Exploring the Use of Leaf Shape Frequencies for Plant Classification

Ana Carolina Quintão Siravenha
Federal University of Para
Vale Institute of Technology
Belem, Brazil
Email: siravenha@ufpa.br

Schubert R. Carvalho
Vale Institute of Technology
Belem, Brazil
Email: schubert.carvalho@itv.org

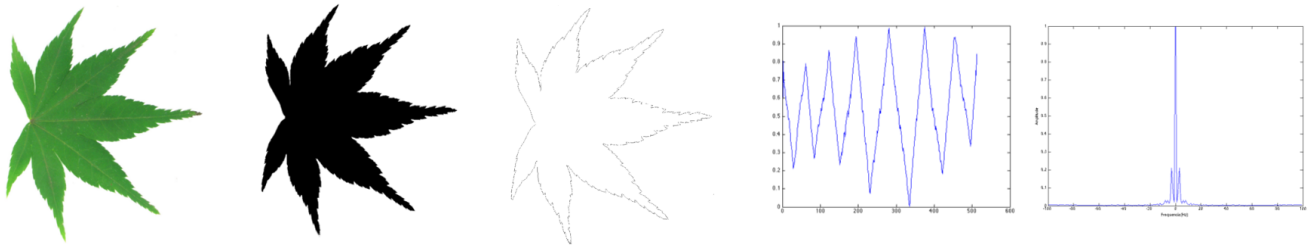


Fig. 1. Leaf shape modeling in the frequency domain. This figure shows the digital image process in combination with the Fourier descriptor for extracting a leaf shape signature. From left to right: original image; binary representation; extracted leaf contour; edge-centroid distance measure; normalised Fourier coefficients.

***Abstract*—Plant identification and classification play an important role in ecology, but the manual process is cumbersome even for experimented taxonomists. Technological advances allows the development of strategies to make these tasks easily and faster. In this context, this paper describes a methodology for plant identification and classification based on leaf shapes, that explores the discriminative power of the contour-centroid distance in the Fourier frequency domain in which some invariance (e.g. rotation and scale) are guaranteed. In addition, it is also investigated the influence of feature selection techniques regarding classification accuracy. Our results show that by combining a set of features vectors - in the principal components space - and a feedforward neural network, an accuracy of 97.45% was achieved.**

***Keywords*-Plants classification; Shape features; Fourier transform; Feature selection**

I. Introduction

In tropical regions, among all the organs that are used in plant identification, leafs are the most used ones. Mainly because they are generally present when compared to flower and fruits. For example, taxonomists can rely on leaves to search for patterns that can be used to identify a plant specie. To do so, they often analyze leaf patterns such as venations and/or shape. However, the manual process of plant identification is highly dependent on expert knowledge. In addition, it is notable the shortage of expert taxonomists, which increases the demand for new tools capable to recognize plants, for example, from images, which can be useful in field tasks [1]. In this work, the identification of an unknown plant from a given leaf database involves leaf shape modeling and classification. To give an example of how a plant specimen is represent from its leaf shape consider Figure 1. In this figure, five steps are performed until a leaf signature describing a plant specie is extracted and represented in the Fourier domain.

Contributions: This paper presents a methodology for plant identification and classification that combines the power of shape descriptors, feature selectors techniques and classification models. In our framework, shape descriptors are rotation, translation, scale and start point invariant by using the normalized Fourier transform (FT) applied over the edge-centroid distance signatures. Feature selectors are used for both reducing the data dimensionality and increasing classification accuracy. From this study, we observed several interesting results that validates and encourages the approach used in this work.

A. Related work

The literature presents many approaches directed to pattern identification for image classification and retrieving. Applications for botanical purposes have been taking advantage of these technological advances [2], [3], [4], [5]. Plants identification and classification, 3D reconstruction of leaves, or species characterization were some tasks that could be tested in real applications, through the years.

Each plant organ gives informations to identify and to characterize species. In particular, flowers, barks and leaves are the most common structures used to obtain the most representative features from a specimen. The approximately two-dimensional aspect of leaves, in contrast with flowers and fruits, makes those eligible to a plant identifying system based on pictures. Further, in most of living tropical plants, flower

and fruit are seasonal, while leaves are seen throughout all the seasons. In general, three aspects are chosen to describe the leaf: color, venation (texture) and shape.

There is several methodologies that includes leaf color as a valuable source of information [6], [7]. Often, they use color histogram to measure the similarity between two images, in association or not to the color co-occurrence matrix. The first method indicates the occurrence frequencies of colors, whilst the second gives information about the color distribution in neighboring pixels. The use of color features is susceptible to environmental conditions and is directly affected by the seasons, thus the recognition performance is limited [7].

Veins patterns are also applied for classifying and identifying purposes [8], [9]. The use of vein patterns as information of texture is based on the assumption that the surface characteristics can be represented by the color or greyscale variations of the veins. The pattern analysis can be done in the spatial or in the frequency domain, in which one can highlight the grey-level co-occurrence matrix (GLCM) [10], Wavelets [11] and local binary patterns [10], [12].

Shape analysis is probably the most usual task for humans, who can easily differ objects based on its outline [13]. For automated systems, this assumption can be applied through all the plants organs. In particular, leaves and flowers, carry the most important contour informations used to distinguish species. Besides extract vein patterns, Sun et. al [14] also obtain from a point cloud data a mesh model used to identify boundary edges of leaves with highly accuracy in a 3D model. On the other hand, Bhandarkar et. al [15] achieve leaf edge detection using binary morphology and eigen transformation. Their tests with 40 samples form 10 species achieve 67.5% of accuracy.

In [16], three techniques are used for comparing the classification of leaves performance. From the tests they concluded that the approach based on Support Vector Machine (SVM) and Binary Decision Tree achieved better results than those based on Fourier moment and probabilistic neural network. They extract twelve digital morphological features from leaves, derived from five basic features (diameter, physiological length and width, leaf area and perimeter). For them, the SVM was more accurate because of its high generalization performance without prior knowledge, and the association with binary trees leads to less number of SVM, improving the classification accuracy.

The improvement of mobile technologies allowed the development of tools to aid in plant recognition. For example, both the free mobile app Leafsnap (http://leafsnap.com) and the application presented by Wang et. al [2], use smartphones platforms to perform plants identification based on leaf properties. The methodology applied in [2] describes the shape contour by a multiscale convexity/concavity measurement scheme. The results illustrate its efficiency over the benchmark methods on the tested leaf databases with a speed improvement of more than 170 times.

Hossain and Amin [17] presented a semi-supervised method for leaf shape identification, including when the leaf if partially damaged or broken. They use several morphological features, such as eccentricity, area and perimeter to describe the leaf and achieved 91.41% of accuracy in a probabilistic neural network. The implementation proposed by Lee et. al [18] classify 1907 leaves using ten morphological features, five vein features and, finally, ten features extracted from the contour-centroid distance vector transformed by a Fast Fourier Transform (FFT). They achieve an average recognition rate of 97.19% overcoming other compared methods.

Distinguishing species by shape features can be a complex issue, mainly considering the contour of leaves have variations even in the same species. Im et al. [19] explore species recognition based in a hierarchically shapes of contours representation of species in the family of Aces (Maple). They indicate the use of statistical measures to better determine the species similarities.

It is well known that information from different sources can improve the classification rate [20], for this reason, many works describe approaches based on a diversity of features vector configurations [17], [7], [6]. For example, Kadir et. al [21] incorporated shape, vein, color, and texture features to classify leaves from Flavia dataset [22] with an average accuracy of 93.75%. Zhang, Yanne and Liang [23] combined local texture features using wavelet decomposition, co-occurrence matrix statistics and global shape features to generate the feature space of plant leaves. Even though there exist redundancies in the feature space, when classified by an SVM supervised model (one-versus-one strategy), they achieve 93.84% of accuracy and 0.165 of mean square error.

B. Technique overview

Natural shape descriptors are commonly used in classification tasks [6], [21], [24], [5]. This work presents a time-frequency decomposition of leaves contour for plants classification. The signature was extracted by the contour-centroid distance and thus decomposed by Fourier Transform (FT), which is invariant to rotation, reflection, translation, scale an contour sampling initial point variation.

As Figure 2 depicts, our method is composed of three main steps:

- **Dataset and data samples:** We use the Flavia database [22], which is a leaf dataset with 32 plant species. An amount of 1865 leaves were used during this study and the leaf shape is extracted by edge-contour distance measure.
- **Analysis techniques:** The Fourier transformed vector is normalized to promote the desired invariance, making up the shape descriptor vector. At this stage, feature selectors are used for dimensionality reduction, such as: Principal Component Analysis, Relief, Filtered subset by Gain Ratio and Spread Subsample.
- **Learning $\&$ classification:** After data analysis, the selected features are subjected to a learning processes before classification. The discriminative power of the feature vectors is analyzed through a selection of classification models: Pattern Net (feedforward neural network),

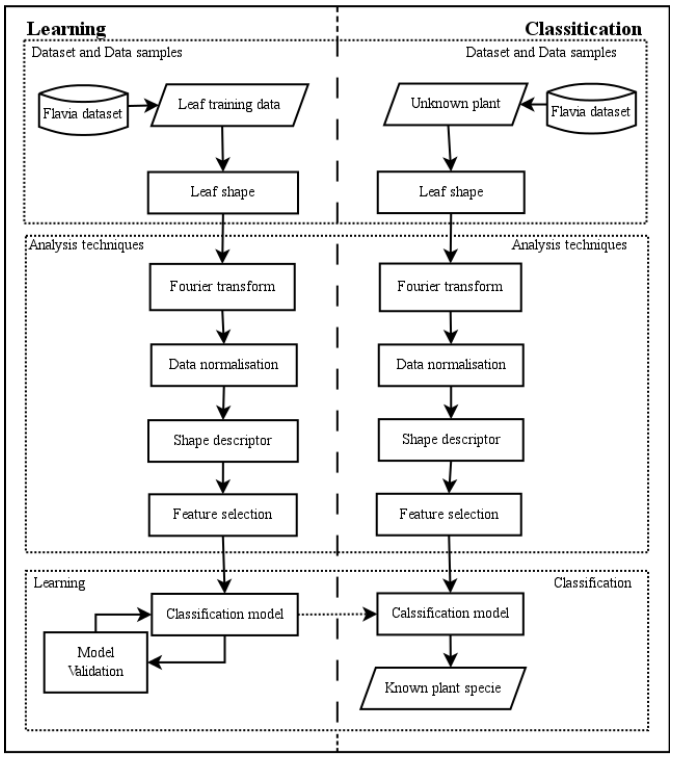


Fig. 2. Architecture of the leaf-based framework for plant classification

Random forest, Rotation forest, Bayes Network, Model trees, Naive Bayes and C4.5 decision tree (Part) .

II. Dataset and Data Samples

A. Leaves dataset

This study used experimental data from a public thirty two-classes database provided by Stephen Gang Wu and colleges [22], which is being considered as a benchmark for comparing the performance of plant classification algorithms. Figure 3 depicts a set of leaf samples for each one of the 32 classes (plants are ordered from left to right and top-down, respectively). One can note the variability in leaf's shape, as well as in rotation.

Fig. 3. Standard leaves from the Flavia dataset [22].

B. Shape signature

Some shape descriptors reduce the complexity of two-dimensional objects by transforming the problem into a one-dimensional contour parameterization. In particular the contour-centroid distance, used in this study, is based on the Euclidean Distance between each vertex and its centroid point. The signature defined in Eq. (1) is intrinsically invariant to translation by considering the information of the object centroid:

$$d(t) = \sqrt{[x(t) - x_c]^2 + [y(t) - y_c]^2}. \quad (1)$$

Where, $(x(t), y(t)), t = 0, 1, \ldots, N-1$ are the boundary co-ordinates and the centroid (x_c and y_c) are defined as:

$$x_c = \frac{1}{N}\sum_{t=0}^{N-1} x(t) \qquad y_c = \frac{1}{N}\sum_{t=0}^{N-1} y(t) \quad (2)$$

III. Analysis techniques

A. Fourier transform

The Fourier transform decomposes a signal into its sine and cosine components, mapping it to the Fourier or frequency domain. More formally, considering a periodic signal $x(t)$, there is a unique transformation function determined by $X(\omega)$ that is completely reversible and defined by:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (3)$$

where ω is the analyzed frequency. The signal representation in the Fourier domain allows to visualize features that were not observed in the original domain.

Fourier descriptors are commonly used for shape representation, mainly because of some important characteristics that simplify its usage, for example, simple derivation and normalization, compact and hierarchical coarse to fine representation, among others [25]. The unidimensional descriptor is obtained applying the Fourier transform through a signature vector, extracted by Eq. (1). The Fourier coefficients $X(\omega)$ in Eq. 3 are called Fourier descriptors.

In order to be unaffected by scale variation the descriptors are normalized by the magnitude of the first coefficient ($X(0)$). When the data is represented in a unidimensional form and with relation to the object centroid, the data is already translation invariant. The variance in rotation and start sample point causes shift or inversion in Fourier coefficients phase, with no effects in the magnitude of $X(\omega)$, then, to achieve such invariance the signal can be represented as $|X(\omega)/X(0)|$.

B. Feature selection and Dimensionality reduction

Real applications for objects representation often require a lot of features to describe them. The efficiency of Fourier shape descriptors can be improved by selecting the lower frequency coefficients, which avoid the influence of the noise caused by the signal discretization. Moreover, in order to achieve the best feature composition able to better represent

the signal and to speed up learning, one can use a set of best features by performing feature selection of the original features.

1) Principal Component Analysis: Principal Component Analysis (PCA) is an unsupervised linear dimensionality reduction technique, which seeks to map data from a high dimensional space to a low one while preserving all the relevant linear structure. It uses an orthogonal transformation in order to convert an amount of observations into a set of uncorrelated variables, so called principal components (PC). This new orthogonal coordinate system optimally describes the variance in a dataset [26]. The results of a PC analysis are presented in terms of component scores (factor scores) and loadings. The scores are the transformed variable values of a data point, while loadings are the weight by which each standardized original variable should be multiplied to get the component score [27].

2) Relief: The Relief feature selector [28] is a statistical method that avoids heuristic search, requiring linear time in number of features and the number of training instances regardless of the target concept to be learned. It is noise-tolerant and robust to feature interactions, as well as being applicable for binary or continuous data. On the other hand, it does not discriminate between redundant features

Overall, given a training data S and a relevance threshold $\tau \in [0, 1]$, the selector detects those features which are statistically relevant to the target. According to [29] Relief is only applicable when the relevance level is large for relevant features and small for irrelevant one, and when τ can be chosen to retain relevant features and discard the others.

3) Information gain and Gain ratio: In a decision tree, non-terminal nodes are represented by tests on one or more features, while terminal nodes are the output decision [30]. To select the test attribute at each node, the information gain measure (IG) is used. IG can be applied to attributes that take on a large number of distinct values, but it needs another method to determine the cut-off point, since IG only determines the sequence of features from most useful to least useful. The sequence is then defined by measuring the expected reduction in entropy.

Gain ration (GR) is a modification of the information gain that reduces its bias by taking into account the number and size of branches during choosing an attribute. Based on entropy, or intrinsic information, it defines how much information is needed to tell to which branch an instance belongs to [31]. According to definition (4), the larger the entropy the smaller the attribute gain ratio.

$$GR(attribute) = \frac{Gain(attribute)}{Entropy(attribute)}. \tag{4}$$

4) Spread subsample: This selector produces a random subsample of a database in which one class distribution is adjusted through a random elimination of objects from the majority class. This distribution can be uniform or defined as a ratio between the classes [32].

IV. Learning & Classification

After the feature extraction, the shape signature vectors are subjected to a learning stage. Artificial neural networks (ANN) [33], [34] are some of the most frequently used machine learning techniques in plant classification applications [22], [35], [36]. Other techniques, such as decision trees and its derivatives are also employed in plant classification [37], [24], [38].

To measure the accuracy of our framework seven (7) classifiers are used. The tests were proceed using both the Weka [39] and Matlab [40] softwares. The model's performance was measured according to:

$$\text{Accuracy} = \left(\frac{N_{correct}}{N_{total}}\right) \times 100. \tag{5}$$

Where, $N_{correct}$, is the number of correct classified samples and N_{total} is the number of total samples.

A. Pattern net (feedforward neural network)

A feedforward neural network is the simplest type of ANN, in which the information only moves forward, from the inputs to the output nodes, through the hidden layers. In this layered architecture, each node is connected to one or more other nodes by real-valued weights (parameters), but not to nodes in the same layer.

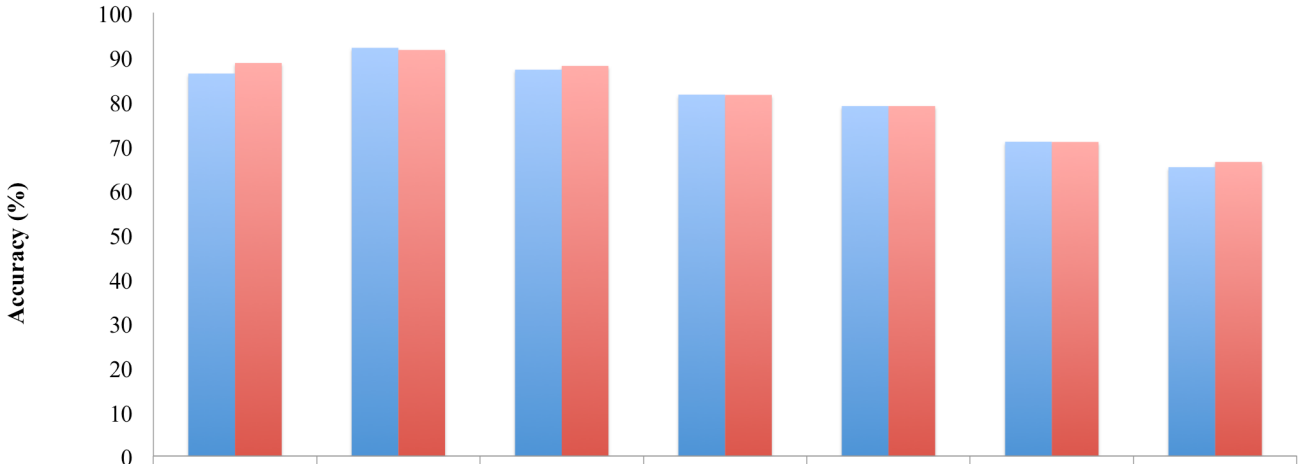
Generally, for each layer (except the output layer) it is implemented an additional node, the bias unit. This node is associated with each vector and node and constrain how input data are related to output data by weighted values determined by the iterative flow of training data through the network. The output layer is composed by a weighted summation of the hidden layer outputs. Its bias is unitary for all patterns in the dataset.

The training phase of a feedforward neural network involves adjusting the network, so that it is able to produce a specific output for each of a given set of input patterns. The supervised training model follows a sequence of steps until presenting an acceptable training error, starting by the randomly setting of weights. Then, the network is fed with the samples of the training set, and the network's output is computed. According to the output response and its distance to the expected output, the nodes weights are reset and the network is once more fed. These steps are repeated until the stop clause being reached.

B. Bayes Network

It is a statistic model that represents a set of random variables and their conditional dependencies by probabilistic relationship. Formally, it is a directed acyclic graph (DAG) whose nodes represent random variables and edges represent conditional dependencies.

Bayesian network structure has two categories of learn methods: the score-and-search-based approach and the constraint-based approach. The first one (used in this work) starts the learning from an initial structure and move to the neighbors until a local maximum of the selected criteria is reached. The movement is according to the best score



	Pattern Net	Random forest	Rotation forest	Bayes Network	Model trees	Naive Bayes	C4.5
256 coefficients	86.26	92.06	87.13	81.5	78.92	70.88	65.14
512 coefficients	88.64	91.58	87.98	81.44	78.92	70.83	66.32

Fig. 4. Measure of Fourier descriptors accuracy. This figure aims to illustrate the influence of the symmetry property of Fourier transform through different classifier models.

in the structure space generated randomly or from domain knowledge [41]. Finally, the result can be improved by re-starting the learning process with different initial structures.

The search method K2, used during the tests, assumes in a DAG the parent nodes are before children nodes in causal ordering. Since nodes ordering is known, it is possible to reduce the search space by finding the best structure to pick the best set of parents for each node, independently. If the score does not increases with the addition of a parent, it stops adding parents to the node.

C. Naive Bayes

Naive bayes classifier is a probabilistic method which apply the Bayes's theorem with strong (naive) independence assumptions between the features. Although is an relatively old methodology, with right data preprocessing, its efficiency is comparable to more sophisticated methods, as support vector machines [42]. This classifier is scalable and the number of parameters required in a learning problem is linear in the number of features. When using maximum-likelihood training, it takes linear time processing by evaluating the closed-form expression. It requires a small amount of training data to estimate the necessary parameters.

D. Random forest

It is an ensemble learning method that construct a multitude of decision trees for classification, regression, among others. The decision trees are created at training time and outputting a class as the classes mode of the individual trees. Random forests reduces overfitting and can handle thousands of input variables without variable deletion giving estimative of what variables are important in the classification [43]. The training set is defined by sampling with replacement, where about one-third of the cases are left out of the sample. These cases are used to get a running unbiased estimate of the classification error when new trees are added to the forest. The data run down through all built tree, and by the proximity measure is increased when two cases occupy the same terminal node. The proximities are normalized by the number of trees and are used to replace missing data e locate outliers.

E. Rotation forest

This classifier is another ensemble learning method that uses independently trained decision trees. In this case, each tree is trained with the whole data set in a rotated feature space producing different trees. To build the training set, the data is split into k subsets from which are extracted the principal components that are all retained in order to preserve the variability information in the data. Each k subset is an axis of rotation that take place to form the new features for a base classifier. Then, all principal components and whole dataset are used to train each classifier [44].

F. Model trees

Model trees are decision trees with linear regression functions at the leaf nodes [45]. First of all it is created an ordinary tree, pruned back by replacing subtrees with linear regression functions. All the attributes present in nodes of a subtree pruned away are variables used during the regression. The learning procedure uses the decision tree to divide the instance space into regions in order to minimize the mean square error between the model's tree and the target.

G. C4.5 (decision tree)

C4.5 is an extension of Iterative Dichotomiser 3 (ID3) algorithm, being both statistical classifiers formulated under decision trees assumptions [46]. It can handle both continuous and discrete attributes, including missing values, that are not

used in gain and entropy calculations. It prunes trees after creation, going back through the tree removing branches that do not increase the accuracy, replacing each removed branch with leaf nodes. Building trees using the concept of information entropy, C4.5 splits branches according differences in entropy in which the highest normalized value is chosen to make a decision.

V. Experiments

To demonstrate the effectiveness of the proposed methodology, we tested seven classifiers and four feature selection techniques. In the first test, it was investigated how the symmetry property of Fourier Transform influences the recognition rate. It is well known that the FT of a real even signal is real and even, and since arbitrary real signals are always a sum of an even and an odd function, the FT of a real signal has an even real part and an odd imaginary part, and the amplitude is thus always symmetrical [47]. Based on these assumptions, two feature vectors were tested. It was chose a Fourier transform with 512 coefficients, being, according to the symmetry property, the amplitude of the first 256 coefficients symmetrical to the other half of points.

The second test aimed to evaluate the influence of seven feature selection techniques and one dimensionality reduction such as the PCA. The goal of this experiment was to identify the most important source of information, which contributes to increase classification accuracy.

In all tests, it was employed a k-fold cross-validation technique, that randomly splits the original tested data into k equal size subsamples from which a single partitioning is retained as the validation data for testing the model. The training data is composed by the remaining subsamples. This process is repeated k times (folds) and the results are averaged in order to produce a singe model accuracy estimation [48]. All results presented in this work is the average value of $k = 10$. This number is normally used for model validation, so we used this number. It is important to say that in all tests with pattern network used an architecture with two hidden layers with 26 neurons each, trained during 100 epochs.

VI. Results and Discussion

Figure 4 shows the models' accuracy tested for the two feature vectors with 512 and 256 coefficients. At this stage, neither feature selection nor dimensionality reduction was performed. The highest accuracy was obtained with the random forest classifier, followed by rotation forest and pattern network. For the first model, the 256 in length vector was 0.48% more accurate than the 512 feature vector.On the other hand, the 512 feature vector produced more accurate results (in comparison with the 256 long one) only when classified by pattern network, rotation forest and C4.5, leading to the assumption it worth nothing, with respect to the classification rates, use the entire vector.

When the data was reduced by the methods described in Section III-B, the results were those pictured in Figure 5. In this experiment, it was used the first 256 Fourier descriptors. The PCA transformed data, taking 95% of variance, reduced the feature vector to 131 coefficients. Relief, GR and Spread subsample selected 126, 97 and 66 coefficients, respectively. On average, the results of PCA coefficients was more accurate through the classifier models (89.49%). Its average was also larger than the majority results obtained by original vectors. Individually, PCA was also more accurate than all the other tests when applied to a pattern network.

The biggest improvement, after feature selection, was using the PCA vector into Naive Bayes classifier (24.73% larger than the result of the original vector). In opposition, when using the vector selected by gain ratio into a pattern network the accuracy was approximately 4.5% lower than using the original one.

By comping the accuracy of selected vector with the original vector with 512 coefficients, the classification effectiveness increases at a lower rate. For PCA the average raise of well classification was by 8.68 percentage points. Otherwise, when using gain ratio the efficiency was reduced by 2.35%. In particular, when classified by C4.5 tree, the accuracy was reduced by 5.09%. The best individual raise of accuracy, when compared with 256 vector was using the Naive Bayes model with the PCA vector, from 70.88% to 88.41%. In comparison with other methods, this proposal is in consonance with the state-of-the-art methods in shape recognition (see Table I), achieving better results, including when non-optimal scenarios are in comparison.

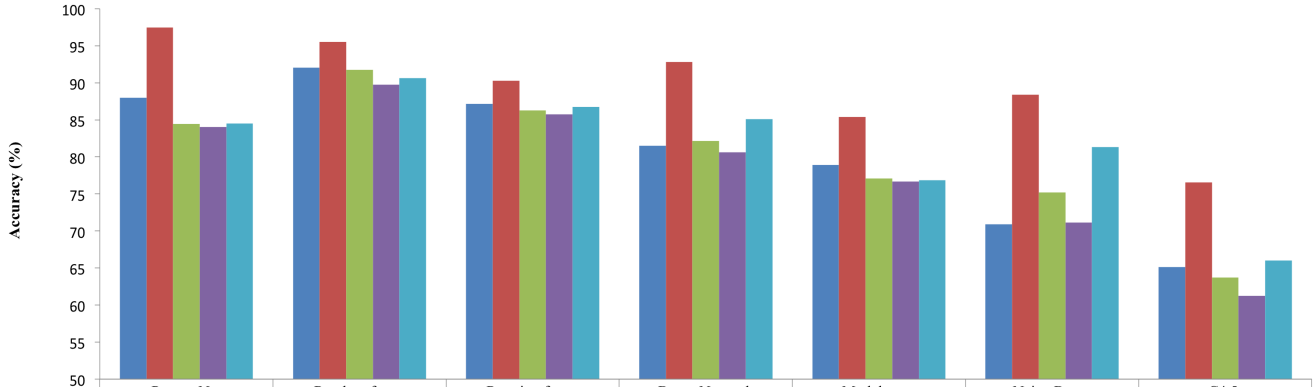
TABLE I
Accuracy comparison with state-of-the-art in shape recognition methods.

Proposal	Accuracy (%)
Wu et al.(2007) [22]	90.3
Hossain and Amin(2010) [17]	91.41
Kadir et al.(2011) [21]	93.75
Zang et al.(2012) [23]	93.8
Singh et al.(2010) [16]	96
Lee et al.(2013) [18]	97.19
PCA selected vector	97.45

The table includes methods which use not only shape features, but also texture and color combined to form the feature vector, besides other classification models, as probabilistic neural network, for example.

VII. Conclusion

In this work, we explored the use of leaf shape frequencies obtained from the Fourier transform for automatic plant identification and classification. The performance of the studied method has been evaluated on the well-known Flavia dataset by using two feature vectors with 256 and 512 coefficients. It was also analyzed the use of four feature selection and one dimensionality reduction techniques, applied over the smaller feature vector. Finally, the results was also tested through seven classifier models. When no feature selection or reduction was used, the 256 long vector classified by random forest presented the best accuracy results. On the other hand, the



	Pattern Net	Random forest	Rotation forest	Bayes Network	Model trees	Naive Bayes	C4.5
256 coefficients	88	92.06	87.13	81.5	78.92	70.88	65.14
PCA	97.45	95.54	90.29	92.81	85.41	88.41	76.56
Relief	84.44	91.74	86.27	82.14	77.1	75.17	63.69
GR	84.04	89.75	85.73	80.64	76.67	71.15	61.23
Spread subsample	84.5	90.61	86.75	85.09	76.83	81.34	66

Fig. 5. Measure of accuracy of selected feature vectors. The feature selectors where applied over 256 Fourier descriptors.

pattern network achieved the highest accuracy with the vectors in the PCA space.

ACKNOWLEDGMENT

Authors would like to thank the funding from the call 59/2013 MCTI/CT-Info/CNPq (process number 440880/2013-0).

REFERENCES

[1] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, "Plant species identification using digital morphometrics: A review," *Expert Systems with Applications*, pp. 7562–7573, 2012.

[2] B. Wang, D. Brown, Y. Gao, and J. L. Salle, "Mobile plant leaf identification using smart-phones," in *20th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2013, pp. 4417 – 4421.

[3] T. Santos and J. Ueda, "Automatic 3d plant reconstruction from photographies, segmentation and classification of leaves and internodes using clustering," in *International Conference on Functional-Structural Plant Models*, 2013, pp. 95–97.

[4] R. Binsfeld, J. Gamboa, and M. Walter, "Visual patterns in the plant kingdom," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2011.

[5] L. A. Loss and C. L. Tozzi, "Discrimination of natural contours by means of time-scale-frequency decompositions," in *Advances in Visual Computing*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, vol. 3804, pp. 684–689.

[6] H. Kebapci, B. Yanikoglu, and G. Unal, "Plant image retrieval using color,shape and texture features," *The Computer Jornal*, 2010.

[7] B. S. Anami, S. S. Nandyal, and A. Govardhan, "A combined color, texture and edge features based approach for identification and classification of indian medicinal plants," *International Journal of Computer Applications*, vol. 6(1), 2010.

[8] V. Metre and J. Ghorpade, "An overview of the research on texture based plant leaf classification," *CoRR*, vol. abs/1306.4345, 2013.

[9] A. N. Hussein and M. I. S. S. Mashohor, "A texture-based approach for content based image retrieval system for plant leaves images," in *IEEE 7th International Colloquium on Signal Processing and its Applications*, 2011, pp. 11–14.

[10] Y. Ben Salem and S. Nasri, "Rotation invariant texture classification using suport vector machines," in *International Conference on Communications, Computing and Control Applications (CCCA)*, 2011, pp. 1–6.

[11] Y. Shang, T. An, and Z. Meng, "A study on methods for rotation invariant image recognition based on texture characteristic," in *International Conference on Computer Application and System Modeling (ICCASM 2010)*, 2010, pp. 568–571.

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

[13] G. C-Y and W. C-H., "Plant identification through images: Using feature extraction of key points on leaf contours," *Applications in Plant Sciences*, vol. 1(11), 2013.

[14] Z. Sun, S. Lu, X. Guo, and Y. Tian, "Leaf vein and contour extraction from point cloud data," in *International Conference on Virtual Reality and Visualization*, 2011, pp. 11–16.

[15] P. Bhandarkar, H. Doshi, M. Kadakia, and R. Ahmed, "Leaf identification using morphology and structural decomposition," in *International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2014, pp. 125–129.

[16] K. Singh, I. Gupta, and S. Gupta, "SVM-BDT PNN and Fourier moment technique for classification of leaf shape," *International Journal of Signal Processing Image Processing and Pattern Recognition*, vol. 3, no. 4, pp. 67–78, 2010.

[17] J. Hossain and M. A. Amin, "Leaf shape identification based plant biometrics," in *International Conference on Computer and Information Technology (ICCIT 2010)*, 2010, pp. 458–463.

[18] K.-B. Lee and K.-S. Hong, "An implementation of leaf recognition system based on leaf contour and centroid for plant classification," in *International Journal of Bio-Science and Bio-Technology*, vol. 5 (2), 2013, pp. 57–66.

[19] C. Im, H. Nishida, and T. Kunii, "Recognizing plant species by leaf shapes-a case study of the acer family," in *International Conference on Pattern Recognition-Volume 2 - Volume 2*, ser. ICPR '98. IEEE Computer Society, 1998.

[20] A. Mansano, J. Matsuoka, L. Afonso, J. Papa, F. Faria, and R. Torres, "Improving image classification through descriptor combination," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2012, pp. 324 – 329.

[21] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, "Leaf classification using shape, color, and texture features," *International Journal of Computer Trends and Technology*, pp. 225–230, 2011.

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

[23] H. Zhang, P. Yanne, and S. Liang, "Plant species classification using leaf shape and texture," in *International Conference on Industrial Control and Electronics Engineering*, 2012, pp. 2025–2028.

[24] H. Liu, D. Coquin, L. Valet, and G. Cerutti, "Leaf species classification based on a botanical shape sub-classifier strategy," in *International Conference on Pattern Recognition (ICPR)*, 2014, pp. 1496–1501.

[25] D. Zhang and G. Lu, "Shape based image retrieval using generic fourier descriptors," *Signal Processing Image Communication*, 2002.

[26] C. Boutsidis, M. W. Mahoney, and P. Drineas, "Unsupervised feature selection for principal components analysis," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2008, pp. 61–69.

[27] P. Shaw, *Multivariate statistics for the Environmental Sciences*. Hodder-Arnold, 2003.

[28] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the Ninth International Workshop on Machine Learning*, 1992, pp. 249–256.

[29] ——, "The feature selection problem: Traditional methods and a new algorithm," in *National Conference on Artificial Intelligence - AAAI*, 1992.

[30] A. G. Karegowda, A. S. Manjunath, and M.A.Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *International Journal of Information Technology and Knowledge Management*, vol. 2(2), 2010.

[31] R. P. Priyadarsini, M.L.Valarmathi, and S. Sivakumari, "Gain ratio based feature selection method for privacy preservation," *ICTACT Journal on soft computing*, vol. 01 (04), 20011.

[32] O. Loyola-Gonzlez, M. Garca-Borroto, M. A. Medina-Prez, J. F. Martnez-Trinidad, J. A. Carrasco-Ochoa, and G. Ita, "An empirical study of oversampling and undersampling methods for lcmine an emerging pattern based classifier," ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 7914, pp. 264–273.

[33] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1995.

[34] E.-P. M., de Ridder D., and H. H., "Image processing with neural networksa review," *Pattern Recognition*, vol. 35, no. 10, pp. 2279 – 2301, 2002.

[35] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, "Neural network application on foliage plant identification," *International Journal of Computer Applications*, no. 9, pp. 15–22, 2011.

[36] Z. Zulkifli, P. Saad, and I. Mohtar, "Plant leaf identification using moment invariants and general regression neural network," in *2011 11th International Conference on Hybrid Intelligent Systems (HIS)*, 2011, pp. 430 – 435.

[37] B. Pant, K. Pant, and K. Pardasani, "Decision tree classifier for classification of plant and animal micro rna's," in *Computational Intelligence and Intelligent Systems*, ser. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2009, vol. 51, pp. 443–451.

[38] J. Fan, J. Peng, L. Gao, and N. Zhou, "Hierarchical learning of tree classifiers for large-scale plant species identification," in *International Conference on Semantic Computing (ICSC)*, 2015, pp. 389 – 396.

[39] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[40] MathWorks. Neural network toolbox.

[41] D. Chickering, "Optimal structure identification with greedy search," *Journal of Machine Learning Research*, vol. 3, pp. 507–554, 2002.

[42] J.Rennie, L. Shih, J. Teevan, and D. Karger, "Tackling the poor assumptions of naive bayes classifiers," in *International Conference on Machine Learning*, 2003.

[43] L. Breiman, "Random forests," pp. 5–32, 2001.

[44] J. Rodriguez, L. Kuncheva, and C. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1619–1630, 2006.

[45] E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. Witten, "Using model trees for classification," *Machine Learning*, vol. 32, no. 1, pp. 63–76, 1998.

[46] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[47] R. Bracewell, *The Fourier Transform and Its Applications*. McGraw-Hill Science/Engineering/Math, 1999.

[48] R. Picard and D. Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, vol. 79 (387), p. 575583, 1984.