Leaf Classification using Structure Features and Support Vector Machines

Sarin Watcharabutsarakham
National Electronics and Computer Technology Center,
Klong 1, Klong Luang, Pathumthani, 12120 Thailand
e-mail: sarin.watcharabutsarakham@nectec.or.th

Wasin Sinthupinyo[1] and Kantip Kiratiratanapruk[2]
National Electronics and Computer Technology Center,
Klong 1, Klong Luang, Pathumthani, 12120 Thailand
e-mail: wasin.sinthupinyo@nectec.or.th[1] and
kantip.kiratiratanapruk@nectec.or.th

***Abstract*—In Thailand, there are a lot of near extinction herbs, because they are used as an essential resource for food and medicine industry. Before they become extinct, the herb information needs a systematic collection. This paper is proposed a herb leaf classification method for making an automatic categorization. In digitization step, the leaves with white background are photographed with digital camera. As a preprocessing step, apex and leafstalk are removed with a histogram-based method. To describe leaf characteristic, three types of ratio; aspect ratio, slice ratio and radius ratio are measured. These features are scale in variant and -/+15 degrees tolerance of rotation. In experimental results, our technique performs 95 percent accuracy on 10 classes of leaf type.**

***Index Terms*— SVM, Classification, Leaf , Image**

I. Introduction

In Thailand, there are a lot of near extinction herbs, because they are used as an essential resource for food and medicine industry. Before they become extinct, botanists collect herb information systematically. In general, the herb collection and classification are done by biologist. These processes are complicate and heavy time consumption. Recently, image processing and pattern recognition are applied to make an automatic herb leafs classification.

Contrast with many categories by biologists, image recognition uses appearance inform such as shape, color, structure, texture and etc. to make a classifier. Hossain, J. and Amin, M.A. proposed an extraction method of leaf features such as eccentricity, area, perimeter, major axis, minor axis, equivalent diameter, convex area and extent [2]. This paper used a probabilistic neural network for training 1200 leaf samples to classify 30 different plant species. Stephen Gang Wu et al. proposed a method for extracting 12 common morphological features. They employed a probabilistic neural network to train 1800 leaf samples from 32 kinds of plants with an accuracy rate greater than 90% [3]. Zheng Xiao-dong et al. introduced a new idea on feature extraction named as feature extraction based on visual consistency (FEBVC). Many leaf shapes have been evaluated and the results showed a good feasibility [4].

This paper presented a methodology for leaf shape classification. As show in Fig. 1, there are five processes in this system and details of each process shown below. The variation outer edge distance measurements are used to describe leaf characteristic. As an advantage, the features are robust with an incomplete leaf which damaged by insects or diseases. As an efficient classifier, SVMs are used to categorization leaf images.

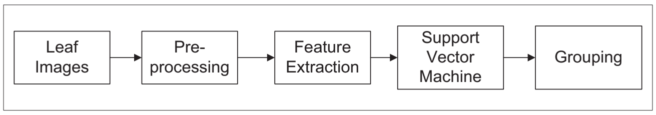


Figure 1. Flow Process

II. Image acquistion

Leaf images are photographed from a digital camera with JPEG format and represented in RGB model. However, the background is clearly white or single color with contrast the leaf color. Image size and image resolution are not restricted and the image should focuses on a leaf. The image of leaf viewed in the vertical direction with leafstalk on the bottom and the apex on top. Experimentation is also used to use 2500 images for training set and 500 images for testing set.

III. PreProcess

In this process, the leaf is extracted by removing background, leafstalk and apex. Normally, the leaf is composed of leafstalk and apex which they make the variance feature values.

A. RGB to Gray Conversion)

The leaf images are digitized with a camera in RGB model. A leaf with white background lays in the vertical direction view with apex on top. Thus given a color input image has to convert to a grayscale image. Let R, G, B be the three color channels of the image. Classically, the gray-scale image is obtained as in Eq. 1.

$$J(x, y) = \alpha.R(x, y) + \beta.G(x, y) + \gamma. B(x, y) \quad (1)$$

Where α, β and γ are the weights corresponding to the three color channels, R, G, and B, respectively, and (x, y) are the pixel location in the input image. Summation of all values in an image becomes a threshold value. The value used to segment between background and foreground of the image. If the pixel values over threshold value, it is the background and otherwise is the foreground. Then, leaf area is cleaned up and sharp.

(a) (b) (c)

Figure 2. Leaf Segmentation (a) An Original Leaf (b) Result of Background Segmentation (c) Result of Pre-processing

B. Leaf Boundary

A natural leaf consists of leafstalk and apex, so we need to remove both of them from the leaf image. The feature values from a leaf will be bias if the leaf has leafstalk and apex. A remove technique is a histogram-based method with horizontal direction. It is calculated and sorted to find the local minimal threshold value. At top of list, if the histogram value less than the local minimum value, it will be removed as an apex of leaf. In contrast, if the histogram value less than the local minimum value, it will be removed as a leafstalk of leaf as shown in Fig. 2.

Next, Sobel's edge detection is used to estimate a leaf boundary, because it is the most popular edge detection operator. Then, dilation and erosion technique are used to improve an incomplete edge. They will increase or decrease the leaf edge to connect or disconnect with near pixels. Then the closed edge of leaf area is filled black and the outside is filled white.

IV. FEATURE EXTRACTION

Feature extraction converts data patterns to the features space, the good features were represented pattern with minimal error for best classification. Our feature extractions are calculated on a major axis and a minor axis. A vertical direction line is a major axis and its perpendicular line is a minor axis. The major line, called Md, is a longest distance line which is started on top of leaf area and end on bottom of leaf area. A datum obtained from the feature extraction is an aspect ratio form because the scaled data obtains an effective training of neural networks.

A. Aspect Ratio of Leaf(AR)

An aspect ratio feature is calculated from width and height of the leaf. This feature explains narrow or wide leaf characteristics as shown in Fig. 3(a)

$$AR = W/H \tag{2}$$

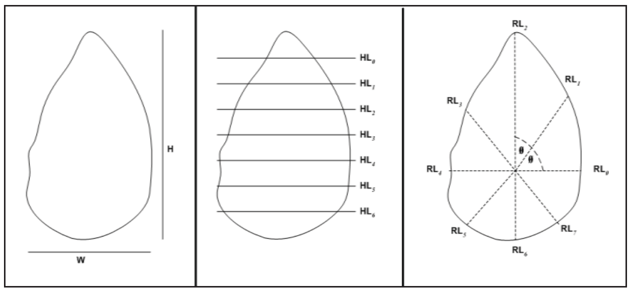
Where W is the leaf width and H is the leaf height.

B. The slice ratio of leaf(HL)

The leaf is sliced into 7 portions in a vertical direction line. Each piece is equal length and parallels a minor axis. The features are calculated by ratio of width in each piece. These features explain the physical characteristics of leaf in vertical direction as shown in Fig. 3(b). The ratio features uses an equation, as in (2). J. Hossain and M. A. Amin decided to use 11 slices which enough their method. We decided to use 7 slices because the number of slices between 7 and 13 are not difference with significant.

$$HL_i = W_i / Md \text{ when } i = 1,2,3,4,5,6,7 \tag{3}$$

Where W_i is length of slice on a minor axis at i position, Md is a longest distance line in vertical direction.



(a) (b) (c)

Figure 3. Extracted Features

C. The radius ratio of Leaf(RL)

The features are ratios of radial length in 8 angles which are explain the shape characteristics of leaf as shown in Fig. 3(c). A radial line is a straight line from a center point to a point on leaf edge in an angle. Let a center point is a point that the longest line in vertical direction perpendicular with the longest line in horizontal direction. Our proposed method used 8 degrees (0°, 45°, 90°,135°,180°,225°,270°,315°) are enough to classify.

$$RL_i = R_i / Rd \tag{4}$$

Where i is a set of degrees between 0° and 315°, R_i is length from a center point to a edge point at i angle, Rd is the longest distance from a center point to a edge point.

D. Ellipse Equilibrium(EA)

The feature is calculated from a leaf area and a created ellipse area. The created ellipse area is created by using the center point of leaf to be a center point of ellipse. Moreover, the major axis of leaf becomes a major axis of ellipse and the leaf minor axis becomes a minor axis of ellipse. So, area of interest is overlap between leaf area and created ellipse area as shown in Fig. 4(a).

$$EA = IA / EC \quad (5)$$

Where IA is a number of pixels in an interested area, EC is a number of pixels in a created ellipse area.

E. Circle Equilibrium(CA)

The feature is calculated from the leaf area and a created circle area. The created circle area uses the center point of leaf to be a center of circle. The radius of circle is a shortest length between width and height. The area of interest is overlap area within the leaf area and created circle area as shown in Fig. 4(b).

$$CA = IA / CC \quad (6)$$

Where IA is a number of pixels in the interested area, CC is a number of pixels in the created circle area.

V. CLASSIFICATION

The proposed method, we used Support Vector Machines (SVM) to classification. SVM is an supervised approach based on statistical learning theory which tries to map the present information to optimal boundary in feature space and seeks a minimum distance between 2 difference classes. The advantage of SVM is ability dealing with large sample and many dimensionality patterns.

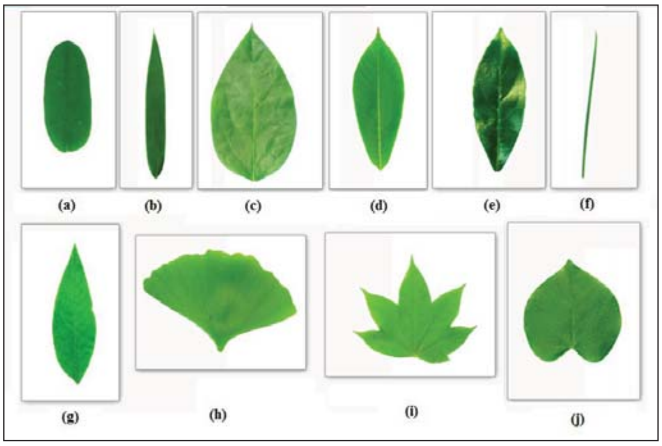


Figure 4. Samples of leaf images (a) oblong (b) linear (c) ovate (d) oblanceolate (f) acicular (g)lanceolate (h)reniform(i)palmatifid(j)cardate

Two thousand and five hundred input images correspond to 10 classes such as acicular, linear, oblong, cordate, ovate, lanceolate, oblanceolate, spathulate, reniform and palmatifid, shown in Fig. 5. Each leaf image is extracted 19 features and set in a feature vector $X_i = (x_1, \ldots, x_{19})$. All input images are set in a pattern format $H = \{X_1, \ldots, X_{2500}\}$ and a matrix size is 2500x19 . Then, the pattern matrix is sent to train with SVM and return the model. Our proposed method applied in various the leaf images and an incomplete leaf which damaged by insects or diseases.

VI. EXPERIMENT RESULT

The proposed method was implemented by Microsoft Visual C++ and SVM library. 500 leaf images are used for testing set and leaves are varied in type, size, color and flaw. The testing images are photographed by a digital camera with JPEG format. Then, they are sent to the preprocess method for background subtraction and removed apex and leafstalk. Next, 19 features are extracted from leaf images. We design experimentations by dividing 19 features into 3 groups, because some features are indeed related. First group uses 4 features: AR, EA, CA and AL. Second group uses 7 features of HL. Then, Third group uses 8 features of RL. In the experiment, SVM parameters are varied in many scales for the best accuracy rate.

In this part, there are 6 experiments. The first experiment uses 4 features: AR, EA, CA and AL. The second experiment uses 7 HL features. The third experiment uses 8 RL features. Next, three experiments are designed by combining across groups. The fourth experiment uses 7 HL features and 4 features: AR, EA, CA and AL. The fifth experiment uses 7 HL feature and 8 RL features. The sixth experiment uses 19 features. In Second Experiment, the proposed technique get more the accuracy rate than biometrics technique‘s result [1]. All experiment results are shown in Table I.

TABLE I. THE EXPERIMARNTAL RESULTS

Features	Accuracy Rate
First: 4 features: AR, EA, CA and AL	74 %
Second: 7 HL features	94 %
Third: 8 RL features	85 %
Fourth: 7 HL features, AR, EA, CA and AL	95 %
Fifth: 7 HL features and 8 RL features	89 %
Sixth: AR, EA, CA, AL, 7 HL features and 8 RL features	94 %

The results of experiments shown that 11 features are suitable for classify the leaf because an accuracy rate is percentage of 95.

VII. CONCLUSION

This paper proposed a method to a leaf classification using SVM. All features are based on structure of leaf characteristic. In experimentation, we found that more features may decrease an accuracy rate but the selected suitable feature is important. The 3 results of experiments are compared with each other and found that the accuracy rates are a few difference as shown in Fig. 6. Our proposed method get an accuracy rate higher than biometrics technique because of noise removing process. As an advantage, the features are robust with an incomplete leaf which damaged by insects or diseases.

ACKNOWLEDGMENT

This paper could not have been written without Dr. Ithipan Methasate who not only served as my consultant but also encouraged and challenged me throughout my English language.

REFERENCES

[1] X.F. Wang, J.X. Du, and G.J. Zhang, “Recognition of leaf images based on shape features using a hypersphere classifier,” International Conference on Intelligent Computing, pp. 87-96, 2005.

[2] J. Hossain, and M.A. Amin,"Leaf shape identification based plant biometrics," 13th International Computer and Information Technology (ICCIT), 2010

[3] S. G. 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, pp. 11-16, 2007.

[4] Zheng Xiao-dong and Wang Xiao-jie, "Feature Extraction of Plant Leaf Based on Visual Consistency," Computer Network and Multimedia Technology, pp.1-4, 2009.

[5] Chih-Chung Chang and Chih-Jen Lin (2012). Available:http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[6] 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.

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

[8] X. Wang, D. Huang, J. Dua, H. Xu and L. Heutte, "Classification of plant leaf images with complicated background," Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation, vol. 205, pp. 916-926, November 2008.

[9] J-X Du, D-S Huang, X-F Wang and X Gu, "Computer-aided plant species identification (CAPSI) based on leaf shape matching technique," Transaction of the Institute of Measurement and Control, vol. 28, pp. 275-284, 2006.

[10] J. Chaki and R. Parekh, “Plant Leaf Recognition using Shape based Features and Neural Network classifiers,” International Journal of Advanced Computer Science and Applications , vol. 2, pp.41-47, 2011.

[11] Z. Zulkifli, P. Saad and I.A. Mohtar, "Plant Leaf Identification using Moment Invariants & General Regression Neural Network," 11th International Conference on Hybrid Intelligent Systems (HIS), pp. 430 – 435, 2011.