P, Wing S (1999) Manual of leaf architecture—morphological description and categorization of dicotyledonous and net-veined monocotyledonous angiosperms by leaf architecture. http://www.peabody.yale.edu/collections/pb/mla/mla.pdf. Accessed 12 June 2008

Backes AR, Casanova D, Bruno OM (2009a) A complex network-based approach for boundary shape analysis. Pattern Recognit 42(1):54–67

Backes AR, Casanova D, Bruno OM (2009b) Plant leaf identification based on volumetric fractal dimension. Int J Pattern Recognit Artif Intell 23:1145–1160

Bailey IW (1951) The use and abuse of anatomical data in the study of phylogeny and classification. Phytomorphology 1:67–69

Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight C (2008) Genome size is a strong predictor of cell size and stomatal density in Angiosperms. New Phytol 179:975–986

Blanco MA, Whitten WM, Penneys DS, Williams NH, Neubig KM, Endara L (2006) A simple and safe method for rapid drying of plant specimens using forced air spaces heaters. Selbyana 27:83–87

Bridson D, Forman L (1998) The herbarium handbook. Royal Botanic Gardens, Kew

Bruno OM, Plotze RO, Falvo M, Castro M (2008) Fractal dimension applied to plant identification. Inf Science 178(12):2722–2733

Camargo EA (2008) O gênero *Leandra*, seções *Carassanae*, *Niangae* e *Secundiflorae* (Melastomataceae) no Paraná. MSc thesis, Departament of Botany, Paraná Federal University

Castro-Esau K, Sánchez-Azofeifa G, Caelli T (2004) Discrimination of lianas and trees with leaf-level hyperspectral data. Remote Sens Environ 90:353–373

Dallwitz MJ (1974) A flexible computer program for generating diagnostic keys. Syst Zool 26:50–57

Dean M, Ashton PA (2008) Leaf surfaces as a taxonomic tool: the case of *Carex* section *Phacocystis* (Cyperaceae) in the British Isles. Plant Syst Evol 237:97–105

DeWolf G (1968) Notes on making an herbarium. Arnoldia 28:69–111

Dickison WC (2000) Integrative plant anatomy. Academic Press, San Diego

Durigan G, Baitello JB, Franco GADC, Siqueira MF (2004) Plantas do Cerrado Paulista: imagens de uma paisagem ameaçada. Editora Página e Letras, São Paulo

Endress PK (2003) Morphology and angiosperm systematics in the molecular era. Bot Rev 68:545–570

Ferson S, Rohlf FJ, Koehn RK (1985) Measuring shape variation of two-dimensional outlines. Syst Zool 34(1):59–68

Goldenberg R (2004) O gênero *Miconia* (Melastomataceae) no Estado do Paraná, Brasil. Acta Bot Bras 18(4):927–947

Goldenberg R, Martin C (2008) Taxonomic notes on South American *Miconia* (Melastomataceae). HPB 13:223–227

Goldenberg R, Penneys DS, Almeda F, Judd WS, Michelangeli FA (2008) Phylogeny of *Miconia* (Melastomataceae): patterns of stamen diversification in a megadiverse Neotropical genus. Int J Plant Sci 169:963–979

Gratani L, Covone F, Larcher W (2006) Leaf plasticity in response to light of three evergreen species of the Mediterranean maquis. Trees 20:549–558

Hetherington AM, Woodward FI (2003) The role of stomata in sensing and driving environmental change. Nature 424:901–908

Hlwatika CNM, Bhat RB (2002) An ecological interpretation of the difference in leaf anatomy and its plasticity in contrasting tree species in Orange Kloog, table Mountain, South Africa. Ann Bot 89:109–114

Holgren PK, Holgren N (1992) Plant specialists index. Koeltz Scientific Books, Königstein

Judd WS, Campbell CS, Kellog EA, Stevens PF, Donoghue MJ (2008) Plant systematics: a phylogenetic approach. Sinauer Associates, Sunderland

Kaplan LM (1999) Extended fractal analysis for texture classification and segmentation. IEEE Trans Image Process 8(11):1572–1585

Klucking EP (1987) Leaf venation patterns v.2: Lauraceae. J. Cramer, Berlin

Lee CL, Chen SY (2006) Classification of leaf images. Int J Imag Syst Tech 16(1):15–23

Leenhouts PW (1968) A guide to the practice of herbarium taxonomy. International Bureau for Plant Taxonomy and Nomenclature of the International Association for Plant Taxonomy, Utrecht

Mancuso S (2002) Discrimination of grapevine (*Vitis vinifera* L.) leaf shape by fractal spectrum. Vitis 41(3):137–142

Martin CV, Little DP, Goldenberg R, Michelangeli FA (2008) A phylogenetic evaluation of *Leandra* (Miconieae, Melastomataceae): a polyphyletic genus where the seed tells the story, not the petals. Cladistics 24:315–317

McLachlan GJ (1992) Discriminant analysis and statistical pattern recognition. Wiley, New York

Mendonça RC, Felfili JM, Walter BMT, Silva-Junior M, Rezende AV, Filgueiras TS, Nogueira PE, Fagg CW (2008) Flora vascular do bioma cerrado: checklist com 12.356 espécies. In: Sano SM, Almeida SP, Ribeiro JF (eds) Cerrado: Ecologia e Flora, vol 2. Embrapa Cerrados, Brasília, Brazil, pp 421–1279

Michelangeli FA, Penneys DS, Giza J, Soltis D, Hils MH, Skean J (2004) A preliminary phylogeny of the tribe Miconieae (Melastomataceae) based on nrITS sequence data and its implications on inflorescence position. Taxon 53:279–290

Mitchell T (1997) Machine learning. McGraw–Hill, New York

Mugnai S, Pandolfi C, Azzarello E, Masi E, Mancuso S (2008) *Camellia japonica* L. genotypes identified by an artificial neural network based on phyllometric and fractal parameters. Plant Syst Evol 270:95–108

Nam Y, Hwang E (2005) A shape-based retrieval scheme for leaf images. In: Ho YS, Kim H (eds) Advances in multimedia information processing. Lecture notes in computer science, vol 3767. Springer, Berlin, pp 876–887

Nam YY, Hwang EJ, Kim DY (2008) A similarity-based leaf image retrieval scheme: joining shape and venation features. Comput Vis Image Understand 110(2):245–259

Neto JC, Meyer GE, Jones DD, Samal AK (2006) Plant species identification using elliptic fourier leaf shape analysis. Comput Electron Agr 50:121–134

Pankhurst RJ (1978) Biological identification. University Park Press, Baltimore

Pearce DW, Millard S, Bray DF, Rood SB (2006) Stomatal characteristics of riparian poplar species in a semi-arid environment. Tree Physiol 26:211–218

Persson HA, Gustavsson BA (2001) The extent of clonality and genetic diversity in lingonberry (*Vaccinium vitis-idaea* L.) revealed by RAPDs and leaf-shape analysis. Mol Ecol 10:1385–1397

Plotze RO, Falvo M, Pádua JG, Bernacci LC, Vieira MLC, Oliveira GCX, Bruno OM (2005) Leaf shape analysis by the multiscale minkowski fractal dimension, a new morphometric method: a study in *Passiflora* L. (Passifloraceae). Can J Bot 83(3):287–301

Ramos E, Fernández DS (2009) Classification of leaf epidermis microphotographs using texture features. Ecol Informat 4(3): 177–181

Ramos VS, Durigan G, Franco ADC, Siqueira MF, Rodrigues RR (2008) Árvores da floresta estacional semidecidual: guia de identificaçào de espécies. Edusp, São Paulo

Rossatto DR, Kolb RM (2009) An evergreen neotropical savanna tree (*Gochnatia polymorpha*, Asteraceae) produces different dry- and wet-season leaf types. Aust J Bot 57:439–443

Salatino A, Montenegro G, Salatino MLF (1986) Microscopia eletrônica de varredura de superfícies foliares de espécies lenhosas do cerrado. Rev Bras Bot 9:117–124

Slaughter DC, Giles DK, Downey D (2008) Autonomous robotic weed control systems: a review. Comput Electron Agr 61:63–78

Souza MLDR, Baumgratz JFA (2009) *Leandra* Raddi (Melastomataceae). In: Wanderley MGL, Shepherd GJ, Melhem TS, Giulietti AM, Martins SE (eds) Flora Fanerogâmica do Estado de São Paulo, vol 6, Instituto de Botânica, FAPESP, São Paulo, pp 32–68

Wang XF, Du JX, Zhang GJ (2005) Recognition of leaf images based on shape features using a hypersphere classifier. In: Huang DS, Zhang XP, Huang GB (eds) Advances in intelligent computing. Lecture notes in computer science, vol 3644, pp 87–96

Wang X, Huang DS, Du JX, Xu H, Heutte L (2008) Classification of plant leaf images with complicated background. Appl Math Comput 205(2):916–926

Wang Z, Chi Z, Feng DD (2003) Shape based leaf image retrieval. IEE Proc Vis Image Signal Process 150(1):34–43

Zhenjiang M (2000) Zernike moment-based image shape analysis and its application. Pattern Recognit Lett 21(2):169–177