Combination of Chain Code and Leaf Morphology Approach for Plant Identification

Oktaviani Oktaviani, Sarifuddin Madenda, Rodiah Rodiah,
Diana Tri Susetianingtias, Fitrianingsih, Dea Adlina, Rini Arianty

Gunadarma University, Margonda Raya Street 100, Depok, Indonesia

***Abstract* – This research will identify plants based on 16 basic leaf shapes. It uses 3 leaf datasets, namely Flavia, Folio and Swedish dataset. The researcher used 12 features that consist of 5 basic geometry features and 7 digital morphological features. Geometry features include physiology length, physiology width, diameter, area, and leaf perimeter while digital morphological features involve smooth factors, aspect ratio, circularity, rectangularity, narrow factor, the ratio between perimeter and diameter combined with chain code in determining the ratio between perimeters and lastly, the length and width of physiology. Accuracy result showed that 98.48% of the plants were correctly identified.**

***Keywords* – Chain Code Extraction, Feature, Leaf.**

1. Introduction

More than 30000 plant species in the world have at least one documented usefulness. Most plants–as many as 17810 species– are used as medicine, followed by 11365 species that are used as building materials and textiles [1]. Some plant species are threatened with extinction. In 2017, the International Union for Conservation of Nature and Natural Resources (IUCN) noted in the IUCN Red List of Threatened Species that out of 310442 described plant species there were 12505 endangered plant species. The threatened species are listed as Critically Endangered (CR), Endangered (EN), and Vulnerable (VU). In the same status category, in Indonesia, as many as 437 plant species are declared to be threatened with extinction [2]. In order to increase the potential of plant resources diversity with diverse species, good biodiversity management and utilization is needed. One procedure that can be carried out for protection activities of large species variety of plants is by identifying the types of the plants. In future research, this identification software can be embedded into the plant species identification tool installed on airports.

The most common way to recognize one plant with another is to identify leaf features in plants. Leaf-based identification is the most effective alternative way for identification since the one part of the plant that can be identified regardless of the season is the part of the leaf. The leaf features of each plant have its own characteristics, thus can be used by researchers to identify the types of plants based on leaf features extraction results.

Several studies related to the plant species identification based on leaf characteristics were carried out by previous researchers. Research by [3] identified plants using features extraction from key points on leaf contours. The classification is done by researchers using the bayesian classifier. Research by [4] used image processing techniques to identify fruit trees through leaf structure gray level images approach. Research conducted by [5] in tea leaves detection process used image processing techniques and real-time techniques. Research of [6] implemented a leaf recognition system for plant classification by extracting 21 features from a leaf image. Research by [7] calculates a similarity vector between leaf image feature vectors and feature vectors of each plant class in the database using the Euclidean distance method. The result is that the class of plants that have the smallest Euclidean distance is expressed as the plant class of the test image. Research of [8] compared the use of Artificial Neural Network (ANN) and Euclidean distance

DOI: 10.18421/SAR23-01
https://dx.doi.org/10.18421/SAR23-01

Corresponding author: Rodiah Rodiah,
Gunadarma University, Margonda Raya Street 100, Depok, Indonesia
Email: *rodiah@staff.gunadarma.ac.id*

Received: 25 July 2019.
Revised: 01 September 2019.
Accepted: 07 September 2019.
Published: 28 September 2019.

method (KNN) in classifying plants. The result stated that the classification using Artificial Neural Network shows a higher level of accuracy and performs with a faster time than the classification using the Euclidean distance method. Research by [9] identified plant species using Multilayer Perceptron Neural Network (MLP) with one weight layer, which only had one estimated linear function of input in 6 species from 197 species. Research of [10] conducted a study to identify leaves by extracting different features of leaf images and classifying different classes of leaf images based on features extracted from the leaf image, studying the prominent features of a leaf image, creating leaf identification using Artificial Neural Network. Research by [11] used Invariant Moments and City Block Distance methods to identify the types of medicinal plants based on their leaf shape features. Medicinal plants recognition was also carried out by researchers [12] using the Localized Arc Pattern method based on 42 pattern models formed from a leaf image. Research of [13] identifies plants based on perimeter (circumference) which is the number of pixels that are on the edge of the leaf. The leaf image edge detection conducted by the researcher used the first gradient operator which is done by calculating the first derivative in the discrete spatial coordinates. The researcher also used alternative image leaf features to determine the variability of object edges on leaf images based on roundness of form and the slenderness of leaf image form.

In this research, the identification of plant species will be carried out based on the results of feature extraction from leaf images. Leaf feature extraction was conducted by researchers based on the global features of leaf features. Based on the 16 basic forms of leaf morphology using 5 basic geometry features and 7 digital morphological features combined with chain code. The extraction results are then used to form training sets and testing sets using Support Vector Machine (SVM). Plant species identification based on the combination of chain code, basic geometric features, and digital morphological features is expected to reduce the threat of extinction in species, by identifying each plant species.

2. Method

In this research, the plant species identification process is divided into several process stages. This research consists of three stages: the first stage is the preprocessing stage of data input, namely the leaf images, the second stage is the form feature extraction stage, and the final stage is plant species identification based on the result of leaf feature extraction.

2.1. Leaf Images Dataset

This research uses the Flavia dataset leaf images (source: flavia.sourceforge.net/), Folio Dataset (source: https://data.world/uci/folio), and Swedish Leaf Dataset (www.cvl.isy.liu.se). As an example, leaf image data from the Flavia Dataset, Dataset Folio and Swedish Leaf Dataset can be seen in Figure 1. This research used 56 plant species with a total of 2540 leaf images. The research was conducted based on 16 leaf morphology (Encyclopeddica Britanica, 2007), namely: Acicular, Cordate, Deltoid, Elliptic, Ovate, Oblong, Obovate, Oval, Hastate, Lanceolate, Linear, Lyrate, Obcordate, Reniform, Spatulae and Sagittate.

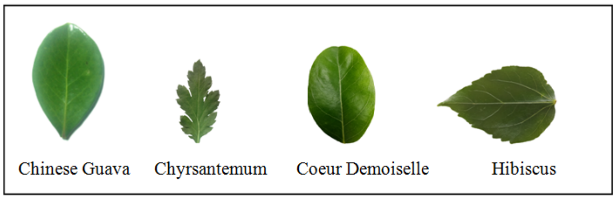


Figure 1. Leaf image data set example

2.2. Leaf Image Preprocessing

The initial stage of leaf image processing is carried out in several stages which are:

1. Extracting the red channel, green channel and blue channel component as can be seen in Figure 2. The leaf image channel extraction process was carried out by the researcher to obtain the grayscale image which is done by calculating the average value of the red, green, and blue components in each pixel of the colored images.

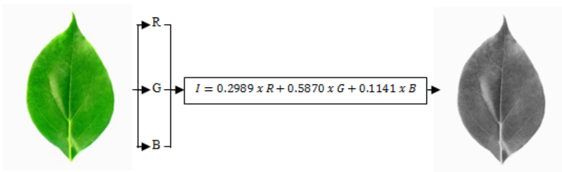


Figure 2. Leaf image preprocessing scheme

2. Performing binarization to separate the leaf image (foreground) from the background to facilitate the process of object identification. Binarization is the process of converting images into binary forms, which only has the numbers 0 and 1. Number 0 represents the background, number 1 represents the foreground leaf. As seen in Figure 3., the image is converted into binary form so that it can be processed at the next stage.

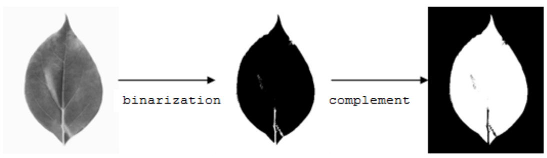


Figure 3. Leaf image binarization scheme

3. Performing filling operations on the resulted complement image of stage two. The image resulted from the binary process producing imperfect binary images, such as the binary image shown in Figure 4., there are black background pixels in the white leaf object pixel area.

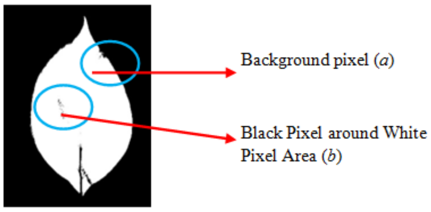


Figure 4. Illustration of filling operation on complement leaf image

Filling operations are carried out to close holes in the object image. The filling operation is done by setting black pixels in the white area (Figure 4b) into object pixels and then checking the above, under, right and left neighbouring pixels as pixel a. Figure 5. shows a comparison of binary complement images before and after the filling process.

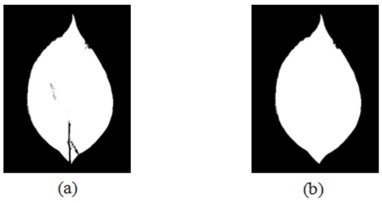


Figure 5. (a). Image before filling operation, (b) Image after filling operation

2.3. Leaf Image Feature Extraction

Morphological characteristics of leaf images are divided into two, namely the basic characteristics and the derivative characteristics. The basic characteristics of leaves consist of [14]: diameter which is the farthest point between the two points of the leaf boundary, physical length which is the distance of the two base points of the leaf, physical width which is calculated based on the longest line length that cuts the physical line length orthogonally, leaf area value and leaf perimeter (circumference). Previous research [11] on plants identification used 7 invariant moment extraction results as leaf feature values. Furthermore, the distance between features is calculated using the City Block Distance method. This study shows that the level of identification proposed by the researchers based on leaf shape is only around 67%.

This research proposes utilization of 12 features in leaf feature extraction consisting of 5 basic geometry features and 7 digital morphological features. The features included in basic geometry features are physiology length, physiology width, diameter, area, and perimeter (circumference) of a leaf, while digital morphological features consist of smooth factors, aspect ratio, circularity, rectangularity, narrow factor, the ratio between perimeters in diameter, and the ratio between perimeters and the physiology length and width. The feature extraction of leaf features is done by the following algorithm:

1. Marking two point coordinates on the image of the leaf. Two point coordinates on leaf images are needed to calculate the physiological length feature. These two points are the base and tip of the leaf which are connected by the main vein of the leaf. Figure 6. shows the location of the two coordinate points that need to be marked.

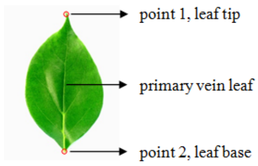


Figure 6. Marking coordinates points on leaf images

The process of marking and retrieving coordinates on the leaves is conducted. The researcher uses parameter value 1 to state the number of taken points. Afterward, the point coordinate value on the *x*-axis stored in the variable x_i, the point coordinates on the *y*-axis stored in the y_i variable are determined which then produce two point coordinates of (x_1, y_1) and (x_2, y_2). Next, the two points coordinates are stored in the *points* variable. Table 1. illustrates the value storing structure in *points* variable in the form of a 2 x 2 matrix.

Table 1. Value storing structure in points variable

	Point 1	**Point 2**
Axis x	x_1	x_2
Axis y	y_1	y_2

2. Calculating the physiology length of the leaf image, where the beginning of the process is done by marking two points on the leaf image. The two points in question are the base and tip of the leaf that is connected by the main vein of the leaf. Figure 7. shows how the two-pixel marking, the end and the base point of the leaf image, is done. From these two pixel coordinates, the physiology length can then be calculated.

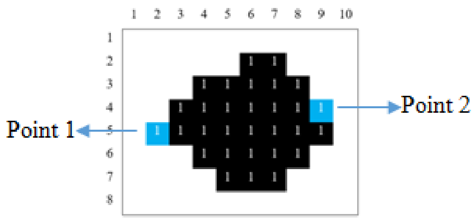


Figure 7. Two starting and end points of leaf image

The physiology length is the marked distance between two coordinate points pixels, calculated using the Euclidean distance formula [15]:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

3. Determining the physiology width of a leaf where the physiology width can also be searched with coordinate points length constituent. Fundamentally, the lines formed from the length points' constituent will be mutually perpendicular to the width line [14] as can be seen in Figure 8.

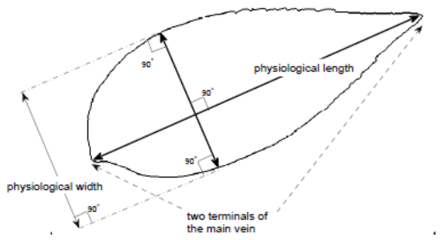


Figure 8. Relation between physiology length and physiology width [14]

First, the line gradient is calculated by utilizing the two coordinates of the physiology length constituent using the formula:

$$L = \frac{(y_2 - y_1)}{(x_2 - x_1)} \quad (2)$$

The width lines are mutually perpendicular to the length lines. Two perpendicular lines have opposed and inversed gradients, thus the gradients for the width lines can be searched utilizing a length line gradient using the formula:

$$W = -\frac{1}{L} \quad (3)$$

Where there is the Length Line Gradient and there is the Width Line Gradient. The search process is carried out by the researchers using Brute force method and calculating the distance and line gradient formed between a contour point and other contour points. The illustration of the obtained length and width line can be seen in Figure 9.

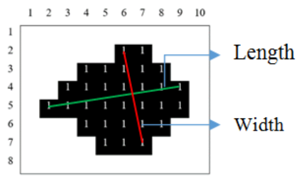


Figure 9. Illustration of length and width line

4. Determining the leaf image diameter, which is the largest distance between two points. Therefore, the search process is the same as the search for physiology width that uses Brute force method by looking at the table of the distance between a point and other points calculation result. The area is calculated with the help of a chain code that has been formed using the following calculation conditions:

   Code 0 : Area = Area + Y
   Code 1 : Area = Area + (Y + 0,5)
   Code 2 : Area = Area
   Code 3 : Area = Area – (Y + 0,5)
   Code 4 : Area = Area – Y
   Code 5 : Area = Area – (Y – 0,5)
   Code 6 : Area = Area
   Code 7 : Area = Area + (Y – 0,5)

5. Determining the basic geometry of the leaf image which consists of:

1) The perimeter of leaf image. It is the circumference or length of the edge of an object. The calculation uses chain code with the formula [16]:

$$P = N_e + N_o\sqrt{2} \quad (4)$$

Where P is the perimeter, N_e is the number of the even chain code and N_o is the number of the odd chain code

2) Smooth factor. It is the comparison between image areas that are smoothed by a $5x5$ rectangular average filter and image areas that are smoothed by a $2x2$ rectangular average filter. Both filters types are shown in the following matrix:

$$\begin{bmatrix} 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 \\ 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 \\ 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 \\ 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 \\ 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 & 0{,}04 \end{bmatrix} \quad \begin{bmatrix} 0{,}25 & 0{,}25 \\ 0{,}25 & 0{,}25 \end{bmatrix}$$

Firstly, the binary image is smoothed using a $5x5$ rectangular average filter, then the value of the area is calculated. Secondly, the binary image is smoothed using a $2x2$ rectangular average filter and then the value of the area is calculated. Lastly, the two area values results are used to get the smooth factor value using the following formula [15] :

$$SF = \frac{F_a}{F_b} \quad (5)$$

Where SF is the smooth factor which is the comparison between the resulted area of filter a (F_a) and the resulted area of filter b (F_b).

3) Aspect ratio. The ratio of physiology length (L) to the physiology width (W) is calculated using the following formula [15] :

$$AspectRatio = \frac{L}{W} \quad (6)$$

4) Circularity is used to compare the shape of a leaf object with the shape of a circle. It is calculated using area values (A) and perimeter values (P) with equation (Madenda, 2015) :

$$Circularity = \frac{4\pi A}{P^2} \quad (7)$$

5) Rectangularity describes the similarity between the shape of a leaf object and a rectangle. (L), physiology width (W) and area (A) by formula [15] :

$$Rectangularity = \frac{LW}{A} \quad (8)$$

6) Narrow factor which is the ratio between diameter (d) and the physiology length (L) which is calculated using the formula [15] :

$$NarrowFactor = \frac{d}{L} \quad (9)$$

7) The ratio between perimeter (P) and diameter (D) is calculated using the formula [15] :

$$Ratio = \frac{P}{D} \quad (10)$$

8) The ratio between perimeter (P) and the physiology length (L) and physiology width (W) is calculated using the following formula [15] :

$$Ratio_{ppw} = \frac{P}{(L+W)} \quad (11)$$

2.4. Leaf Image Identification

The results of leaf image feature extraction will be formed as a training set and testing set, stored in the form of a matrix which structure can be seen in Figure 10.

The training set and testing set storage matrix are stored in a .mat file so that it can be reused during the training and testing process. Matrix saving is done using the following pseudocode:

```
save('trainingset.mat', 'leaf_feature');
```

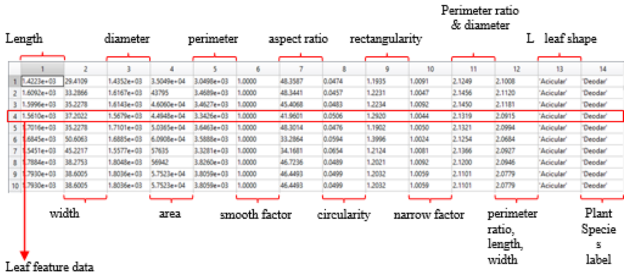


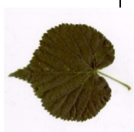
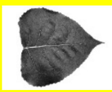
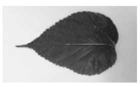
Figure 10. Training set and testing set matrix structure

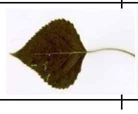
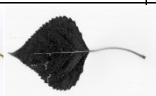
This research uses the LIBSVM library to identify the species types of multi-class plant using the Support Vector Machine method. LIBSVM uses the one-against-one method.

3. Result

Table 2. is the result of leaf image preprocessing which is the step for improving the image by removing noise and getting a binary image from the input image then to be used for form feature extraction.

Table 2. Leaf image preprocessing result

Species	Leaf Image	Grayscale Leaf Image	Filling Operation Leaf Image
Deodar			
Betel			
Betula Pubescens			
Chinese Redbud			
Populus			
Beaumier Du Perou			
Canadian Poplar			
Mulberry Leaf			

Species	Leaf Image	Grayscale Leaf Image	Filling Operation Leaf Image
Tilia			
Camphortree			

Feature extraction produces 12 leaf image form features, which consist of 5 basic geometry features and 7 digital morphological features. In order to get the twelve features, it is necessary to do several processes, namely marking the leaf two points coordinates and forming a chain code.

The two points marked on the cross section of the leaf image are the base and the tip of the leaf that are connected by the main leaf vein, as shown in Figure 11. The process of marking the base and end points of Canadian Poplar leaves produces coordinates (307, 608) and (1447, 533).

Figure 11. Two points marking on leaf Image

3.1. Results of Chain Code Formation

Chain Code is formed using Freeman's algorithm that utilizes the coordinates of leaf image contours that have been obtained in the previous process. By following the contour, the chain code is sequentially obtained in as many as 3629 pieces of code as shown in Figure 12.

Figure 12. Leaf image chain code

3.2. Accuracy Evaluation Results

Evaluation of classification is done using confusion matrix, by calculating the measure of accuracy, precision of each class, and the level of true positive (recall) respectively. For instance, classification with 10 classes or plant species in the test data will produce a confusion matrix. The values on the diagonal of the matrix are true positive, which is the number of test data correctly labeled by classification. Conversely, the values in other positions are the number of test data that is incorrectly labeled by classification. Using the results of the confusion matrix, the accuracy value of identification is 98.48%, where 1.52% of plant species are incorrectly labeled by the classification system.

4. Conclusion

5 basic geometry features successfully extracted from leaf images that include physiology length, physiology width, diameter, area, and perimeter (circumference) of leaf, by marking two point coordinates and edge detection on leaf images using a combination of chain codes (Freeman code). This research succeeded in developing 5 basic geometry features into 7 digital morphological features of a leaf image, consisting of smooth factors, aspect ratios, circularity, narrow factors, the ratio between perimeters and diameters, and the ratio between perimeters and the physiology length and width. Leaf images with 56 plant species were successfully classified using Multiclass Support Vector Machine (SVM) with a classification accuracy value of 98.48%. There are 1.52% of misidentified species due to some similarities in the owned form features value.

Acknowledgements

Thank you to DP2M RistekDikti. We also express our sincere gratitude for Gunadarma University Research Institute.

References

[1]. Royal Botanic Gardens Kew. (2016). Plant Species In The World. Retrieved from: https://www.kew.org/ . [accessed: 10 June 2019].

[2]. IUCN - The International Union for Conservation of Nature and Natural Resources. (2018). Red List of Threatened Species. Retrieved from: https://www.iucnredlist.org/.[accessed: 10 June 2019].

[3]. Gwo, C. Y., & Wei, C. H. (2013). Plant identification through images: Using feature extraction of key points on leaf contours1. *Applications in plant sciences*, *1*(11), 1200005.

[4]. Pande Ankita, V., & Shandilya, V. K. (2013). Digital image processing approach for fruit and flower leaf identification and recognition. *International Journal Of Engineering And Computer Science*, *2*(4), 1280-1285.

[5]. Karunamoorthy, B., & Somasundereswari, D. (2013). A Defect Tea Leaf Identification Using Image

Processing. *Przegląd Elektrotechniczny*, *89*(9), 318-320.

[6]. Lee, K. B., & Hong, K. S. (2013). An implementation of leaf recognition system using leaf vein and shape. *International Journal of Bio-Science and Bio-Technology*, *5*(2), 57-66.

[7]. Sangle, S., Shirsat, K., & Bhosle, V. (2013). Shape based plant leaf classification system using android. *Int J Eng Res Technol*, *2*(8), 1900-1907.

[8]. Satti, V., Satya, A., & Sharma, S. (2013). An automatic leaf recognition system for plant identification using machine vision technology. *International journal of engineering science and technology*, *5*(4), 874.

[9]. Sumathi, C. S., & Kumar, A. S. (2014). Neural network based plant identification using leaf characteristics fusion. *International Journal of Computer Applications*, *89*(5).

[10]. Singh, S., & Bhamrah, M. S. (2015). Leaf identification using feature extraction and neural network. *IOSR Journal of Electronics and Communication Engineering*, *5*, 134-140.

[11]. Suryawibawa, I. W. A., Putra, I. K. G. D., & Wirdiani, N. K. A. (2015). Herbs Recognition Based on Android using OpenCV. *International Journal of Image, Graphics and Signal Processing*, *7*(2), 1.

[12]. Wirdiani, N. K. A., & Sudana, A. A. K. O. (2016). Medicinal plant recognition of leaf shape using Localized Arc Pattern Method. *Int. J. Eng. Technol.*,, *8*(4), 1847-1854.

[13]. Hidayat, T., & Asyaroh, R. N. (2018). Identification of Plant Types by Leaf Textures Based on the Backpropagation Neural Network. *International Journal of Electrical and Computer Engineering*, *8*(6), 5389-5398.

[14]. Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007, December). A leaf recognition algorithm for plant classification using probabilistic neural network. In *2007 IEEE international symposium on signal processing and information technology* (pp. 11-16). IEEE.

[15]. Madenda. S. (2015). Pengolahan Citra & Video Digital: Teori, Aplikasi dan Pemrograman Mengunakan MATLAB. *Erlangga Publisher*.

[16]. RC Gonzalez, RE Woods. (2008). *Digital Image Processing*, Third Edition. Pearson Prentice Hall.