Shape Based Plant Leaf Classification System Using Android

Mr. Shailesh Sangle[1], Mrs. Kavita Shirsat [2], Mrs. Varsha Bhosle [3]
[1] Assistant Professor, Rajiv Gandhi Institute of Technology, Mumbai.

[2] Assistant Professor, Vidyalankar Institute of Technology, Mumbai.

[3] Associate Professor, Vidyalankar Institute of Technology, Mumbai.

Abstract

Plants play an important role in our environment. There are a huge number of plant species worldwide. To handle such volumes of information, development of a quick and efficient classification method has become an area of active research. Designing a convenient and automatic recognition system of plants is necessary and useful since it can facilitate fast classifying plants, and understanding and managing them. Here we describe the development of an Android application that gives users the ability to identify plant species based on photographs of the plant's leaves taken with a mobile phone. In this project, those salient features of plant leaves are studied that may be used as a basis for plant classification and recognition. These features are independent of leaf maturity and image translation, rotation and scaling and are studied to develop an approach that produces the best classification algorithm. The system will be first trained against several samples of known plant species and then used to classify unknown query species.

Keywords

Aspect ratio, rectangularity, sphericity, circularity, eccentricity, form factor

1. Goals

Our main goal is to define an automated plant recognition system, focusing on real-world usability. Our main focus will be on the following:

- **Performance** - we wish to improve on the results of current state-of-the-art methods
- **Portability** - the system should be easily portable, with good down-scalability computational-wise, without loss of performance so that it can provide good results on modern portable devices.
- **Robustness** - small variations in leaf shape, photographing style and background should not affect the outcome of the system.
- **OS independence and Speed** - the system will be written as much as possible from scratch in order to achieve best performance for the given task. Library dependencies will be greatly limited so that the current system would be easily integrated in a multitude of operating systems such as Linux, MS. Windows, iOS or Android.

2. Prior and related work

Many methodologies have been proposed to analyze plant leaves in an automated fashion. A large percentage of such works utilize shape recognition techniques to model and represent the contour shapes of leaves, however additionally, color and texture of leaves have also been taken into consideration to improve recognition accuracies. One of the earliest works [1] employs geometrical parameters like area, perimeter, maximum length, maximum width, elongation to differentiate between four types of rice grains, with accuracies around 95%. Both color and geometrical features have been used in [2] to detect weeds in crop fields employing k-NN classifiers. In [3] the authors propose a hierarchical technique of representing leaf shapes by first their polygonal approximations and then introducing more and more local details in subsequent steps. Fuzzy logic decision making has been utilized in [4] to detect weeds in an agricultural field. In [5] the authors propose a two step approach of using a shape characterization function called centroid-contour distance curve and the object eccentricity for leaf image retrieval. The centroid-contour distance (CCD) curve and eccentricity along with an angle code histogram (ACH) have been used in [6] for plant recognition. The effectiveness of using fractal dimensions in describing leaf shapes has been explored in [7]. In contrast to contour-based methods, region-based shape recognition techniques have been used in [8] for leaf image classification. Wang et al. [9] describes a method based on fuzzy integral for leaf image retrieval. In [10] the authors used an improved CSS method and applied it to leaf classification with self intersection. Elliptic Fourier harmonic functions have been used to recognize leaf shapes in [11] along with principal component analysis for selecting the best Fourier coefficients. In [12] the authors propose a leaf image retrieval scheme based on leaf venation, represented using points selected by the curvature scale scope corner detection method on the venation image and categorized by calculating the density of feature points using non parametric estimation density. In [13] the authors propose a new classification method based on hypersphere classifier based on digital morphological feature. In [14] 12 leaf features are extracted and orthogonalized into 5 principal variables which consist of the input vector to a neural network (NN), trained by 1800 leaves to classify 32 kinds of plants. NNs have also been used in [15] to classify plants based on parameters like size, radius, perimeter, solidity and eccentricity of the leaf shape. In [16] shape representation is done using a new contour descriptor based on the curvature of the leaf contour. Wavelet and fractal based features have been used in [17] to model the uneven shapes of leaves. Texture features along with shape identifiers have been used in [18] to improve recognition accuracies. Other techniques like Zernike moments and Polar Fourier Transform have been proposed in [19] for modeling leaf structures. In [20] authors propose a thresholding method and H-maxima transformation based method to extract the leaf veins for vein pattern classification. Finally in [21] a combination of all image features viz. color, texture and shape, have been used for leaf image retrieval.

3. Proposed system

The algorithm of our system for classifying plant species proceeds as follows. An image of the leaf of interest is acquired and preprocessed to obtain a binary image of the leaf's contour. For several samples of leaves from a given species, morphological features are extracted from the contour image. This data is used to train the algorithm by determining the median value of each feature for a given species. As more samples are added to the training set, the algorithm becomes for effective at identifying a given species. Figure 1 below shows the block diagram for plant leaf recognition.

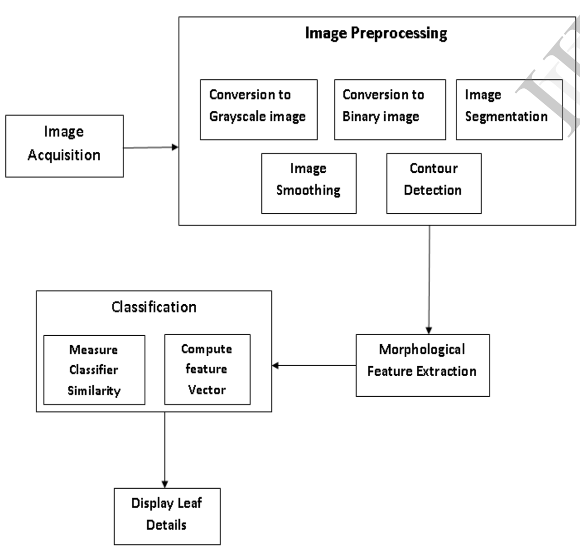

Figure 1 Block Diagram for Plant Leaf Recognition

3.1 Image Acquisition

The input to our algorithm is a photograph of a leaf of unknown species taken with the mobile phones's camera. It is reasonable to assume a certain degree of uniformity in the acquired images, i.e. the picture will be taken at a reasonable distance, in decent lighting, roughly normal to the surface, and against a background which provides sufficient contrast.

3.2 Preprocessing

Before extraction of morphological features can begin, the outline contour of the leaf must be found. The first step in this process is to convert the acquired color image to a grayscale image. Following this, image segmentation is performed to identify leaf pixels and background pixels. After holes have been closed and small regions removed, the segmented image is converted to binary and the interior of the leaf is subtracted, leaving an image of the leaf's outline contour as shown in figure 2.

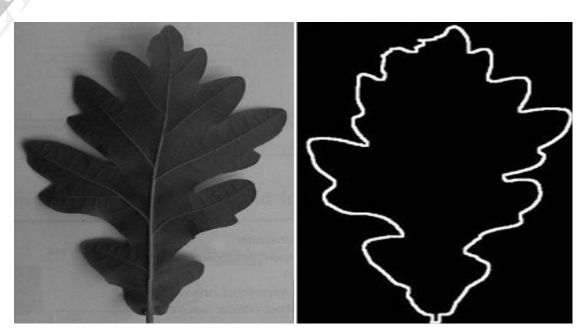

Figure 2 Sample leaf before and after preprocessing

3.3 Morphological Feature Extraction

After an extensive review of literature and experimentation with various combinations of digital morphological features, we decided to include the following features in our final algorithm: aspect ratio, rectangularity, convex area ratio, convex perimeter ratio, sphericity, circularity, eccentricity, form factor, regional moments of inertia, and angle code histogram. The latter two features are based on invariable moments while the rest are geometrical. Each of the features, along with a helper feature, the centroid-contour distance curve is described below.

a. ***Centroid-contour Distance Curve***: The CCD describes the distance from the leaf centroid to the leaf contour as the contour is traced in either a clockwise or counter-clockwise direction. Although the CCD is translation-invariant and can be made scale-invariant by normalization, it requires accurate rotational alignment.

b. ***Aspect Ratio (AR):*** The aspect ratio is the ratio between the maximum length D_{max} and the minimum length D_{min} of the minimum bounding rectangle (MBR).

$$AR = \frac{D\max}{D\min}$$

c. ***Rectangularity (R):*** Rectangularity is defined as the ratio between the region-of-interest (ROI) area and the MBR area

$$R = \frac{A_{ROI}}{(D_{max}/D_{min})}$$

d. ***Convex Area Ratio (CAR):*** The convex area ratio is the ratio of the ROI area and the convex hull area (AC).

$$CAR = \frac{A_{ROI}}{AC}$$

e. ***Convex Perimeter Ratio (CPR):*** The convex perimeter ratio is the ratio of the ROI perimeter (P_{ROI}) and the convex hull perimeter (PC).

$$CPR = \frac{P_{ROI}}{PC}$$

f. ***Sphericity (S):*** Sphericity is the ratio of the radius of the incircle of the ROI (r_i) and the radius of the excircle of the ROI (r_c).

$$S = \frac{r_i}{r_c}$$

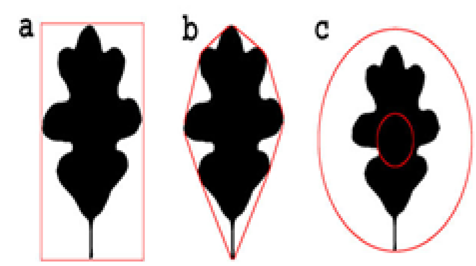

Figure 3 a) MBR, b) Convex Hull, c) Incircle and Excircle

g. ***Circularity (C):*** Circularity is based on the bounding points of the ROI and is the ratio of the mean distance between the center of the ROI and all of the bounding points (μR)and the quadratic mean deviation of the mean distance (sR).

$$C = \frac{\mu R}{sR}$$

h. ***Eccentricity (E):*** Eccentricity is the ratio of the length of the main inertia axis of the ROI (EA) and the length of the minor inertia axis of the ROI (EB).

$$E = \frac{EA}{EB}$$

i. ***Form Factor (FF):*** Form factor is a well-known shape description characteristic given by

$$FF = \frac{4\pi A_{ROI}}{P_{ROI}^2}$$

j. ***Regional Moments of Inertia (RMI):*** Regional moments of inertia (RMI) capture spatial information about the weight distribution of the leaf at different positions along its vertical axis. Because our leaves maintain constant orientation and are generally symmetric about their vertical axis, Ixx quantities for four different regions are used as a descriptor. In order to make these descriptions scale-invariant, leaf area images are first cropped with a bounding box and uniformly resized to a height of 240 pixels before calculations are performed.

$$I_{xx} = \frac{1}{N} \sum_{x_i \in R} (x_i - x_{centroid})^2$$

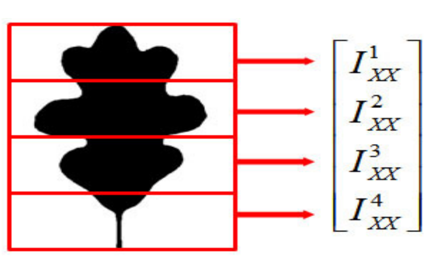

Figure 4 RMI

k. ***Angle Code Histogram (ACH):*** Points along the leaf contour are joined to make line segments and angles between adjacent line segments are measured. A histogram consisting of five uniformly-sized angle bins, each designated by a 1-5 angle "code," is populated with all angles measured along the contour. The histogram is normalized and used as a five-dimensional leaf classifier.

4. Training, Classification, and Matching

Fifteen samples of each leaf class [Ground Nut, Mango, Tulsi, Rose, Ashoka, Gulmohar] were acquired for class training. They were segmented from the background and their contours were extracted in the same manner that query leaves were pre-processed before matching. The five scores for each of the features within a class were averaged to determine the feature vector for the class.

For each morphological feature, a weight was manually assigned. First, the four regional moments of inertia were scaled to 1/4 each, and the five ACH bins were scaled to 1/5 each, since the regional moments of inertia and ACH essentially describe one feature with multiple parts. Total weights for the ten features are shown in Table 1.

To determine a query leaf's correct match, the query leaf's feature vector is computed and its difference from each leaf class represents its feature similarity vector. The feature similarity vectors are then normalized and weighted, and their

Euclidean distance is calculated in the high-dimensional space. The order of the resulting vector of class distances represents the order of similarity.

Table 1 Feature Weights

Aspect Ratio	1
Rectangularity	1
Convex Hull Area Ratio	1
Convex Hull Perimeter Ratio	1
Sphericity	1
Circularity	1
Eccentricity	0.5
Form Factor	1
Regional Moment of Inertia	[0.5, 0.5, 0.5, 0.5]
Angle Code Histogram	[0.4,0.4,0.5,0.4.0.4]

5. Android Implementation

A live viewfinder displays the overlay of the current leaf segmentation. This allows the user to review the results of the segmentation and take the image when the region is acceptably highlighted as shown in Figure 5.

Figure 5 Highlighted Leaf Region

Then the touch-screen based rotation interface lets the user orient the leaf shape in an upright position as shown in Figure 6.

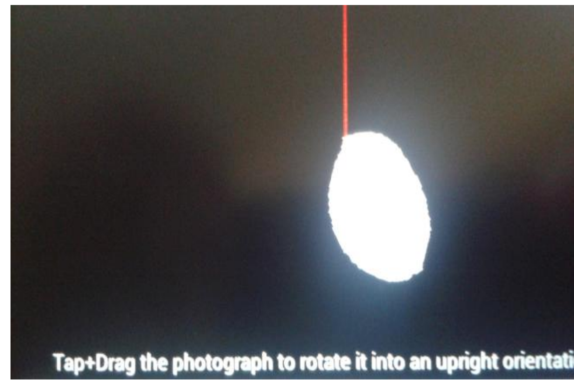

Figure 6 Leaf Orientation in Upright Position

Once the user corrects the leaf shape orientation, a busy animation is displayed while the application generates a feature vector and performs distance calculations against class centers as shown in Figure 7. Finally, the distance between the generated feature vector and each of the six class centers is displayed in a list.

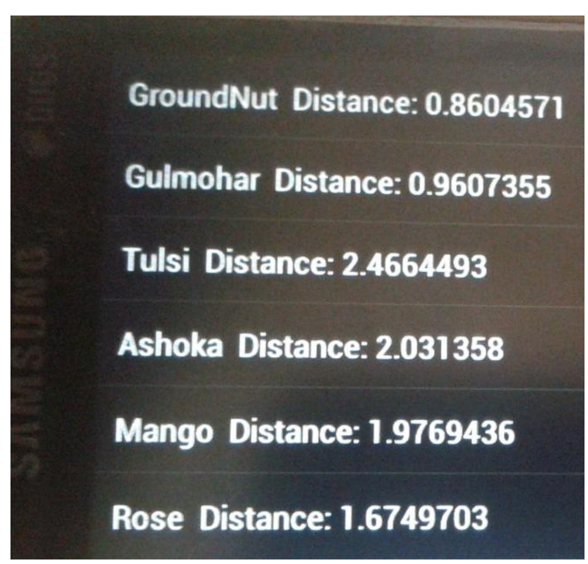

Figure 7 Euclidean Distance

The class with the lowest reported distance is the class the algorithm has identified as the likely species of the query leaf as shown in Figure 8.

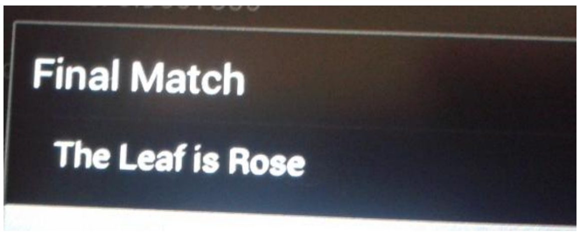

Figure 8 Leaf Match UI

6. Conclusion and Future Scope

In this paper, a digital morphological feature based automatic recognition method for plant images was proposed and performed. We classify plant species based on the nearest neighbor distance of the query leaf's features from the median features of each species in the training set. Our method proved quite robust under reasonable conditions.

To improve usability, an improvement would be to optimize the speed of the live viewfinder segmentation algorithm. The current implementation uses a large amount of iteration that might be avoidable, and the low frame rate and slight delay of the current viewfinder overlay can make it difficult to point precisely at skinny leaf regions.

7. References

[1] N. Sakai, S. Yonekawa, & A. Matsuzaki (1996), "Two-dimensional image analysis of the shape of rice and its applications to separating varieties", *Journal of Food Engineering*, vol 27, pp. 397-407.

[2] A. J. Perez, F. Lopez, J. V. Benlloch, & S. Christensen (2000), "Color and shape analysis techniques for weed detection in cereal fields", *Computers and Electronics in Agriculture*, vol. 25, pp. 197-212.

[3] C. Im, H. Nishida, & T. L. Kunii (1998), "A hierarchical method of recognizing plant species by leaf shapes", *IAPR Workshop on Machine Vision Applications*, pp. 158-161.

[4] C-C Yang, S. O. Prasher, J-A Landry, J. Perret, and H. S. Ramaswamy (2000), "Recognition of weeds with image processing & their use with fuzzy logic for precision farming", *Canadian Agricultural Emgineering*, vol. 42, no. 4, pp. 195-200.

[5] Z. Wang, Z. Chi, D. Feng, & Q. Wang (2000), "Leaf image retrieval with shape feature", *International Conference on Advances in Visual Information Systems (ACVIS)*, pp. 477-487.

[6] Z. Wang, Z. Chi, & D. Feng (2003), "Shape based leaf image retrieval", *IEEE Proceedings on Vision, Image and Signal Processing (VISP)*, vol. 150, no.1, pp. 34-43.

[7] J. J. Camarero, S. Siso, & E.G-Pelegrin (2003), "Fractal dimension does not adequately describe the complexity of leaf margin in seedlings of Quercus species", *Anales del Jardín Botánico de Madrid*, vol. 60, no. 1, pp. 63-71.

[8] C-L Lee, & S-Y Chen (2003), "Classification of leaf images", *16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP)*, pp. 355-362.

[9] Z. Wang, Z. Chi, & D. Feng (2002), "Fuzzy integral for leaf image retrieval", *IEEE Int. Conf. on Fuzzy Systems*, pp. 372-377.

[10] F. Mokhtarian, & S. Abbasi (2004), "Matching shapes with self-intersection: application to leaf classification", *IEEE Trans. on Image Processing*, vol. 13, pp. 653-661.

[11] J. C. Neto, G. E. Meyer, D. D. Jones, & A. K. Samal (2006), "Plant species identification using elliptic Fourier leaf shape analysis", *Computers and Electronics in Agriculture*, vol. 50, pp. 121-134.

[12] J-K Park, E-J Hwang, & Y. Nam (2006), "A vention – based leaf image classification scheme", *Alliance of Information and Referral Systems*, pp. 416-428.

[13] J. Du, X. Wang, & G. Zhang (2007), "Leaf shape based plant species recognition", *Applied Mathematics and Computation*, vol. 185, pp. 883-893.

[14] S. G. Wu, F. S. Bao, E. Y. Xu, Y-X Wang, Y-F Chang, & Q-L Xiang (2007), "A leaf recognition algorithm for plant classification using probabilistic neural network", *The Computing Research Repository (CoRR)*, vol.1, pp. 11-16.

[15] J. Pan, & Y. He (2008),"Recognition of plants by leaves digital image and neural network", *International Conference on Computer Science and Software Engineering*, vol 4, pp. 906 – 910.

[16] C. Caballero, & M. C. Aranda (2010), "Plant species identification using leaf image retrieval", *ACM Int. Conf. on Image and Video Retrieval (CIVR)*, pp. 327-334.

[17] Q-P Wang, J-X Du, & C-M Zhai (2010), "Recognition of leaf image based on ring projection wavelet fractal feature", *International Journal of Innovative Computing, Information and Control*, pp. 240-246.

[18] T. Beghin, J. S. Cope, P. Remagnino, & S. Barman (2010), "Shape and texture based plant leaf classification", *International Conference on Advanced Concepts for Intelligent Vision Systems (ACVIS)*, pp. 345-353.

[19] A. Kadir, L.E. Nugroho, A. Susanto, & P.I. Santosa (2011), "A comparative experiment of several shape methods in recognizing plants", *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 3, no. 3, pp. 256-263

[20] N. Valliammal, & S. N. Geethalakshmi (2011), "Hybrid image segmentation algorithm for leaf recognition and characterization", *International Conference on Process Automation, Control and Computing (PACC)*, pp. 1-6.

[21] B. S. Bama, S. M. Valli, S. Raju, & V. A. Kumar (2011), "Content based leaf image retrieval using shape, color and texture features", *Indian Journal of Computer Science and Engineering*, vol. 2, no. 2, pp. 202-211.