Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 1, January 2017, pg.92 – 99

Leaf Recognition based on Neural Network Feed-Forward and Support Vector Machine Classifiers

Abbas H. Hassin AlAsadi [1], Eman Qais Anduljalil [2], Amal Hameed Khaleel [3]

[1]Faculty of Computer Science and Information Technology, Basrah University, Iraq.
[2]Faculty of Computer Science and Information Technology, Basrah University, Iraq.
[3]Faculty of Computer Science and Information Technology, Basrah University, Iraq.
[1] abbas.hassin@uobasrah.edu.iq; [2] iman.abduljaleel@uobasrah.edu.iq; [3] amal.khaleel@uobasrah.edu.iq

Abstract— ***Leaf recognition is very important in plant classification and its key subject to distinguish different kinds of leaves. In this paper, A computer aided leaf recognition system is suggested. The main steps involved in this work are image preprocessing, edge detection, feature extraction and the classification. Preprocessing step consists of four sub-steps; color converting, thresholding, binary operations, and filtering (segmentation). Edge detection is strongly needed for locating features. Whereas, feature extraction depends on three types of features; shape, and texture. Finally, two approached of classifiers are used; Neural Network feed-forward and Support Vector Machine. The experimental results are applied to two different datasets. According to the results, the classifiers succeeds in getting accuracy of 91%. The results are acceptable.***

Keywords— ***Plant Identification; Digital Morphological Feature; Support Vector Machine; Features Extraction***

I. Introduction

Plants species plays a major role in our life. Plants are an essential resource for the development of human society human welfare. Plants are of plenty of use in several areas, such as food chain, medical science, industry, and extremely important for environmental protection. So, the study of plants classification is extremely needed.

Leaf recognition is very important in plant classification and its key subject to distinguish different kinds of leaves. Comparing with other methods, such as cell and molecule biology methods, classification based on leaf image is the first choice for leaf plant classification. Someone can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques.

Computer aided leaf recognition is still very challenging task in computer vision because of improper models and inefficient representation approaches. Therefore, the main aim of leaf recognition is to evaluate and extract the morphological features of the leaf. The extraction features from a leaf is a key step in the plant recognition process.

Plant leaves are approximately two-dimensional in nature and the shape of plant leaf is one of the most important features for characterizing various plants species. [1] Therefore, it is necessary to develop an easy and automatic method that can correctly identify and recognize leaf shapes of different species. Most of the existing plant recognition methods are based on both the global shape feature and the whole plant leaves [2]. The effectiveness of a shape-based image retrieval system depends on the types of shape representation used, the types of queries allowed, and the efficiency of the shape matching techniques implemented [3].

Fig. 1 presents the various external leaf parts and a description of each of these parts is given in Table 1.

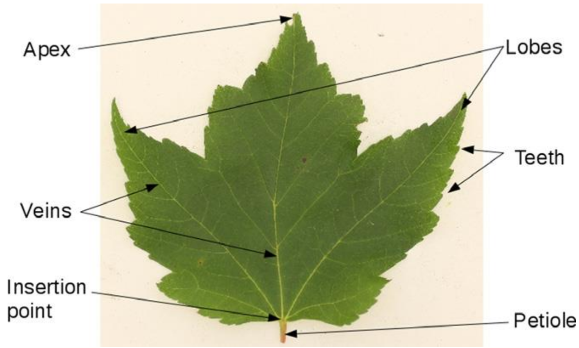


Fig. 1 Structure of Leaf [4].

TABLE 1

THE VARIOUS EXTERNAL LEAF PARTS AND THEIR DESCRIPTION

Leaf's part	Description
Apex	Leaf tip or the outer end of the leaf. It is the point farthest from the point of attachment.
Base	Leaf part that connects to the stem (petiole)
Teeth or Margin	Edges of leaves
Vein	Vascular tissue of the leaf, located in the spongy layer of the mesophyll.
Venation	The pattern of the veins is called venation, and is typically characterized by hierarchical structures with abundant closed loops.
Lobe	Part of a leaf, often rounded, formed by incisions to about halfway to the midrib.
Petiole	The stalk of a leaf.

The rest of this paper is arranged as follows: Section 2 surveys the different approaches and methodologies related with the goal of this paper. Current Methodology is covered in Section 3, whereas Section 4 is devoted to proposed algorithms. Experimental results are given in Section 5. Section 6 completes the conclusions with some perspectives and future work on improving our algorithm.

II. RELATED WORKS

Several approaches have been introduced to classify a leaf, such as k-Nearest Neighbor Classifier (k-NN), Neural Network (NN), Genetic Algorithm (GA), Support Vector Machine (SVM), and Principal Component Analysis (PCA). Some of researchers used green color leaves, others ignored color information on leaves [5].

For instance, Pérez et.al [3] used the color and shape feature of leaf image to discriminate soil, weed, and crops. K-Nearest Neighbor (KNN), Bayes rule and heuristic approaches are used in classifying the leaf image.

A new classification method called Move Median Centres (MMC) hyper sphere classifier is proposed [6]. From the experimental results of this paper, the methodology save both storage space and reduces the classification time.

Ehsanirad et al. [7] proposed a method to extract the texture feature of the leaf image and classification. Two different algorithms namely Principal Component Analysis Algorithms (PCA) and Gray Level Co-occurrence Matrix (GLCM) are used to achieve an accuracy of 78% in extracting the texture feature. While, Propagation Neural Network (PNN) for classifying the plants with broad flat leaves is used [8, 9].

Actually, many of suggested algorithms for identification using multiclass classification based on color, texture, vein, shape of the leaves are performed [1, 9-12].

III.PROPOSED LEAF RECOGNITION SYSTEM

The phase structure of the proposed system of leaf recognition is essentially consists of two stages: the training and the testing stages. Each stage has specific functions; all functions have explained in detail via the following subsections. The testing phase is the same of training phase, but, in the testing phase, the features do not store in database just entered to one of two classifiers separately. Fig. 2 describes the block diagrams of the proposed systems.

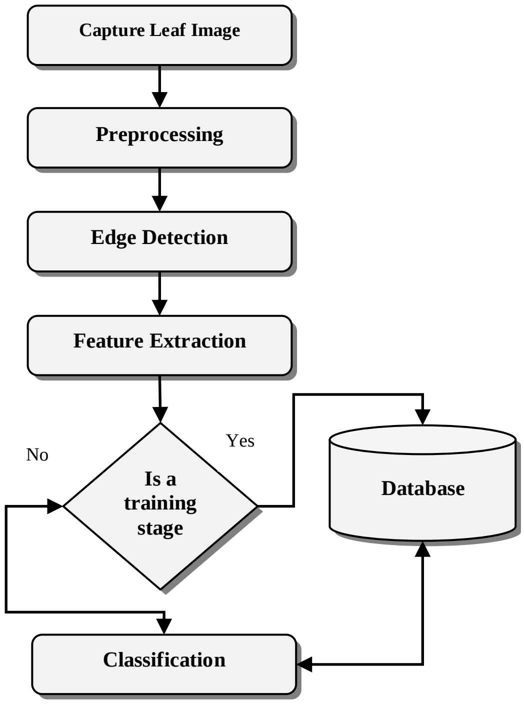


Fig. 2 The main block diagram of proposed system.

3.1 Leaf Image Capturing

In this paper, the leaves images were collected by two procedures; (1) manually, captured 10 leaves from each 10 different species. (2) The dataset of leaf-snap, which currently contains 188 tree species from the Northeastern United States. Ten leaves images from each 10 species are used separately in the testing stage. All images are in the JPEG format, with different resolution of pixels, which were later resized to 256 x 256 pixels. Fig. 3 displays a sample of collected datasets.

Fig. 3 Samples of collected datasets

3.2 Preprocessing

Depending on the capturing device, leaf images are difficult to proceed to the next step without improving and enhancing. As a result, it is necessary to improve the quality of the image and makes the feature extraction phase easier and more reliable. The main objective of this phase is to enhance and suppress the undesired distortion of leaf images. In this paper, preprocessing step is already split into five substeps;

1) RGB to HSV Conversion: In such situations where color description plays an integral role, the HSV color model is preferred over the RGB model. The HSV model describes colors similarly to how the human eye tends to perceive color. HSV describes color using more familiar comparisons such as color, vibrancy and brightness[13].

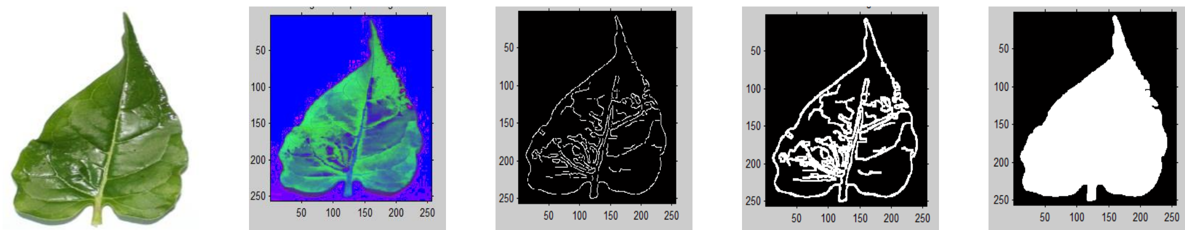
2) Threshold: According to the leaf histogram, there are two peaks; one for the leaf and other for the background. The main purpose of taking the blue band was to increase the distance between the two peaks, so that the thresholding can be done easily. Moreover, there would be no overlap between the pixels corresponding to the two peaks. A value between the two peaks has been taken as a threshold for the gray scale (blue band) image.

3) *Edge Detection :*Edge detection reduces the amount of image data while maintaining the structural properties [14]. In this paper, Applying Canny edge detection method to the leaf gray scale images to extraction the borders of the image and Remove the background edges keeping only leaf edge details.

*4) Filtering :*After binary conversion, the output image contains some amount of noise and therefore, enhancement techniques that reduce its effect are always desirable. To remove this noise, the standard median filtering (5x5 window) is applied. If the number of white pixels in a window is greater than the number of black pixels, the mid-pixel is changed to white; and vice versa.

5) Binary Conversion: Taking the threshold as a level to separate the background from the leaf, the values less than threshold is taken as white and the values greater is taken as black. The leaf image is now white color and the background is black color.

Fig. 4 shows the results of all sub-steps of image preprocessing.



Input image RGB to HSV Edge detection Filtering Binary Image

Fig. 4 The results of all sub-steps of image preprocessing.

3.4 Feature Extraction

An effective leaf recognition system requires a set of leaf features that best define the leaf image and which can provide maximum difference between different leaves. For that reason, the quality of a feature vector is related to its ability to recognize examples from different classes.

1) Shape features: Shape features on the basis of morphological features are defined as [15] :

A. Geometrical features

1. *Diameter of leaf* :The diameter is the longest distance between any two points on the margin of the leaf. It is represented as D.
2. *Length of leaf*: The distance between the two terminals of the main vein is the physiological length. It is represented as L.
3. *Leaf Area*: The value of leaf area is obtained by Counting the number of pixels of binary 1s on smoothed leaf image. Leaf area is denoted as A.
4. *Leaf Perimeter*: leaf perimeter is obtained by counting the number of pixels present at the leaf margin. It is represented as P.
5. *Smooth factor*: Smooth factor is the ratio between area of leaf image smoothed by 5×5 rectangular averaging filter and the one smoothed by 3×3 rectangular Averaging filter.
6. *Circularity*: This feature illustrates the difference between a leaf and a circle. It is defined perimeter of the leaf margin.
7. *Rectangularity*: Rectangularity illustrates the similarity between a leaf and a rectangle.

8. *Ratio of diameter to length* : Narrow factor is the ratio of the diameter D and length L, thus D/L.

9. *Ratio of Perimeter to diameter*: Ratio of perimeter to diameter, denoting the ratio of leaf perimeter P and leaf diameter D, thus P/D.

10. *Ratio of Perimeter to length*: This feature is the ratio of leaf perimeter P and length L thus P/L.

B. Digital Morphological Features

Based on 5 basic features introduced previously, we can define 12 digital morphological features used for leaf recognition.

1. *Smooth factor*: We use the effect of noises to image area to describe the smoothness of leaf image. In this paper, smooth factor is defined as the ratio between area of leaf image smoothed by 5 × 5 rectangular averaging filter and the one smoothed by 2 × 2 rectangular averaging filter.

2. *Aspect ratio*: The aspect ratio is defined as the ratio of physiological length Lp to physiological width Wp, thus Lp/Wp.

3. *Form factor*: This feature is used to describe the difference between a leaf and a circle. It is defined as 4_A/P2, where A is the leaf area and P is the perimeter of the leaf margin.

4. *Rectangularity*: Rectangularity describes the similarity between a leaf and a rectangle. It is defined as LpWp/A, where Lp is the physiological length, Wp is the physiological width and A is the leaf area.

5. *Narrow factor*: Narrow factor is defined as the ratio of the diameter D and physiological length Lp, thus D/Lp.

6. *Perimeter ratio of diameter*: Ratio of perimeter to diameter, representing the ratio of leaf perimeter P and leaf diameter D, is calculated by P/D.

7. *Perimeter ratio of physiological length and physiological width*: This feature is defined as the ratio of leaf perimeter P and the sum of physiological length Lp and physiological width Wp, thus P/(Lp +Wp).

C. Moment Invariants

Moment invariants have been used as feature extractions in variety of object recognition applications during the last 40 years [16]. The moment invariants are used to evaluate seven distributed parameters of a numeral image. Moment invariants are work under translation, rotation, scaling and reflection [17]. Regular moments are defined as mentioned in [18].

2) Texture Features :Naturally, the most well-known statistical tools for extracting texture information from images is the Gray Level Co-occurrence Matrix (GLCM). In this paper, five texture features of leaf image, called feature vector, could be directly computed from the GLCM. These measures which based-on the statistical features are described below [19, 20]:

1. Angular Second Moment (ASM) $= \sum_{i,j=0}^{N} P(i,j)^2$,

2. Contrast $= \sum_{i}\sum_{j}(i-j)^2 P(i,j)$,

3. Inverse Different Moment (IDM) $= \sum_{i}\sum_{j}\frac{1}{1+(i-j)^2}P(i,j)$,

4. Entropy $= \sum_{i}\sum_{j}P(i,j)\log P(i,j)$

5. Correlation $= \dfrac{\left\{\sum_{i}\sum_{j}(i,j))P(k,j)\right\} - \mu_x\mu_y}{\sigma_x\sigma_y}$.

3.5. Classification

Together with feature extraction, the most crucial phase in the process of leaf recognition is the classification. In this paper, two approaches are used to classify our dataset as follows:

1) Feed-Forward Neural Network: A Feed-Forward Neural Network (newff) is a type of neural network architecture where the connections are "fed forward", i.e. do not form cycles (like in recurrent nets). The term "Feed forward" is also used when you input something at the input layer and it travels from input to hidden and from hidden to output layer [21].

2) Support Vector Machines : Support Vector Machines (SVM) is a machine essentially used for the purpose of classification, it consists of associated supervised learning techniques. In order to increase the boundary between two data sets SVM generates a two parallel hyper plane [22].

A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one "target value" (i.e. the class labels) and several "attributes" (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

IV. EVALUATION METHODS

Performance evaluations of the proposed system calculated in terms of sensitivity, specificity, and accuracy. The three terms defined as follows [23]:

- *Sensitivity* (also called the true positive rate or the recall in some fields) measures the proportion of positives that are correctly identified; the result indicates positively. It is calculated through the following equations:

$$Sensitivity = \frac{TP}{(TP+FN)} * 100\% \qquad (1)$$

- *Specificity* (also called the true negative rate) measures the proportion of negatives that are correctly identified; the result indicates negatively (non-disease). It is calculated through the following equations:

$$Specificity = \frac{TN}{(TN+FP)} * 100\% \qquad (2)$$

- *Accuracy* measures the probability that the diagnostic test is performed correctly. It is calculated through the following equations:

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} * 100\% \qquad (3)$$

Where TP (True Positives) correctly classified positive cases; TN (True Negative) correctly classified negative cases, FP (False Positives) incorrectly classified negative cases, and FN (False Negative) incorrectly classified positive cases.

The higher the sensitivity and specificity, the more accurate the classifier is.

V. EXPERIMENTAL RESULTS

5.1. Configurations

The framework of project work is designed by MATLAB on a PC with the following configurations; 2.50 GHZ core i5 processor, 4 GB of RAM, run under Microsoft Windows 8_64 bits.

5.2. Testing Results

As mentioned previously, the input data of the our method is a leaf image, which collected by two ways. (1) manually, captured 10 leaves from each 10 different species. (2) The dataset of leaf-snap, which currently contains 188 tree species from the Northeastern United States.

For assuring, we have applied our method over different types and different number of features. Table 2 shows the overall results of the testing stage through two classifiers.

From Table 2, we can notice that, the results in row one with SVM classifier realizes 97.57% as accuracy.

TABLE 2

THE RESULTS OF TESTING WITH NEWFF AND SVM METHODS

I	Features	Classifiers	Accuracy	Dataset	Training	Testing
1	Invariant Moments(7), Area, Perimeter, Centroid, Equvidiamete, Roundness, Bounding Box, Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, RMS, IDM Variance, Smoothness, Kurtosis, Skewness,	Newff	92.80%	1980	1320	660
		SVM	97.57%	1980	1320	660
2	Invariant Moments(7), color feature, Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness.	Newff	88.65%	1980	1320	660
		SVM	93.08%	1980	1320	660
3	color feature, Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness.	Newff	56.65%	1980	1320	660
		SVM	67.77%	1980	1320	660
4	Invariant Moments(7), Area, Perimeter, Centroid, Roundness, Equvidiameter, Bounding Box	Newff	90.66%	1980	1320	660
		SVM	96.17%	1980	1320	660
5	Invariant Moments(7), color feature	Newff	66.12%	1980	1320	660
		SVM	75.65 %	1980	1320	660

Table 3 shows performance evaluation results of experimental results.

TABLE 3

PERFORMANCE PARAMETERS OF TESTING.

Performance Parameters	Performance%
True Positive rate	97.57
False Positive rate	2.43
True Negative rate	97.57
False Negative rate	2.43
Sensitivity	97.57
Specificity	97.57
Accuracy	97.57

5.3. Proposed System vs. Related Work Systems

The implementation of proposed system has illustrated by different types of features and compared with other related system as in Table 4.

TABLE 4

PROPOSED SYSTEM VS. RELATED WORK SYSTEMS

Supervisors	Accuracy
Wu, Stephen Gang, et al. [24].	90.31%
Singh et al. [25].	81.50%
Proposed System	97.57%

VI. CONCLUSIONS

In the experimental results, the proposed leaf recognition system showed a performance of 97.57%. From the experimental results, we can confirm that the recognition rate of the proposed advanced leaf recognition system was better than that of the existing leaf recognition system.

In future work, we improve the proposed system and further improve its recognition performance. In addition, we are continuing to research to find a correct leaf contour extraction method in the complex background.

REFERENCES

[1] Beghin, Thibaut, et al. "Shape and texture based plant leaf classification." International Conference on Advanced Concepts for Intelligent Vision Systems. Springer Berlin Heidelberg, pp. 345-353, 2010.

[2] Perez, A. J., et al. "Colour and shape analysis techniques for weed detection in cereal fields." Computers and electronics in agriculture, 25.3 (2000): 197-212.

[3] H. Kebabci, B. Yanikoglu and G. Unal, "Plant Image Retrieval using Color, Shape, & Texture Features", The Computer Journal Advance Access, 53.1 (2010):1-16.

[4] Cope, J.S., Corney, D., Clark, J.Y., Remagnino, P. and Wilkin, P. "Plant Species Identification using Digital Morphometrics: a Review", Expert Systems with Applications, 39.8 (2012): 7562–7573.

[5] Sharma, Sapna, and Chitvan Gupta. "A Review of Plant Recognition Methods and Algorithms." International Journal of Innovative Research in Advanced Engineering 6.2 (2015):111-116.

[6] Du, Ji-Xiang, Xiao-Feng Wang, and Guo-Jun Zhang. "Leaf shape based plant species recognition." Applied mathematics and computation 185.2 (2007): 883-893.

[7] Ehsanirad, Abdolvahab, and Y. H. Sharath Kumar. "Leaf recognition for plant classification using GLCM and PCA methods." Oriental Journal of Computer Science and Technology 3.1 (2010): 31-36.

[8] Javed Hossain, M. Ashraful Amin, "Leaf Shape Identification Based Plant Biometrics", Proceedings of 13th International Conference on Computer and Information Technology (ICCIT), 2010, pp. 458-463.

[9] Kulkarni, A. H., et al. "A leaf recognition technique for plant classification using RBPNN and Zernike moments." International Journal of Advanced Research in Computer and Communication Engineering 2.1 (2013): 984-988.

[10] Pavan Kumar Mishra, Sanjay Kumar Maurya, Ravindra Kumar Singh, Arun Kumar Misra, "A semi-automatic plant identification based on digital leaf and flower Images", International Conference On Advances In Engineering, Science And Management (ICAESM), 2012, pp. 68-73.

[11] Satti, Vijay, Anshul Satya, and Shanu Sharma. "An automatic leaf recognition system for plant identification using machine vision technology." International Journal of Engineering Science and Technology 5.4 (2013): 874.

[12] Pallavi, P., and V. S. Veena Devi. "Leaf recognition based on feature extraction and zernike moments." International Conference On Advances in Computer & Communication Engineering (ACCE). 2014., (pp. 67-73).

[13] Gonzalez, R. C., and Woods, R. E., "Digital Image Processing", 3rd ed., Pearson Education, Inc., New Jersey, 2008.

[14] Pushpa, B. R., C. Anand, and P. Mithun Nambiar. "Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images." International Journal of Applied Engineering Research 11.7 (2016): 5142-5147.

[15] Du, Ji-Xiang, Xiao-Feng Wang, and Guo-Jun Zhang. "Leaf shape based plant species recognition." Applied mathematics and computation 185.2 (2007): 883-893.

[16] Zulkifli, Zalikha, Puteh Saad, and Itaza Afiani Mohtar. "Plant leaf identification using moment invariants & general regression neural network." Hybrid Intelligent Systems (HIS), 2011 11th International Conference on. IEEE, 2011.

[17] Neto, J.C., Meyer, G.E., Jones, D.D. and Samal, A.K., "Plant species identification using Elliptic Fourier leaf shape analysis". Computers and electronics in agriculture, 50.2 (2006):121-134.

[18] Zhang, Yudong, Shuihua Wang, Ping Sun, and Preetha Phillips. "Pathological brain detection based on wavelet entropy and Hu moment invariants." Bio-medical materials and engineering 26.1 (2015): 1283-1290.

[19] Ali Basim Yousif AL-Khafaji," Content-Based Image Retrieval System Using Three Set of Features", M.Sc. Thesis, Department of Computer Science, Faculty of Science, University of Basra, 2012.

[20] Haralick, Robert M. "Statistical and structural approaches to texture", Proceedings of the IEEE 67.5 (1979): 786-804.

[21] Priya, C. Arun, T. Balasaravanan, and Antony Selvadoss Thanamani. "An efficient leaf recognition algorithm for plant classification using support vector machine." Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on. IEEE, 2012.

[22] Larese, Mónica G., et al. "Automatic classification of legumes using leaf vein image features." Pattern Recognition 47.1 (2014): 158-168.

[23] Cha, Sung-Hyuk. "Comprehensive survey on distance/similarity measures between probability density functions", *International Journal of Mathematical Models and Methods in Applied Sciences*, 1.4 (2007):300–307.

[24] Wu, Stephen Gang, et al. "A leaf recognition algorithm for plant classification using probabilistic neural network." 2007 IEEE international symposium on signal processing and information technology. IEEE, 2007.

[25] 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, 3.4 (2010):67-78.

