RECOGNITION OF FRACTAL DIMENSION OF MEDICAL PLANT LEAVES BY USING TECHNIQUES OF IMAGE PROCESSING

G. GOPALA KRISHNAN[1], DR.D. CHITRA[2], MONICA.S[3]
1 ASSISTANT PROFESSOR, 2 ASSOCIATE PROFESSOR 3 PG STUDENT
ELECTRONICS AND COMMUNICATION ENGINEERING
MAHENDRA ENGINEERING COLLEGE
(AUTONOMOUS), NAMAKKAL, TAMILNADU, INDIA

ABSTRACT

The leaves of a plant having unique properties and it will be different from that of other leaves. So leaf recognition can represent a tree or plant recognition. This report describes some of the findings of currently published work in the area of leaf recognition. From these we are aiming to create a good leaf recognition system. Leaf Images acquired first goes for preprocessing. Then some relevant features are extracted. Our objective is to extract feature based on fractal dimensions. A fractal has been defined as a rough or fragmented geometric shape that can be split into parts, each of which is reduced-size copy of the whole. For classification a new learning algorithm called extreme learning machine (ELM) for single-hidden layer feed forward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs will be experimented using for the better result.

Keywords: - Leaf recognition, Extreme Learning Machine, Single-Hidden Layer Feed Forward Neural Networks

1. INTRODUCTION

1.1 DIGITAL IMAGE PROCESSING

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. An image is an array, or a matrix, of square pixels (picture elements) arranged in columns and rows.

Pictures are the most common and convenient means of conveying or transmitting information. A picture is worth a thousand words. Pictures concisely convey information about positions, sizes and inter-relationships between objects. They portray spatial information that we can recognize as objects. Human beings are good at deriving information from such images, because of our innate visual and mental abilities. About 75% of the information received by human is in pictorial form. In the present context, the analysis of pictures that employ an overhead perspective, including the radiation not visible to human eye are considered.

A digital image differs from a photo in that the x, y, and f (x, y) values are all discrete. Usually they take on only integer values, so an image will have x and y ranging from 1 to 256 each, and the brightness values also ranging from 0 (black) to 255 (white). A digital image can be considered as a large array of discrete dots, each of which has a brightness associated with it. These dots are called picture elements, or more simply pixels. The pixels surrounding a given pixel constitute its neighborhood. A neighborhood can be characterized by its shape in the same way as a matrix: we can speak of a 3x3 neighborhood, or of a 5x7 neighborhood.

1.2 ASPECTS OF IMAGE PROCESSING

Image Enhancement: Processing an image so that the result is more suitable for a particular application. (Sharpening or deblurring an out of focus image, highlighting edges, improving image contrast, or brightening an image, removing noise). The greatest difficulty in image enhancement is quantifying the criterion for enhancement and, therefore, a large number of image enhancement techniques are empirical and require interactive procedures to obtain satisfactory results. Image enhancement methods can be based on either spatial or frequency domain techniques.

Spatial domain enhancement methods: Spatial domain techniques are performed to the image plane itself and they are based on direct manipulation of pixels in an image. The operation can be formulated as g (x, y) =T [f (x, y)], where g is the output, f is the input image and T is an operation on f defined over some neighborhood of (x, y). According to the operations on the image pixels, it can be further divided into 2 categories: Point operations and spatial operations (including linear and non-linear operations). Frequency domain enhancement methods: These methods enhance an image f (x, y) by convoluting the image with a linear, position invariant operator. The 2D convolution is performed in frequency domain with DFT.

• **Image Restoration: This may be con**sidered as reversing the damage done to an image by a known cause.

(Removing of blur caused by linear motion, removal of optical distortions). All natural images when displayed have gone through some sort of degradation: during display mode, acquisition mode, or processing mode. The degradations may be due to sensor noise, blur due to camera miscues, relative object-camera motion, random atmospheric turbulence, others. Image restoration differs from image enhancement in that the latter is concerned more with accentuation or extraction of image features rather than restoration of degradations Image restoration problems can be quantified precisely, whereas enhancement criteria are difficult to represent mathematically

• **Image Segmentation:** This involves subdividing an image into constituent parts, or isolating certain aspects of an image. (Finding lines, circles, or particular shapes in an image, in an aerial photograph, identifying cars, trees, buildings, or roads.

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

In recent years a great deal has been published on numerical classification techniques by which the structure of populations can effectively be studied. The conventional techniques as for instance canonical analysis assume that a classification exists and that the problem is to test whether or not it is a valid classification or to assign individuals to the classes in which they best fit. In ecology the problem is frequently more basic the classes themselves need to be established. There had been rapid development in this field, with a great proliferation technique. The strategies proposed are however either divisive or agglomerative. In divisive classification subgroups may be formed by splitting a large group into 2 at the discontinuity of a single characteristic. Such monolethetic classifications unfortunately have the disadvantage that uninformative subdivisions on the basis of unimportant attributes are likely. Agglomeration is carried out in successive cycles in a manner such that the within-group sum of squares is minimized and, accordingly, the differences between the groups are maximized at each clustering cycle. The process results in the construction of a hierarchy of dichotomous branching. Two hierarchies are presented based on separate analyses of absolute distance and standard distance, utilizing data from a previous study of plant communities on the Squamish Flood-plain, British Columbia. The results indicate that the structure imposed on the sample is dependent on the distance measure by which stand relationships are defined. The hierarchies, however, are similar and at one level the classification units closely follow the major environmental divisions on the flood-plain. These units are described as vegetation types. The corresponding class centroids are ordinated in a principal components analysis.

2.2 PROPOSED METHOD

This project is mainly used for the identification of medicinal plants. It contains two parts. First one tries to make a classification based on the geometrical features. Around fifteen feature values are extracted. Second part is to is an attempt find fractal dimensions of preprocessed images. It implements algorithms such as Blanket method to find fractal dimension. The images are preprocessed using 2D wavelet transform. In this project a robust and computationally efficient method for medicinal plant specie recognition from leaf image is obtained. The system is intelligent enough to identify a plant from a partially damaged or broken leaf. This system shows 90% accuracy when compared to other systems.

3. SYSTEM SPECIFICATION

3.1. HARDWARE REQUIREMENTS

Processor	:	Intel (R) core(i3) CPU
Processor speed	:	2.5 GHZ
Memory (RAM)	:	256MB

Hard disk	:	3000GB
Monitor	:	13” color monitor
Keyboard	:	Standard 104 keys
Mouse	:	Optical mouse

3.2. SOFTWARE REQUIREMENTS

Platform	:	Windows 7
Simulation tool	:	MATLABR2010A

4. PROJECT DESCRIPTION

4.1. OVERVIEW OF THE PROJECT

It is quite clear that leaves of a plant having unique properties and it will be different from that of other leaves. So leaf recognition can represent a tree or plant recognition. This report describes some of the findings of currently published work in the area of leaf recognition. From these we are aiming to create a good leaf recognition system. This project contains two parts. First one tries to make a classification based on the geometrical features. Around fifteen feature values are extracted. Second part is to is an attempt find fractal dimensions of preprocessed images. It implements algorithms such as Blanket method to find fractal dimension. The images are preprocessed using 2D wavelet transform.

Leaf Images acquired first goes for preprocessing. Then some relevant features are extracted. Our objective is to extract feature based on fractal dimensions. A fractal has been defined as a rough or fragmented geometric shape that can be split into parts, each of which is reduced-size copy of the whole. For classification a new learning algorithm called extreme learning machine (ELM) for single-hidden layer feed forward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs will be experimented using for the better result.

4.1.1 BLOCK DIAGRAM DESCRIPTION

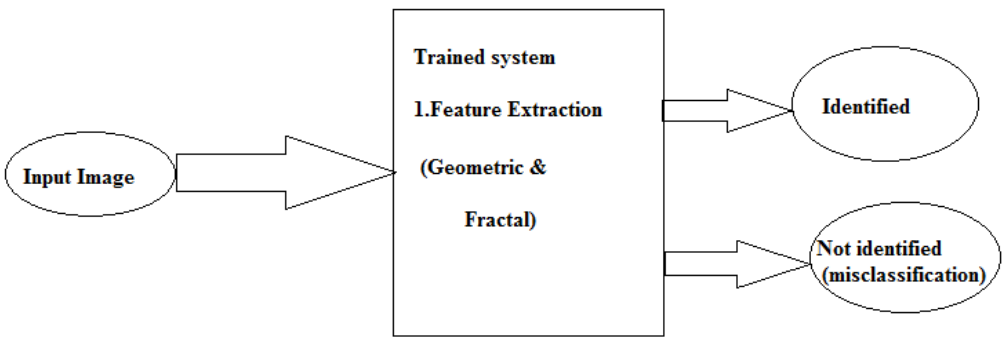

Fig: 1 Block Diagram

The basic steps involved in the proposed system are:

Image Acquisition
Image Preprocessing
Feature Extraction
Classification

4.2 AIM AND EXPECTED OUTCOME

The main objective of this project is to propose an efficient method to classify leaf images. Here classification refers to identification and annotation of leaves when the image is given as input to the system. So the leaf images present in the huge database must be able to classify with several classes according to the shape of the leaf by the developing system. Of the 2, 50,000 higher plant species on earth, more than 80,000 are medicinal. India is one of the world’s 12 biodiversity centers with the presence of over 45000 different plant species. However, only 7000-7500 species are used for their medicinal values by traditional communities.

Our first aim was to create a medicinal leaf image database. The medicinal leaves in the rural area of Kerala were covered.

The border of the leaf was being extracted.

Feature extraction: Several morphological features such as area, perimeter, major axis, minor axis, equivalent diameter, convex area, aspect ratio etc. are extracted.

Some fractal values are found using methods like blanket method.

These features are used as inputs to the train the network.

Now the final task is to feed the classifier with these features and compare the result.

5.IMPORTANCE OF MEDICINAL PLANTS

Plants exist everywhere we live, as well as places without us. Plants play a critical role in preserving the delicate balance of the environment. Many of them carry significant information for the development of human society. Unfortunately, the overwhelming development of human civilization has disrupted this balance to a greater extent than we realize. The urgent situation is that man plants are at the risk of extinction. It is one of our biggest responsibilities to save the plants from various threats, restore the diverseness of the plant community and put everything back to balance. So it is very necessary to set up a database for plant protection. We believe that the first step is to teach a computer how to classify plants. A computerized plant identification system can be very helpful in botanical garden or natural reserve park management, new plant species discovery, plant taxonomy, exotic plant detection, edible/poisonous plant identification and so on. Plants play a critical role on human life. This role includes food, medicine, industry and environment. The earth hosts a huge number of plant species and thus identification of each species is hard. A quick and accurate recognition and identification of a plant is essential for scientists in the area of agriculture, botany and environment. Plant species identification has been carried out by botanists and plant specialist for many years using different plant features like leaf shapes, fruits, flowers and color.

6. PRE-PROCESSING STAGES

6.1 INTRODUCTION

The preprocessing is done to identify the leaf in an image and discarding all other information other than the leaf shape. This can be done with a little help from the user. The user can help identify the base-point and some reference points of the leaf. The leaf is extracted from the background and a binary image is produced where the background pixels are set to 0 or black and the pixels within the leaf is set to 1 or white. The remaining black pixels within the leaf blade are removed to produce an enhanced binary image. Then the tip of the leaf is located by finding out the furthest point (which is, in most cases, the tip of the leaf) from the base-point (selected by the user).

6.2 DATA ACQUISITION

This is the initial stage of the project, where the given data i.e. image that is already in digital form or should convert that into a digital image by an analog to digital converter. Finally, the outcome must be a digital image for further processing. In order to train our model and test results it provides, we need a large number of leaf images. Here I made a database of forty classes of medicinal leaves and each class consists of thirty images to train the system. I test the system by randomly taking images from the dataset. The test set consists of totally 50 images, where each class contains 75 images. Below shown is the the snapshot of sample database.

Figures below shows the snapshot of some of the leaf images in database

1. Sandal

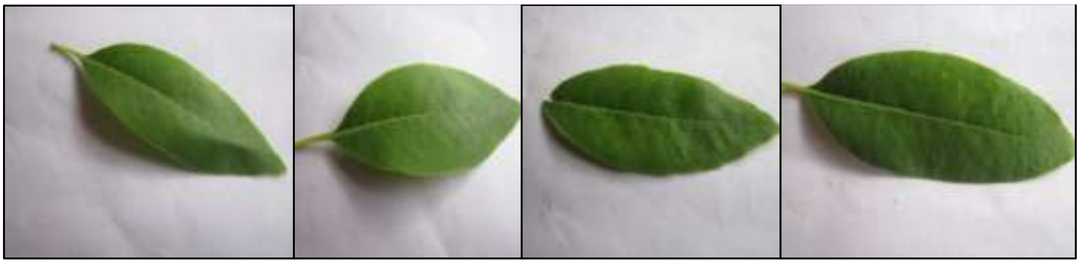

Fig 2: Sandal leaf images database

2. Bird's eye chilli

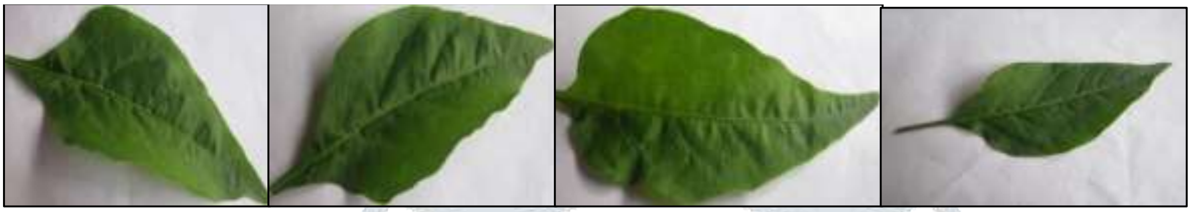

Fig 3: Bird's eye chilli Leaf image database

3. Tapioca

Fig 4: Tapioca leaf image database

6.3 THRESHOLDING

Most of the image processing techniques are carried out using a binary image, because the processing will be easy by taking only two level of values. The method of converting a color images (3channel 0-255 level) or gray level image (0-255 level) into binary image (0-1 level) is known as thresholding. Thresholding helps in separation of foreground from its background. We can choose a single threshold for an image if the foreground and background intensities are uniform. For no uniform foreground and background intensities adaptive thresholding is used.

Otsus Method: Its a nonparametric approach for thresholding. Calculations are first made of the ratio of between-class variance to within-class variance for each potential threshold value. The classes here are: foreground and background pixels. The purpose is to find the threshold that maximizes the variance of intensities between the two classes, and minimizes them within each class. We exhaustively search for the threshold that minimizes the intra-class variance, defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t)$$

Which is expressed in terms of class probabilities in and class means μ i which in turn can be updated iteratively. This idea yields an effective algorithm.

Fig 5: Input image & Binary image

6.4 ROTATING THE LEAF

This preprocessing step is mainly using for the detection of features like leaf width factor. The tip of the leaf is located. This is done by finding out the major axis positions in the leaf. Then the slope of the line connecting the major axis t and the horizontal axis is calculated. Finally, the enhanced binary image is rotated according to the angle of inclination to make the leaf horizontally aligned.

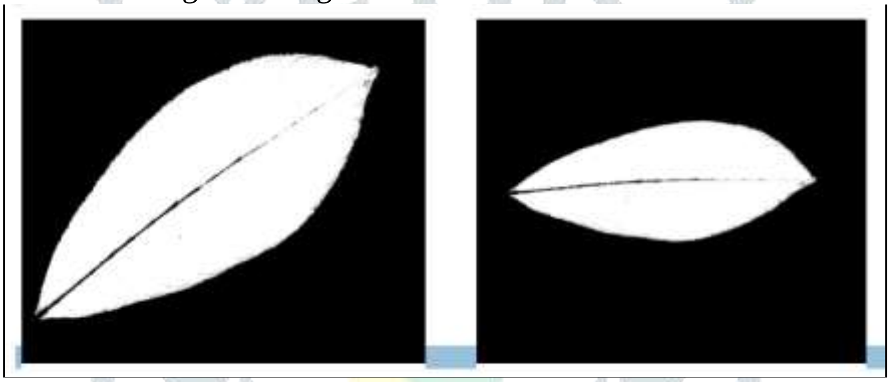

Fig 6 Input image & Binary image

6.5 FINDING THE EDGE

The purpose of edge detection in general is to significantly reduce the amount of data in an image, while preserving the structural properties to be used for further image processing. The canny edge detector first smoothes the image to eliminate and noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum. The gradient array is now further reduced by hysteresis. Hysteresis is used to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero. If the magnitude is above the high threshold, it is made an edge. And if the magnitude is between the two thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient above T2. These are the following steps:

Step 1: Smoothing:

The first step is to filter out any noise in the original image before trying to locate and detect any edges. And because the Gaussian filter can be computed using a simple mask, it is used exclusively in the Canny algorithm. Once a suitable mask has been calculated, the Gaussian smoothing can be performed using standard convolution methods. A convolution mask is usually much smaller than the actual image. As a result, the mask is slid over the image, manipulating a square of pixels at a time. The localization error in the detected edges also increases slightly as the Gaussian width is increased.

Step 2: Finding the border

By doing continuous iteration from all sides of the smoothened binary image we will get the border of leaves. This is an iterative procedure using the help of for loops. This method is working perfectly for leaves with less folding.

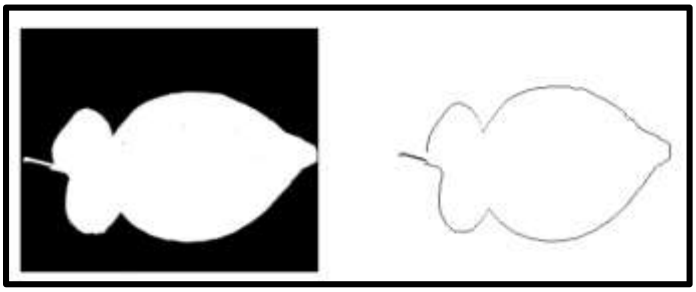

Fig 7 Rotated Image

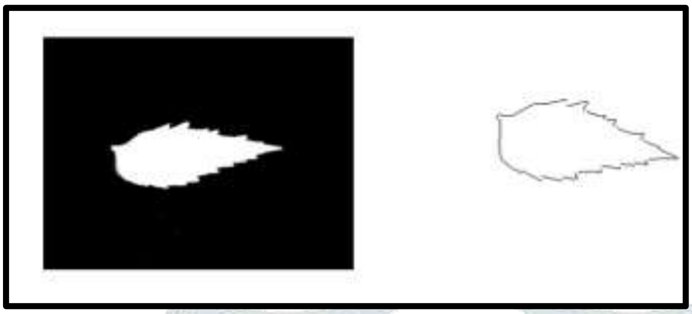

Fig 8: Border of the Image

7. EXPERIMENTAL RESULTS

7.1 USING 10 GEOMETRICAL FEATURES

Ten geometric features are extracted from each leaves. They are physiological length, width, area of leaf, perimeter, diameter, aspect ratio, form factor, rectangular factor, perimeter ratio, perimeter ratio of length and width. The results obtained are as follows

Sl no	Leaf Name	No of Samples For testing	Correct classification using	
			Minimum Distance Classifier	Nearest Neighbor
1	Sandal	10	8	8
2	Neem	10	8	8
3	Hibiscus	10	6	6
4	Capsicum	10	8	8
5	Grambu	10	8	8
6	Teak	10	10	10
7	Philanthus	10	10	10
8	Tapioca	10	10	8
9	Bael tree	10	8	10
10	Amaranthus	10	6	6

Table : 1 Using 10 geometrical features

Accuracy of the classification: Out of 100 testing samples percentage of correctly classified using
minimum distance classifier is 82%
nearest neighbor classifier is 80%

7.2 USING 30 GEOMETRICAL FEATURES

In addition to the above given 10 features we are finding 20 more values for each leaf. We are taking 10 values of leaf width factor for each leaf and 10 values of the contour distance between leaf border and the center of leaf.

Sl no	Leaf Name	No of Samples For testing	Correct classification using	
			Minimum Distance Classifier	Nearest Neighbour
1	Sandal	10	8	8
2	Neem	10	8	8
3	Hibiscus	10	7	8
4	Capsicum	10	8	6
5	Grambu	10	8	8
6	Teak	10	10	10
7	Philanthus	10	10	8
8	Tapioca	10	10	8
9	Bael tree	10	9	10
10	Amaranthus	10	7	8

Table: 2 Using 30 geometrical features

Accuracy of the classification: Out of 100 testing samples percentage of correctly classified using
minimum distance classifier is 86%
nearest neighbor classifier is 82%

Results obtained are better for combinational features. The main source of error comes from the unbalanced database and the proper preprocessing methods. The extraction of leaf from its background is one of the challenging tasks. By using geometrical features, we are getting around 85 to 90% accuracy. In the end, from this study and the results obtained, we can say that a classification based on the geometric features is one of the best ways to identify leaf images.

8. CONCLUSION

Experiments on 10 leaves with 30 samples were conducted. Our objective is to extract geometrical features. A classification based on different combinations of geometrical features was conducted. Whenever the number of distinguishable features are increasing the accuracy of our system is also increasing propotionally. So the difficult task is to find the best combination of features. The other factors which effecting the classification are the preprocessing methods and the type of the classifier. For classification nearest neighbor and a minimum distance classifier algorithms are applied here. In addition to this a classification based on fractal geometry is also developed. A fractal has been defined as a rough or fragmented geometric shape that can be split into parts, each of which is reduced-size copy of the whole. Finally, we can tell our system doesn't match to the human ability of plant identification. Human brain is far too powerful to be compared with our system. But when it comes to identifying a plant from hundreds and thousands leaf samples, a system like this can definitely be of use.

9. REFERENCES

[1]. Mark S. Hamell, Przemyslaw Prusinkiewicz, and Brian Wyvill "Modelling Compound Leaves Using Implicit Contours", Department of Computer Science University of Calgary Calgary, Alberta, Canada T2N 1N4June 1992. Proceedings of Computer Graphics International '92 (Tokyo, Japan, 22–26 June 1992).

[2]. Z. Wang, Z. Chi and D. Feng "Shape based leaf image retrieval" @ IEEE. 2003 Paper first received 19th December 2001 and in revised form 16September 2002

[3]. Yan Li, Zheru Chi, Member, IEEE, and David D. Feng, Fellow, IEEE" Leaf Vein Extraction Using Independent Component Analysis" 2006 IEEE Conference on Systems, Man, and Cybernetics October 8-11, 2006, Taipei, Taiwan.

[4]. Stephen Gang Wu1, Forrest Sheng Bao2, Eric You Xu3, Yu-Xuan Wang4, Yi-Fan Chang5 and Qipao-Liang Xiang4 1 Institute of Applied Chemistry, Chinese Academy of Science, P. R. China 2 Dept. of Computer Science, Texas Tech University, USA "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", 29 Jul 2007.

[5]. Hassan Hajjdiab and lman AI Maskari, College of Engineering and Computer Science; Abu Dhabi University," Plant Species Recognition Using Leaf Contours", ©2011 IEEE

[6]. Jived Hossain, M. Ashraful Amin Department of Electrical Engineering and Computer Science, North South University, Dhaka, Bangladesh; School of Engineering and Computer Science, Independent University, Dhaka, Bangladesh. "Leaf Shape Identification Based Plant Biometrics", Proceedings of 13th International Conference on Computer and Information Technology (ICCIT 2010) 23-25 December, 2010, Dhaka, Bangladesh

[7] Qi Guo · Jiaqing Shao · Virginie F. Ruiz" Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms"

[8] S. Wu, F. Bao, E. Xu, Y. Wang, Y. Chang, and Q. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," IEEE 7th International Symposium on Signal Processing and Information Technology, December 2007.

[9] Z. Wang, Z. Chi, and D. Feng, "Shape based leaf image retrieval," IEEE Proceedings-Vision, Image and Signal Processing, vol. 150, no. 1, February 2003.

[10] H. Fu, Z. Chi, D. Feng, and J. Song, "Machine learning techniques for ontology-based leaf classification," IEEE 2004 8th International Conference on Control, Automation, Robotics and Vision, Kunming, China, 2004.