

# Leaf Shape Recognition using Centroid Contour Distance

**Abdurrasyid Hasim<sup>1</sup>, Yeni Herdiyeni<sup>1</sup>, Stephane Douady<sup>2</sup>**

<sup>1</sup>Department of Computer Science, Faculty of Mathematics and Natural Science, Bogor Agricultural University, West Java, Indonesia

<sup>2</sup>Laboratoire : Matière et système complexe, Université Paris Diderot, France

E-mail: abdurrasyid.hasim@gmail.com, yenihardiye@gmail.com

**Abstract.** This research recognizes the leaf shape using Centroid Contour Distance (CCD) as shape descriptor. CCD is an algorithm of shape representation contour-based approach which only exploits boundary information. CCD calculates the distance between the midpoint and the points on the edge corresponding to interval angle. Leaf shapes that included in this study are ellips, cordate, ovate, and lanceolate. We analyzed 200 leaf images of tropical plant. Each class consists of 50 images. The best accuracy is obtained by 96.67%. We used Probabilistic Neural Network to classify the leaf shape. Experimental results demonstrated the effectiveness of the proposed approach for shape recognition with high accuracy.

## 1. Introduction

Plant conservation becomes an urgent necessity for preservation of ecosystems. The growing of human population and climate change increasingly the environmental destruction. Extinction of existing plants may happen in next few years therefore identification of plant is important as conservation needs. Plant identification with traditional approach still be a difficult thing and can only be performed by trained botanists. This has lead to an increasing interest in automating the process of species identification and related tasks. The development and ubiquity of relevant technologies, such as digital cameras and portable computers has bought these ideas closer to reality [1].

In the field of image processing, the leaves are the most appropriate parts of the plant are used for identification as relatively easy to be photographed and analyzed [2]. Leaf is part of plant that is the basic feature in the identification of plants. Types of leaf can be described in terms of leaf shape, leaf margin, leaf venation, leaf texture, leaf vein, leaf color, leaf tip and leaf base. Based on leaf shape, the leaves are grouped into 20 classes [3]. There are two techniques to represent the shape of an object. Contour-based approach that only exploits the information contained in boundary and region-based which involving all parts of an object [4]. Centroid Contour Distance (CCD) is belonging to contour-based approach shape descriptor. Research has been conducted to compare CCD and CCG (Centroid Contour Gradient) in leaf recognition based on leaf tip and leaf base. CCG method is derived from CCD approach. The difference between these two approaches are CCD is used to find out the distance between centroid point and the



boundary points while CCG is used to calculate the gradient between two points on the leaf's contour corresponding to the interval angle. The study included four types of leaf tip and leaf base (acuminate, cuspidate, obtuse, and acute). The average accuracy for recognizing the leaf was 96.6% for CCG and 74.4% for CCD [5].

This research will identify leaf shape. CCD is used to extract the features of the leaf boundary. In classification process, we used Probabilistic Neural Network (PNN) to classify and measure accuracy. PNN is adopted for it has many advantages. Its training speed is many times faster than a BP network. PNN can approach a Bayes optimal result under certain easily met conditions. Additionally, it is robust to noise examples [6].

## 2. Leaf Shape

Botanists observe so many plant features using dichotomous keys just to identify a species. Leaf shape is one of the features to determine plant species. Based on leaf shape, leaves are grouped to 20 classes.

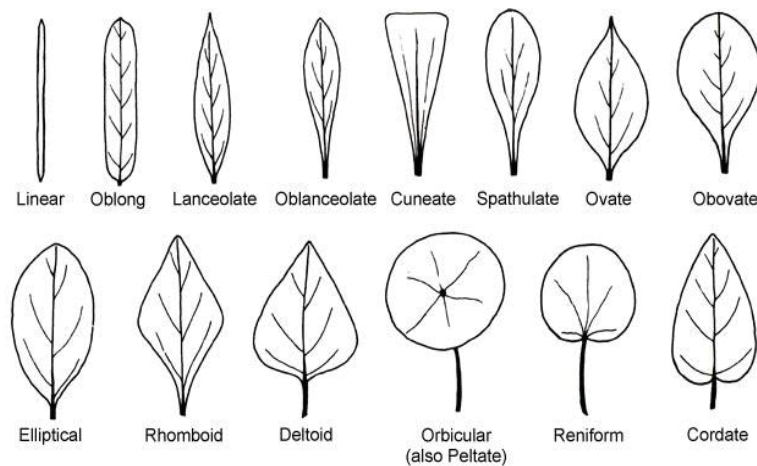


Fig. 1. Examples of leaf shapes

In general, leaf shape characteristics are determined by ratio of leaf height and leaf width and position of the widest part of the leaf [3]. We do not have leaf images of each class therefore we only used the most leaf shape of our collection: elliptical, cordate, ovate, and lanceolate.

Tracing an object contour can be considered as circling around its centroid. Point  $P$  on the boundary is determined by the midpoint of  $C$ , the distance between the point  $P$  and the center point  $C$ , and the angle  $\alpha$ .

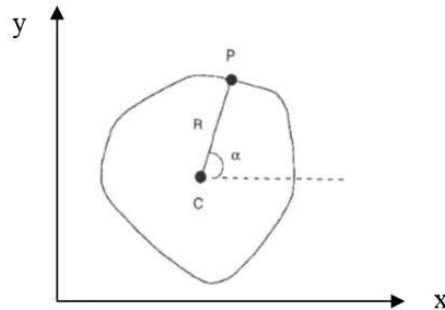


Fig. 2. Illustration of CCD

Centroid-distance function  $R(t)$  is the distance between the boundary points  $(x(t), y(t))$  from the midpoint of the formula (1) [7].

$$R(t) = \sqrt{(x(t) - X_c)^2 + (y(t) - Y_c)^2} \quad (1)$$

#### 4. Method

Our method is to classify the leaf shape with CCD as the feature. We focus on how effective CCD in representing the leaf shape. In this part we explain how to extract the leaf boundary points and use them to determinate the class.

##### 4.1 Experimental Data

We used 200 leaf images of tropical plant in Indonesia. We only used four leaf shapes, i.e., elliptical, cordate, ovate, lanceolate. Each class consists of 50 images. The images were taken from Arboretum of Biopharmaca IPB. The representative's samples of images are shown in Figure 3. The leaf images were not separated from its background and have a random position.



Fig. 3. Samples of taken images from Arboretum of Biopharmaca IPB for each class. From top left, clockwise consecutively, there are ellips, cordate, lanceolate, and ovate.

##### 4.2 Preprocessing

A series of image preprocessing were done to ensure accurate edge and noise-free. In this research the leaf should has a white background and has standing position perpendicular to the horizontal line. At this manual preprocessing, since the leaf images were not separated from its background and have random position, at first we manually separated it from the background and rotated it so the leaf position is perpendicular to the horizontal line. If the leaf has big holes, the holes closed first. The petiol was also

eliminated in order to simplify the process of finding the midpoint.

The images are uniform with length and width of 700 pixels and a resolution of 72 dpi. To change the image to binary Threshold operation is done to change the image into a binary. The leaf image turned to black and this will simplify edge detection process because there will no wrong detection venation as an edge. To ensure there are no slightest noises outside the leaf object, dilation operation was performed. Erosion operation was also performed to close small holes in the leaf object.

#### 4.3 Edge Detection

To perform edge detection, Canny Edge algorithm was used. The image that has been detected edges leaving only the outline of the leaf contour. These images are ready to be processed at the stage of leaf representation.

#### 4.4 Feature Extraction

CCD calculates the distance between each boundary point and the midpoint. Midpoint ( $C_x, C_y$ ) obtained from a simple calculation,  $C_x$  is half of the leaf width and  $C_y$  half of the leaf height. Assuming that the shape of the leaf is always symmetrical, CCD was only done on the right side of the leaf.

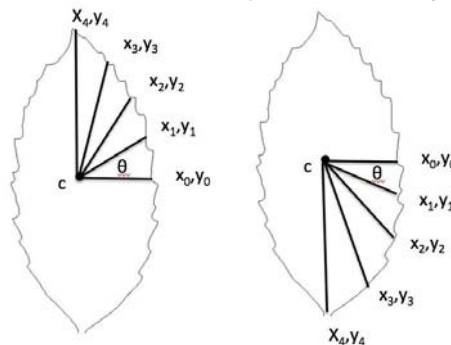


Fig. 4. Illustration of boundary points detection

Point selection was done in two stages, to the top and bottom of leaves. Edge point marked with  $(X_i, Y_i)$  and  $(i = 1, 2, 3, \dots, n-1, n)$  selected only that satisfy equation (2) for the upper leaves and equation (3) for the lower leaf.

$$Y_i = [\tan(\theta) \times (X_i - C_x)] + C_y \quad (2)$$

$$Y_i = [\tan(-\theta) \times (X_i - C_x)] + C_y \quad (3)$$

For the upper leaves, here  $n$  indicates the number of intervals with  $n = (90 / \theta) + 1$ . For example, if the value of  $\theta = 15^\circ$  then the boundary points are selected from the angles  $\theta = \{0, 15, 30, 45, 60, 75, 90\}$ . As for the bottom leaves,  $n = (-90 / -\theta) + 1$ . For angles  $\theta = 15^\circ$  then the selected boundary points are the angles  $\theta = \{0, -15, -30, -45, -60, -75, -90\}$ . Coordinates  $(C_x, C_y)$  shows the central point of the leaf. In this study, we used  $10^\circ$  and  $5^\circ$  for the interval angle. After getting the boundary points, distance was calculated between the midpoint of the distance and the boundary points using the CCD formula.

### 5. Experimental Result

CCD performed with interval angle of  $\theta=10^\circ$  and  $\theta=5^\circ$ . Figure 5 shows with  $\theta=10^\circ$  obtained 19 points, and the  $\theta=5^\circ$  obtained 37 points. The distance between midpoint and boundary points  $R(t)$  calculates consecutively with  $t=1$  at the tip of leaf.

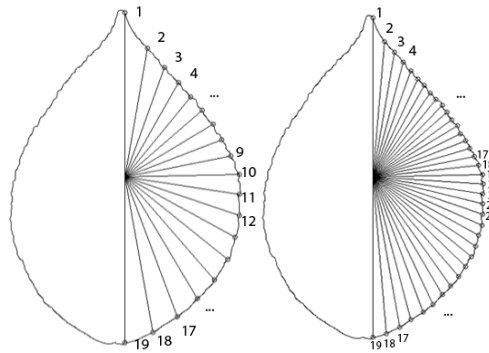


Fig. 5. The number of boundary points depends on interval angle

Figure 6 shows the samples of manually preprocessed leaf images that were used. The images have separated from their background, similar size, and upright position.

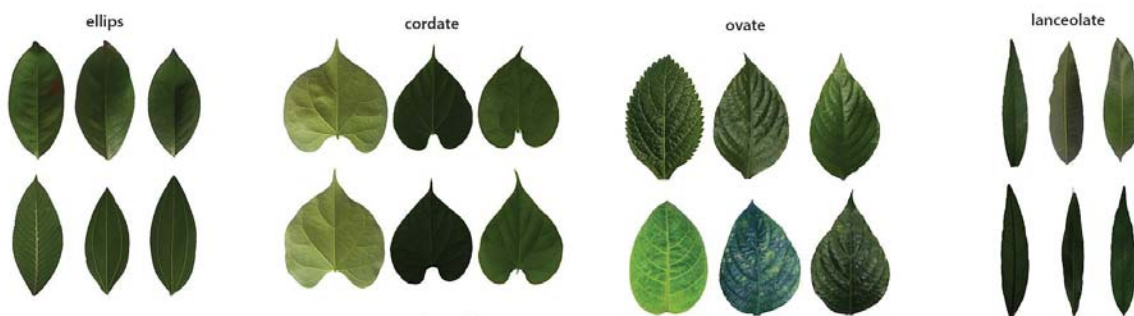


Fig. 6. Preprocessed sample of 4 classes' plant leaves

Graphical representation of the CCD results for four classes leaves with  $\theta=10^\circ$  is shown in Figure 7. Graph shows consecutively distance between midpoint and the boundary points  $R(t)$  with point  $t=1$  starting from a point located at the leaf tip.

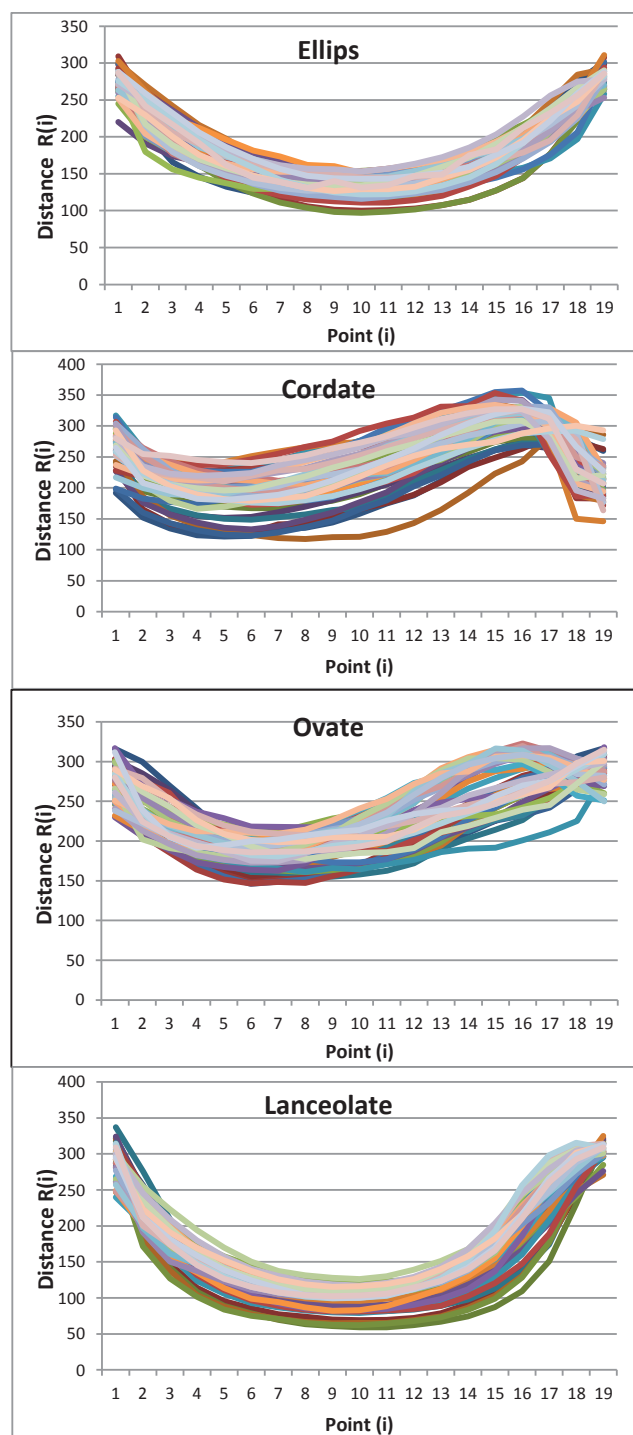


Fig. 7. CCD graphs for each class

In Figure 6, graphics of CCD has a distinctive pattern for each class. The shape of the graph class of

ellips and lanceolate look similar, but lanceolate graph's tend to more concave downward because its shape is thin. The average value of  $R(t)$  for each class shown in Figure 8. Four leaf shape classes that have different graphs show that the CCD feature can distinguish leaf shape well.

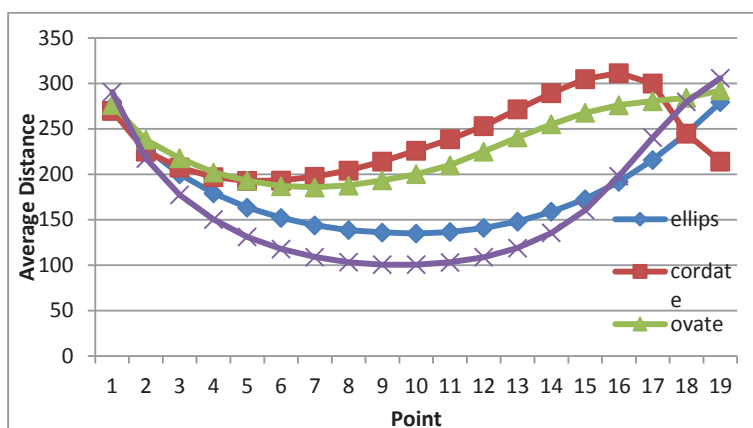


Fig. 8. Average distances of CCD for each class

### 5.1 Classification

Experiment were conducted 15 training data for each class and 140 testing data, the results of the classification of PNN to features of the CCD can be seen in Table 1. The number of CCD features depends on interval angles. Interval  $\theta=10^\circ$  will obtain 19 features and  $\theta=5^\circ$  get 37 features. In order to increase the number of features, those 19 features and 37 features from both interval were merged. In this case, the more features does not increase the accuracy because CCD achieved best accuracy by 96.67% with  $\theta=10^\circ$ .

**Table 1.** PNN Accuracy

Features	19 features ( $\theta=10^\circ$ )	37 features ( $\theta=5^\circ$ )	19+37 features
Accuracy	96.67%	95.00%	93.33%

According to experiment results, CCD are relatively robust under scale change. Figure 9 shows that the distance values of CCD will change if size changed. CCD calculates the distance between midpoint and boundary points. When the leaf size changed, the distance also changed proportionally. Both graph have the relatively similar pattern although the distance values are different.

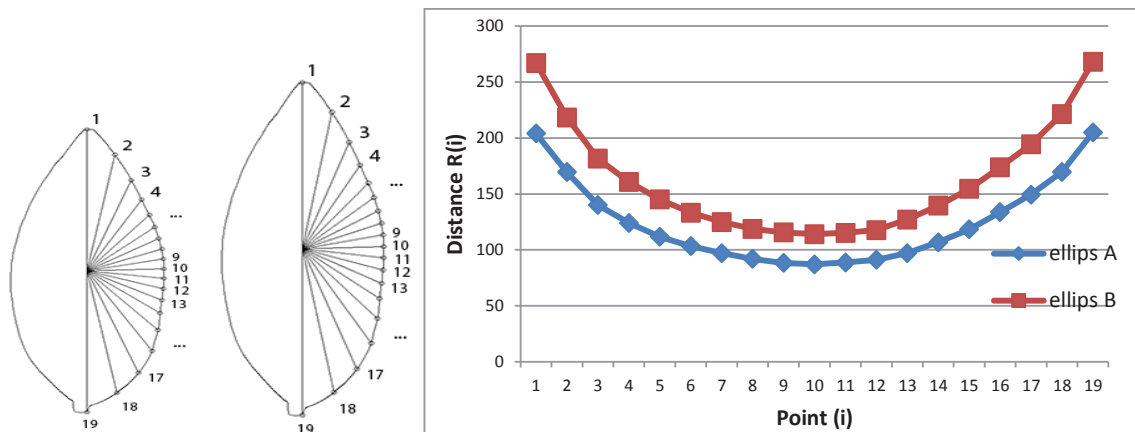


Fig. 9. Ellips A (left) and Ellips B (right) are the same leaf but different in size

## 6. Conclusion

This study successfully implements CCD to represent leaf shape. CCD is good to represent leaf shape because it can distinguish the characteristics of each class. The best accuracy obtained using the CCD 96.67% with interval angle  $\theta=10^\circ$  (19 features). In this experiment, the less interval angle which is the more points extracted does not mean the better accuracy obtained. CCD also relatively robust to the scale change so it could be applicable against leaves with various size.

## 7. Acknowledgment

We specially acknowledge to Prof. Stéphane Douady from Laboratoire: Matière et Système Complexe, Université Paris Diderot for research collaboration. His support is gratefully acknowledged. Also, we are most grateful to a grant from DIKTI (Republic of Indonesia) who made this research possible through International Research Collaboration and Scientific Publication (2014-2015).

## 8. References

- [1] Benson L 1957 *Plant Classification* (Boston : D.C. Heath and Company)
- [2] Bong MF, Sulong GB, and Rahim M S MM 2013 *International Journal of Computer Science Issues* p 477
- [3] Cerutti G, Coquin D, Mille J, Tougne L and Vacavant A 2011 *Conference on Multilingual and Multimodal Information Access Evaluation* p 1
- [4] Cope J S, Cormey D, Clark J Y, Remagnino P, and Wilkin P. 2012 *Expert Systems with Applications* vol 39 p 7562-7573
- [5] Lu G and Zhang D 2004 *Pattern Recognition* 37 p 1-19
- [6] Pahalawatta K 2008 *Plant species biometric using feature hierarchies* (Department of Computer Science and Software Engineering University of Canterbury : Thesis)
- [7] Wu S G, Bao F S, Xu E Y, Wang Y, Chang and Xiang Q 2007 *Signal Processing and Information Technology* IEEE International Symposium