Leaf and Flower Recognition System (e-Botanist)

Chomtip Pornpanomchai, *Member, IACSIT*, Chawin Kuakiatngam
Pitchayuk Supapattranon, and Nititat Siriwisesokul

***Abstract*— The objective of this research is to build a computer system that can recognize a plant by using either its leaf or flower image. The system consists of 4 main modules, 1) image acquisition, 2) image preprocessing, 3) image recognition, and 4) display result. In the image acquisition module, the system captures a leaf or flower image, which is taken with a digital camera. The image is taken by using white paper as a background, with a ping-pong ball at a side of the leaf or flower. In the image preprocessing module, the system applies several image processing techniques to prepare a leaf or flower image for the features extraction process. In the image recognition module, the system extracts eight main features from the leaf or flower image and recognizes it by the Euclidean distance algorithm. In the display result module, the system displays the recognition results. The experiment was conducted on 30 kinds of leaves with a total of 1,075 images, and 30 kinds of flowers with a total of 1,075 images. The precision of the system is 76.8 percent and 74 percent for recognizing leaf and flower images, respectively. The average access time for the system is 6.21 seconds per leaf image and 5.69 seconds per flower image.**

***Index Terms*— leaf recognition, flower recognition, image processing.**

I. Introduction

Thailand is located in Southeast Asia, which is hot, warm and rainy. For this reason, Thailand is rich in natural resources and there are various species of plants, some of which can be found generally, while others can only be found in specific areas. These plants are both useful and dangerous. Memorizing and recognizing species of plants are a difficult and important task for people. As the world achieves a major breakthrough in a digital camera technology, people can take a picture anywhere and any time. Therefore, the objective of this project is to develop the leave or flower images recognition system that allows humans to differentiate the sorts of plants more precisely.

II. Related work

There are more than 250,000- 270,000 plant species that have been named around the world. Therefore, it is very difficult for people to recognize all of them. There are a lot of researchers trying to build an automatic system that can recognize plants by using both leaves and flowers. Some of the researches have the following details [1], [2].

A. *Plant leaf recognition system*

Normally, a plant leaf has the green color. Therefore only the RGB color feature cannot be used to recognize plants leaves. Researchers use more features and techniques to recognize a plant leaf image. For example, they use shape, size, vein and texture of the leaf to recognize plant species. There are many researchers using a neural network method to recognize a leaf image. Moreover, some researchers use the fuzzy logic system or a support vector machine to recognize the plants. However, there is no consensus as to, which method is the best for recognizing a leaf image [2]-[13].

B. *Plant flower recognition system*

There are a lot of researchers using flower features such as color, size, shape, boundary etc. to recognize flowers. T. Saitoh, *et al*. used both the leaf and flower to recognize wild flowers. M. Zhenjiang *et al*. applied modified Fourier descriptor and shape analysis to recognize rose flowers. S. Yeung, and P.T. Lim, demonstrated a virus infection clustering for flower image identification [14]-[20].

Based on the previous researches, the systeme is developed to help people to recognize the plants by using either a leaf or flower image. The method to build the system will be presented in the next section.

III. Methodology

This section presents the methodology of the system. First, the system conceptual diagram is presented. After that the mapping between the system conceptual diagram and the system structure chart is illustrated.

A. *The system conceptual diagram*

The system conceptual diagram is shown in Figure 1. The system starts with taking leaf or flower images, after that stores them in the system database. The system gets an unknown leaf or flower image from a user and extracts several features from the unknown leaf or flower image such as the color, size, ratio, roundness, etc. by using image processing techniques. Finally, the system uses all leaf or flower features to identify the plant properties.

B. *The system structure chart*

The system structure chart is created by mapping the system conceptual diagram in Figure 1. The system structure chart has four processing modules, which are 1) image acquisition module, 2) image preprocessing module, 3)

Manuscript received April 22, 2011.

Dr. Chomtip Pornpanomchai is an Assistant Professor in the Faculty of Information and Communication Technology, Mahidol University, Bangkok 10400, Thailand, e-mail: itcpp@mahidol.ac.th, Tel:662-354-4333.

Chawin Kuakiatngam, Pitchayuk Supapattranon, and Nititat Siriwisesokul are graduated from the Faculty of Information and Communication Technology, Mahidol University, Bangkok 10400, Thailand, e-mail:{u5088126,u5088150,u5088227}@student.mahidol.ac.th, Tel:662-354-4333.

image recognition module, and 4) display result module (as shown in Figure 2). Each module has the following details.

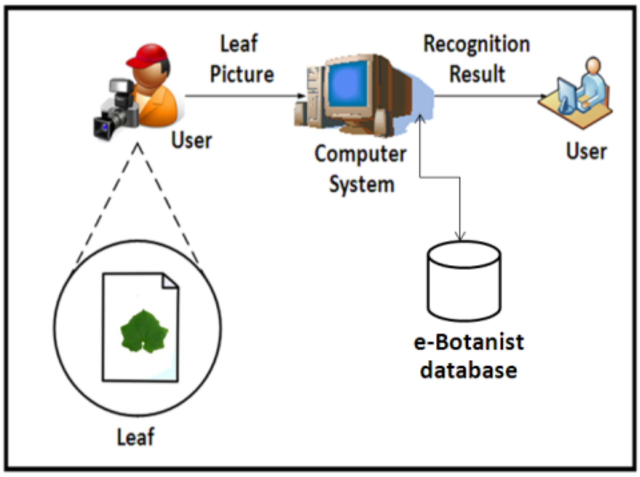


Figure 1. The system conceptual diagram

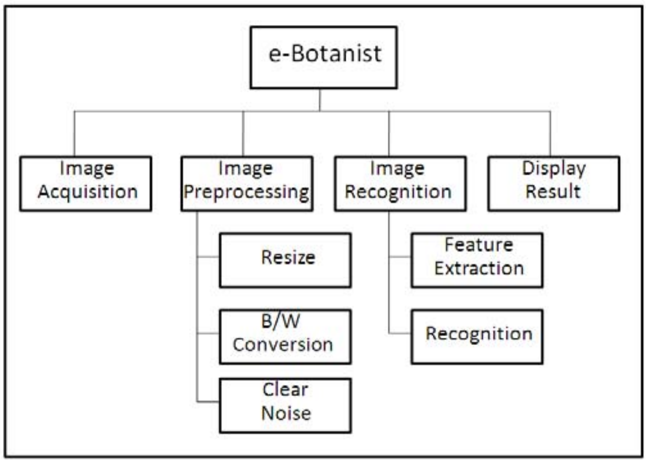


Figure 2. The system structure chart

1) Image acquisition module

This module takes an unknown leaf or flower picture, with a ping-pong ball at its side, as shown in Figure 3. The ping-pong ball is painted in black color to avoid a light reflection problem.

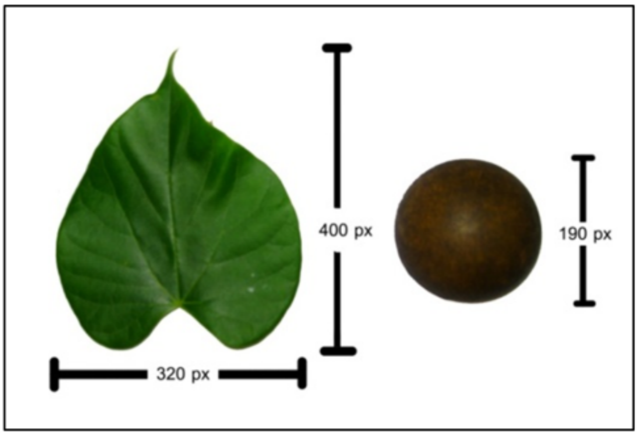


Figure 3. Comparing a leaf and a ping-pong ball

2) Image preprocessing module

This module has three sub-modules, which are:

- *Resizing sub-module*

This sub-module resizes the height of a leaf or a flower to 400 pixels and the width of a leaf or a flower to 320 pixels.

- *Black and white conversion*

This sub-module converts a leaf or flower image to gray-scale color, and then changes a gray-scale image to black-and-white image. The equation used to change an RGB image to a gray-scale image is shown in Equation 1.

$$G = 0.299*R + 0.587*G + 0.114*B \quad (1)$$

Where G = gray, R = red, G = green and B= blue.

- *Clear noise*

This sub-module clears noise in the leaf or flower picture for easily to extract their features. The system uses the erosion and dilation technique to clear the noises.

3) Image Recognition

This module consists of two sub-modules, which are:

- *Feature extraction*

This sub-module extracts eight main features from an object. The eight main features have the following details.

- Height ratio

Height ratio is a ratio between the height of a leaf or a flower and a ball. Based on Figure 3, the height ratio of the leaf and the ball is equal to 400/190 or 2.105.

- Area ratio

Area ratio is a ratio between the area of a leaf or a flower and a ball. Based on Figure 3, the area ratio of the leaf and the ball is equal to 128,000/113,352 or 1.129.

- Width and height ratio

Width and height ratio is a ratio between the width and height of a leaf or a flower. Based on Figure 3, the width and height ratio of the leaf is equal to 320/400 or 0.8.

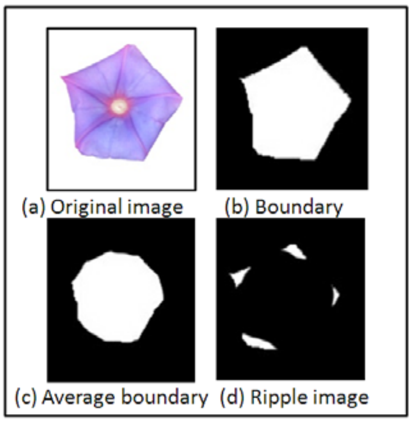


Figure 4. Finding ripple image

- Roundness value

Roundness value is an approximate roundness value of a leaf or a flower. The equation used to calculate a roundness value is shown in Equation 2. The boundary of a leaf or a flower in Equation 2 is approximate length of a leaf or a flower.

$$R = (4\pi * A) / B^2 \qquad (2)$$

Where: R = Roundness value; A = leaf or flower area; B = leaf or flower boundary.

- Ripple feature

A ripple feature value is a difference value between leaf image and average boundary of leaf image as shown in Figure 4 (b) – (c). The ripple feature consists of two sub-features, which are 1) the number of ripples counting, and 2) the ripple area counting.

- Half leaf of flower area ratio

The system divides a leaf or flower image into two equal areas by a horizontal line. After that, the system finds an upper-area ratio by dividing the upper leaf or flower area by the upper-half image area. Similarly, a lower-area ratio is found by dividing the lower leaf or flower area by the lower-half image area.

- Color feature

The system applies average RGB (red, green, blue) color for three sub-features. Then the system transforms the RGB color into average L*a*b* (luminance, chromaticity channel a, chromaticity channel b) color for three other sub-features. The formulas used to calculate the L*a*b*values are shown in Equation 3–8.

$$X = k1R+k2G+k3B \qquad (3)$$
$$Y = k4R+k5G+k6B \qquad (4)$$
$$Z = k7R+k8G+k9B \qquad (5)$$
$$L^* = 116(Y/Yn)1/3 -16 \qquad (6)$$
$$a^* = 500[(X/Xn)\ 1/3 -(Y/Yn)\ 1/3] \qquad (7)$$
$$b^* = 200[(Y/Yn)\ 1/3 -(Z/Zn)\ 1/3] \qquad (8)$$

Where

R, G, B = the gray pixel in three components (red, green and blue);

X, Y, Z = the tri-stimulus values in the Commission Internationale de l' Eclairage (CIE1931) system;

k1–k9 = constants (relating to the standard white and the three primary colors);

Xn, Yn, Zn = tri-stimulus values of standard white color;

L* = averaged value of the luminance;

a* = averaged value of the chromaticity channel a

b* = averaged value of the chromaticity channel b.

- Boundary feature

The system applies the Sobel edge detection algorithm, with threshold values 0.1 and 0.5, to find the leaf or flower boundary as shown in Figure 5(c) and Figure 5(d), respectively. After that, the system counts white pixels on each threshold value.

- *Recognition*

This sub-module recognizes a plant by using a leaf or flower image. The system applies two recognition methods, which are 1) Euclidean distance with the average of all features matching values, and 2) Euclidean distance with popular features matching values.

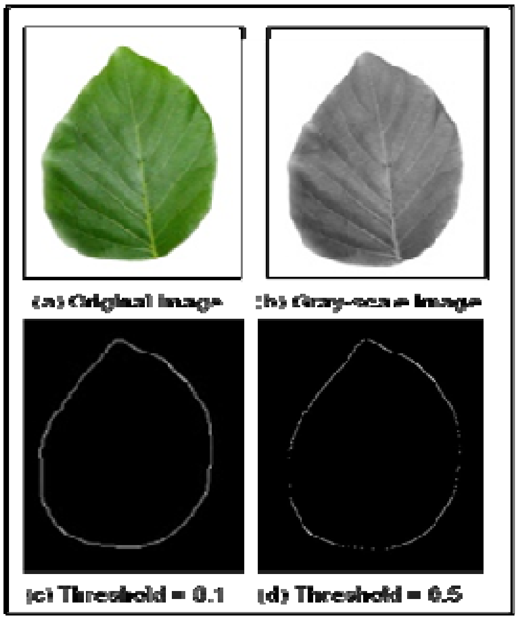


Figure 5. The leaf boundary with thresholds 0.1 and 0.5, respectively.

4) *Display results*

The final stage is a display result module. The system uses MATLAB (R2010a) to develop a graphic user interface (GUI) as shown in Figure 6. There are two image boxes, which are: 1) the display original image box (label number 1), and 2) the recognition image box (label number 2). There are three property text boxes, which are 1) display all features values text box (label number 3), 2) display plant property (label number 4), and 3) add new plant name (label number 5).

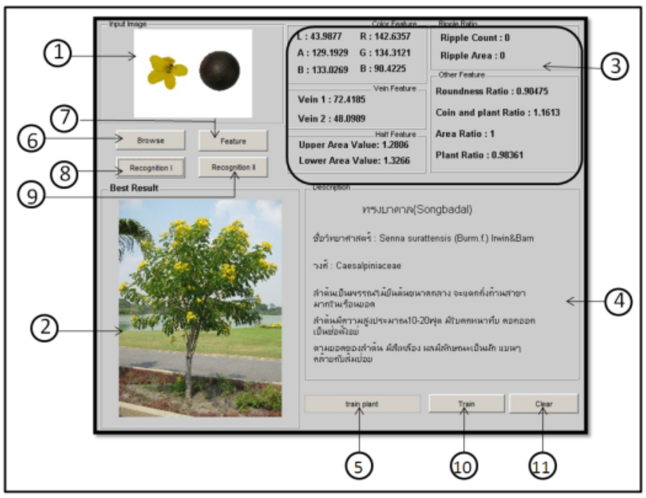


Figure 6. The system user interface screen

Moreover, there are six buttons on the user interface screen, which are: 1) Browse button used for opening a leaf or flower image to be recognized (label number 6), 2) Feature button used for showing all feature values (label number 7), 3) Recognize 1 button used for recognizing the loaded image by applying Euclidean distance with an average of all feature values (label number 8), 4) Recognize 2 button used for recognizing the loaded image by applying Euclidean distance with a popular value in each feature (label number 9), 5) Train button used for training new data and saving it in system database (label number 10), and 6) Clear button used for initializing a new screen and starting to browse a new unknown image (label number 11).

IV. Experiment Results

The system was conducted on 25 species of leafs and 25 species of flowers. The system uses 25 images of each leaf and flower species to train itself and save them in the database. Then, it captures 10 images of each leaf and flower species to test the system performance and efficiency. Moreover, the system applies 5 unknown leaves and 5 unknown flowers to test an un-training data set.

TABLE 1: PRECISION RATE OF THE LEAVES TRAINING DATA SET.

Thai name	Scientific name		Recognition1			Recognition 2		
		N	M	S	U	M	S	U
Borrapetch	Tinospora crispa (L.)	10	10	0	0	10	0	0
Chummalieng	Lepisanthes fruticosa Leenh	10	5	2	3	10	0	0
Fahtalai	Andrographis paniculata	10	10	0	0	10	0	0
Gumpangjedchan	Salacia chinensis L.	10	5	4	1	7	3	0
Gungrao	Fragraea fragrans Roxd	10	9	1	0	4	2	4
Keelek	Cassia siamea (Lamk.)	10	8	1	1	7	3	0
Krapaodang	Ociemum Sanctum L.	10	6	4	0	4	6	0
Kuntongpayabaht	Suregada multiflorum Baill.	10	7	3	1	3	5	2
Lumduan	Melodorum fruticosum Lour.	10	8	2	0	6	2	2
Makhamteth	Pithecolobium Dulce	10	10	0	0	6	4	0
Makwit	Feronia limonia (L.) Swing.	10	6	4	0	5	3	2
Maphood	Garcinia vilersiana Pierre	10	9	0	1	10	0	0
Mayom	Phyllanthus acidus skeels	10	6	3	1	1	2	7
Mangkha	Cynometra iripa Kostel.	10	6	3	1	7	1	2
Namprom	Carissa spinarum L.	10	10	0	0	9	0	1
Nhamtaoton	Averrhoa bilimbi L.	10	7	1	2	7	0	3
Rumpoey	Thevetia peruviana (Pers.)	10	10	0	0	0	4	6
Sa-ae	Capparis horrida Linn.f.	10	8	2	0	9	0	1
Sadao	Azadirachta indica	10	7	0	3	7	0	3
Sarapatpit	Sophora tomentosa L.	10	10	0	0	7	1	2
Sa-yak	Pedilantus tithymaloides (L.)	10	7	1	2	5	1	4
Soiinthanil	Thunbergia grandiflora.	10	4	1	5	6	1	3
Talingping	AverrhoaBilimbiLinn.	10	10	0	0	0	6	4
Teenpednham	Cerbera odollamGaertn.	10	5	2	3	8	2	0
Thonggwaw	Butea monosperma Kuntze.	10	9	0	1	8	1	1
	TOTAL	250	192	34	24	156	47	47

where: N=Number tested, M=Match, S=Mismatch, and U=Unknown

TABLE 2: PRICISION RATE OF THE FLOWERS TRAINING DATA SET.

Thai name	Scientific name		Recognition1			Recognition 2		
		N	M	S	U	M	S	U
Banmairooroey	Gomphrena globosa Linn.	10	10	0	0	10	0	0
Chuanchom	Mock Azalea	10	5	4	1	8	2	0
Dalha	Etlingera elatior	10	2	3	5	10	0	0
Daoroung	Tagetes erecta L.	10	10	0	0	8	1	1
Donya Queen Sirikit	Mussaendaphilippica A.Rich cv.	10	7	3	0	8	1	1
Feungfah	Bougainvillea	10	10	0	0	7	1	2
Khem	Lxora chinensis Lamk. Ixora spp	10	9	1	0	6	4	0
Red-Benjamas	Chrysanthemum morifolium	10	7	0	3	8	0	2
Morningglory	Lpomoea Rorsfalliae. (L.) Roth.	10	6	0	4	8	0	2
Orange-Benjamas	Chrysanthemum morifolium	10	8	0	2	3	2	5
Orange-Feungfah	Bougainvillea	10	8	2	0	7	1	2
Orange-Pangpuay	Catharanthus roseus "G. Don".	10	10	0	0	8	0	2
Peoysiean	Euphorbia millii.	10	8	2	0	8	1	1
Pink-Pangpuay	Catharanthus roseus "G. Don".	10	9	0	1	7	1	2
Pink-Petunia	Petunia x Hybrida hart.Vilm.- Andr.	10	7	1	2	8	0	2
Pink-Rose	Rosa spp.	10	2	6	2	7	1	2
Red-Feungfah	Bougainvillea	10	8	2	0	6	1	3
Red-Orchid	Orchid	10	7	0	3	8	0	2
Red-Pangpuay	Catharanthus roseus "G. Don".	10	9	1	0	7	2	1
Red-Petunia	Petunia x Hybrida hart.Vilm.- Andr	10	6	2	2	9	1	0
Red-Rose	Rosa spp.	10	5	0	5	7	3	0
Songbadan	Senna surattensis (Burm.f.) Irwin&Barn	10	10	0	0	2	3	5
Sritung	Jacaranda	10	6	4	0	8	1	1
Violet-Pangpuay	Catharanthus roseus "G. Don".	10	9	0	1	8	0	2
Yellow-Orchid	Orchid	10	7	0	3	10	0	0
	TOTAL	250	185	31	34	186	26	38

where: N=Number tested, M=Match, S=Mismatch, and U=Unknown

The leaves recognition precision is 76.8 percent and 62.4 percent for recognition methods 1 and 2, respectively (as shown in Table 1). The flowers recognition precision is 74.0 percent and 74.4 percent for recognition methods 1 and 2, respectively (as shown in Table 2). For an un-training data set, the leaf recognition precision is around 88.0 percent and 92.0 percent for recognition methods 1 and 2, respectively (as shown in Table 3). The flower recognition precision is around 52.0 percent and 76.0 percent for recognition methods 1 and 2, respectively (as shown in Table 4). The average access time for leaf and flower recognition is 6.12 seconds/image, and 5.69 seconds/image, respectively.

TABLE 3: PRECISION RATE OF THE LEAVES UN-TRAINING DATA SET.

Thai name	Scientific name		Recognition1			Recognition 2		
		N	M	S	U	M	S	U
Kaena	Dolichandrone rpathacea Schum	5	0	0	5	0	1	4
Kruaorn	Congeo tomentosa Roxb.	5	0	1	4	0	1	4
Madun	Sesbania grandiflora (L.) Pers.	5	0	1	4	0	0	5
Makunkai	*Drypetes roxburghii* (Wall.) Hurusawa	5	0	0	5	0	0	5
Poodjeeb	*Tabernaemontana divaricata* (L.)	5	0	1	4	0	0	5
	TOTAL	25	0	3	22	0	2	23

where: N=Number tested, M=Match, S=Mismatch, and U=Unknown

TABLE 4: PRECISION RATE OF THE FLOWERS UN-TRAINING DATA SET.

Thai name	Scientific name		Recognition1			Recognition 2		
		N	M	S	U	M	S	U
Bannburi	Allamanda cathartica L.	5	0	1	4	0	2	3
Pink-Herbertia	Aloplia drunnondii	5	0	5	0	0	2	3
Preediyatorn	Tabebuia arentea Britt.	5	0	0	5	0	1	4
Peep	Millingtonia hortensis Linn. F.	5	0	4	1	0	0	5
White-Herbertia	Aloplia drunnondii	5	0	2	3	0	1	4
	TOTAL	25	0	12	13	0	6	19

where: N=Number tested, M=Match, S=Mismatch, and U=Unknown

V. Conclusion

In this research, the system is able to fulfill the research objective by extracting seven main features of a leaf or a flower and recognizing it. The system trains 25 kinds of leaves and 25 kinds of flowers with 1,250 images in the system database. The system tests its performance with 25 kinds of leaves and 25 kinds of flowers within 500 images for a training data set. Moreover the system tests another 5 kinds of leaves and 5 kinds of flowers within 50 images for an un-training data set. The precision rate is 71.9 percent for the training dataset and 77.0 percent for the un-training data set. The average access time is 5.91 seconds per image.

This project has many problems and limitations as the following:

1) It is very difficult to take a clear picture without trimming off a leaf or a flower or controlling the environment.

2) Based on a picture taken in an uncontrolled environment, a leaf or flower image needs to have a ping-pong ball put aside it to compare the leaf or flower size.

3) The system can recognize just a very small number of plants, compared with a large number of plants in the world.

The system is just in an initial stage. It needs more time and manpower to train a huge amount of leaf and flower data set used for recognizing the real world's plants with an acceptable precision.

REFERENCES

[1] G.J. Zhang, X.F. Wang, D.S. Huang, Z. Chi, Y.M. Cheung, J.X. Du, and Y.Y. Wan, "A Hypershere Method For Plant Leaves Classification", in *Proceeding of the International Symposium on Intelligent, Multimedia, Video and Speech Processing*, Hong Kong, China, 20-22 October 2004, pp. 165-168.

[2] H. Fu, and Z. Chi, "Combined thresholding and neural network approach for vein pattern extraction from leaf images", *in Proceeding of Vision, Image and Signal Processing*, Vol. 153, Issue. 6, December 2006, pp. 881-892.

[3] H. Fu, and Z. Chi, "A Two-Stage Approach for Leaf Vein Extraction", in *Proceeding on International Conference on Neural Network & Signal Processing*, Nanjing, China, 14-17 October 2003, pp. 208-211.

[4] J. Pan, and Y. He, "Recognition of plant by leaves digital image and neural network", in *Proceeding of International Conference on Computer Science and Software Engineering*, Wuhan, Hubei, China, 12-14 December 2008, pp. 906-910.

[5] S.G. Wu, F.S. Bao, E.Y. Xu, Y.X. Wang, Y.F. Chang, and Q.L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", in *Proceeding of International Symposium on Signal Processing and Information Technology*, Cairo, Egypt, 15-18 December 2007, pp. 11-16.

[6] H. Lin, and H. Peng, "Machine Recognition for Broad-Leaved Trees Based on Synthetic Features of Leaves Using Probabilistic Neural Network", in *Proceeding of International Conference on Computer Sience. and Software Engineering*, Wuhan, Hubei, China, 12-14 December 2008, pp. 871-877.

[7] Y.F. Li, Q.S. Zhu, Y.K. Cao, and C.L. Wang, A leaf Vein Extraction Method Based on Snakes Technique, in *Proceeding of International Conference on Neural Networks and Brain*, Beijing, China, 13-15 October 2005, pp. 885-888.

[8] P. Tzionas, S.E. Papadakis, and D. Manolakis, "Plant leaves classification based on morphological features and a fuzzy surface selection technique", in *Proceeding of International Conference on Technology and Automation*, Thessaloniki, Greece, 15-16 October 2005, pp. 365-370.

[9] X. Gu, J.X. Du, and X.F. Wang, "Leaf Recognition Based on the Combination of Wavelet Transform and Gaussian Interpolation", in *Proceeding of International Conference on Intelligent Computer*, Hefei, China, 23-26 August 2005, pp. 253-262.

[10] L. Han, "Recognition of the Part of Growth of Flue-Cured Tabacco Leaves Based on Support Vector Machine", in *Proceeding of the World Congress on Intelligent Control and Automation*, Chongqing, China. 25-27 June 2008, pp. 3624-3627.

[11] Y.C. Zhang, H.P. Mao, B. Hu, and M.X. Li, "Features Selection of Cotton Disease Leaves Image Based on Fuzzy Feature Selection Techniques", in *Proceeding of the International Conference on Wavelet Analysis and Pattern Recognition*, Beijing, China. 2-4 November 2007, pp. 124-129.

[12] J.S. Casals, C.M. Travieso, J.B. Alonso, and M.A. Ferrer, "Improving a Leaves Automatic Recognition Process Using PCA", in *Proceeding of Inter Workshop on Practical Applications of Computational Biology & Bio Informatics*, Salamanca, Spain, 22-24 October 2008, pp. 243-251.

[13] W. Dake, and M. Chengwei, "The Support Vector Machine-Based Near-Infrared Spectrum Recognition of Leaves Infected by the Leafminers", in *Proceeding of International Conference on Innovation Comp. Information & Control*, Beijing, China, 30 August – 1 September 2006, pp. 448-451.

[14] C. Pornpanomchai, and A. Suppaiboonvong, "Thai Blooming Flower Recognition", in *Proceeding of International Joint Conference on Computer Science and Software Engineering*, Phuket, Thailand, 13-15 May 2009, pp. 219-224.

[15] T. Saitoh, and T. Kaneko, "Automatic Recognition of Wild Flowers", in *Proceeding of International Conference on Pattern Recognition*, Barcelona, Spain, 3-8 September 2000, pp. 507-510.

[16] T. Saitoh, K Aoki, and T. Kaneko, "Automatic Recognition of Blooming Flowers", in *Proceeding of International Conference on Pattern Recognition*, Cambridge, UK, 23-26 August 2004, pp. 27-30.

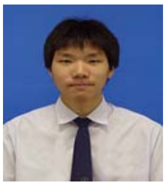
[17] M. Zhenjiang, M.H. Gandelin, and Y. Baozong, "Fourier Transform-Based Image Shape Analysis and Its Application to Flowers Recognition", in *Proceeding of International Conference on Signal Processing*, Beijing, China, 26-30 August 2002, pp. 1087-1090.

[18] J. Zou, and G. Nagy, "Evaluation of Model-Based Interactive Flower Recognition", in *Proceeding of International Conference on Pattern Recognition*, Cambridge, UK, 23-26 August 2004, pp. 311-314.

[19] S.Y. Cho, and P.T. Lim, A Novel Virus Infection Clustering for Flower Images Identification, The *Proceeding of International Conference on Pattern Recognition*, Hong Kong, China, 20-24 August 2004, pp. 1038-1041.

[20] M.E. Nilsback, and A. Zisserman, A Visual Vocabulary for Flower Classification, in *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, USA, 17-22 June 2006, pp. 1447-1454.

Chomtip Pornpanomchai received his B.S. in general science from Kasetsart University, M.S. in computer science from Chulalongkorn University and Ph.D. in computer science from Asian Institute of Technology. He is currently an assistant professor in the faculty of Information and Communication Technology, Mahidol University, Bangkok, Thailand. His research interests include artificial intelligence, pattern recognition and object-oriented systems. Email: itcpp@mahidol.ac.th.



Chawin Kuakiatngam was born in Bangkok Thailand. He received his B.S. in Information and Communication Technology, Mahidol University, Bangkok, Thailand. Email: u5088126@student.mahidol.ac.th.

Pitchayuk Supapattranon was born in Bangkok Thailand. He received his B.S. in Information and Communication Technology, Mahidol University, Bangkok, Thailand. Email: u5088150@student.mahidol.ac.th.

Nititat Siriwisesokul was born in Bangkok Thailand. He received his B.S. in Information and Communication Technology, Mahidol University, Bangkok, Thailand. Email: u5088227@student.mahidol.ac.th.