

A Review and a Comparative Study of Various Plant Recognition and Classification Techniques using Leaf Images

Anand Handa

Computer Science & Engineering Department
PSIT College of Engineering, Kanpur (India)

Rashi Agarwal, PhD

Department of Information Technology
UIET, CSJM University, Kanpur (India)

ABSTRACT

This paper compares various algorithms that are used for plant classification based on leaf images are presented. The paper reviews the main computational, morphological and image processing methods that have been formulated in recent years. At the end, we conclude with the ongoing work in the present area and the other existing problems in the area. The automatic digital plant classification can be done by extracting various features from its leaves and still there exist possibilities to improve plant species identification through the designing of a new digital automatic plant identification and recognition system. In general different ways can be used with following major steps. Firstly, leaf images are acquired with high quality digital cameras and scanners then the user has to select a base point and reference points are marked on the leaf blades. This step leads to the extraction of several morphological features and these features are used as inputs to a machine learning algorithm. The use of the Probabilistic neural network is done for the above process and for this the network is trained with leaves from different plant species. At last the recognition rate and accuracy are tested for each proposed methodology. All the specified methods work only for plant leaves that are broad and flat or they are more or less two dimensional in nature.

Keywords

Plant Leaf Classification, PNN, PCA, Texture Analysis and Radial Basis Function, Moments Invariants, Neural Networks.

1. INTRODUCTION

Plants play an important role in various areas such as medical sciences, industrial growth and the environment. Around 3.5 million plant species exist on the planet and they are closely related to the human beings. However, due to rapid urbanization and development of human society has disturbed the balance to a great extent resulting in the endangerment of many plant species. Hence it is one of our foremost responsibilities to save the plants from various threats and must restore the diverseness of plant family and the ecosystem. Therefore, it is very important to develop a highly sophisticated automatic plant classification and recognition system for plant identification and protection. Leaf identification methodologies plays an important role in plant classification and the important issue lies in the selection of the features for leaf detection. They must be stable and have a good ability to identify and differentiate various kinds of leaves.

The automatic recognition and classification of the plants is the first step towards the conservation of flora and fauna. It is always very difficult to identify a plant correctly by just feeling and looking at it as there are so many kinds of plants that are unknown to humans on earth. Recent classifications are done on the basis of flowers, leaves, barks, etc. in order to deal with such a large volume of information, realization and development of a quick and efficient identification and classification method has

become an area of study. A correct and appropriate way to classify plant is to recognize them on the basis of the features extracted from the plant leaf images. The two most important features which are used for plant recognition based on leaf images are color and shape [2], [6-9]. An automatic plant identification system is of great importance in botanical gardens or natural reserves, plant taxonomy and for new plant species discovery.

An automated plant identification nor classification system can use various different characteristics of the flora which can start from a very basic level, such as shape and color of the leaf, flower and its fruit type, its branching style, root type, seasonality, outlook, etc. to a very complex level such as tissue and cell structure and genetic structure. In present scenario the cell phones are capable of capturing high quality images with their integrated digital camera which makes the usability of such a system even wider and more. For the classification of plant a set of appropriate numerical attributes of features is extracted. In recent years the research on the utilization of moments for object characterization in both in variant and non variant objectives has gained attention [8, 9]. The concept of moments has been used in various fields ranging from engineering mechanics and statistics to image processing and pattern recognition.

2. TECHNIQUES USED IN PLANT RECOGNITION

A substantial amount of work has been done on leaf shape based plant classification, recognition and identification, but still there is a possibility of more investigations to develop more robust plant identification systems. The performance of the designed systems may vary from sample to sample. Sometimes it is also possible that two or more plants have leaves with similar shape but may vary in colors. In such a case color, features cannot be neglected. A typical image based plant identification system is shown in Fig. 1.

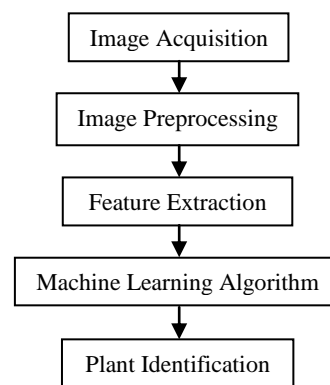


Fig 1: Plant Identification System

Wang et al. [2] proposed a method based on centroid contour distance (CCD) curve, eccentricity and angle code histogram (ACH). They analyzed 1400 leaf images gathered from 140 plants. Their experimental results show that the proposed algorithms can achieve a better retrieval performance rate than both the earlier discussed methods which were curvature scale space (CSS) method and modified Fourier descriptor (MFD) method.

Fu et al. [3] proposed an ontology based leaf classification system that uses the centroid contour distance curve for the representation of leaf shapes and images. They also used machine learning algorithms for the development of an automatic leaf retrieval system.

Wang et al. [4] proposed an efficient computer aided plant species identification (CAPSI) approach which is based on recognition of plant leaf images using a shape matching methodology. They extracted seven geometric moments and sixteen Zernike moments to represent a leaf shape. The above is carried out in following stages. Firstly, an approximation algorithm, namely Douglas-Peucker is used and a new shape representation is considered to form the sequence of invariant attributes. This approximation is categorized as polygon approximation, then a modified dynamic programming (MDP) for shape matching and recognition is used for leaf recognition. In second stage, the experimental facts and results are gathered. The experiment used 50 leaf images which were randomly selected from our image database as the query images and each query can retrieve at most 20 such similar images from the database.

Gu et al. [5] also used an approximation algorithm in order to extract leaf shape features. Gu et al. used the result of leaf segmentation which is based on wavelet transform (WT) and Gaussian interpolation. It is a new and efficient approach for leaf classification. The approach uses various classifiers, a nearest neighbor classifier i.e. 1-NN, a radial basis probabilistic neural network (RBPNN), a k –nearest neighbor classifier (k-NN) are used. The above mentioned features are based on run length features (RLF) which are extracted from the skeleton to recognize the leaves.

Wang et al. [6] for the classification of leaf images used Centroid Contour Distance (CCD). They extracted various geometric features like rectangularity, circularity, eccentricity and seven other moment invariants for the classification of leaf images. The following seven functions are computed and they form central moments with respect to object scale, translation and rotation:

$$\emptyset_1 = \mu_{20} + \mu_{02} \quad (1)$$

$$\emptyset_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (2)$$

$$\emptyset_3 = (\mu_{30} - 3\mu_{12})^2 + 3(\mu_{21} + \mu_{03})^2 \quad (3)$$

$$\emptyset_4 = (\mu_{30} - \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad (4)$$

$$\emptyset_5 = (\mu_{30} + 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} - \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (5)$$

$$\emptyset_6 = (\mu_{20} - \mu_{02}) [(\mu_{30} + 3\mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11} (\mu_{30} + \mu_{12}) - (\mu_{21} + \mu_{03}) \quad (6)$$

$$\emptyset_7 = (3\mu_{21} - \mu_{03}) (\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - (\mu_{30} - 3\mu_{12})(\mu_{21} - \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (7)$$

Apart from the above mentioned features 15 more features are extracted from the preprocessed leaf. In that paper an iterative threshold selection, segmentation method is used to address those leaf images which are having simple background. In case of leaf

images with complicated background Marker controlled watershed segmentation method is selected. In addition to the above specified parameters a moving center hypersphere (MCH) classifier based on shape features is proposed for classifying a large number of leaves. The use of conventional classifiers like K-NN or Neural Network tends the classification process to be time and space consuming. The class of patterns as a series of “hyper-spheres” is used in new approaches while in earlier approaches these patterns from one class are all treated as a set of points. The experimental results classified more than 20 classes of plant leaves successfully with an average recognition rate up to 92.2%

Du et al. [7] applied leaf shape recognition based on radial basis probabilistic neural network (RBPNN). Orthogonal least square algorithm (OLSA) is further applied to train the system and further it is optimized by recursive OLSA. It performs plant recognition and classification through modified Fourier descriptors. The advantage is that it can identify the type of plant from a partially damaged leaf. A set of unique features is used and experimental result shows that the proposed methodology achieves a higher recognition rate and efficiency than both the approaches, namely radial basis function neural network (RBFNN), BP neural network (BPNN).

Hu et al. [8] proposed a methodology based on the combination of wavelet transform (WT) and Gaussian interpolation. They used the result of segmentation of leaf skeleton. Hu [8] proposed the utilization of moment invariants for image analysis and representation of the object in 1961. Hu’s Uniqueness Theorem states that if $f(x,y)$ is piecewise continuous and has non zero values in the finite part of the $f(x, y)$ plane, then geometric moments of all orders exist. It can then be shown that the moment set $\{mpq\}$ is uniquely determined by $f(x, y)$ and conversely $f(x,y)$ is uniquely determined by $\{mpq\}$.

Wang et al. [9] extracted various geometric features like rectangularity, circularity, eccentricity and 7 moment invariants for plan classification. T. H. Resis [9] stated that moment used by Hu is incorrect. The four moments under general linear transformation shows an error.

Du et al [10] proposed shape detection based on radial basis probabilistic neural network, which is trained by OLSA and which is further optimized by recursive OLSA. The modified Fourier descriptors of leaf shape are used for plant recognition. Maitre [10] modified the seven moment invariant proposed by Hu which are independent of change. The categories of change were transformation, scale, rotation and contrast also.

A curvature scale space image is used by Mokhtarian and Abbasi [11] for the representation of leaf shapes and is applied for leaf classification. Similarly a new set of moment invariants with respect to rotation, translation and scaling is proposed by Jan Fusser [11]. The proposed methodology is suitable for recognition of objects having fold rotation symmetry.

Wu et al. [12] extracted digital morphological features which were further orthogonalized into five principal variables. The classification of 32 varieties of plant is done by using PNN through a sample of 1800 leaves. The leaf image is firstly converted into a binary image from the RGB image using the given formula. Wu et al [12] used twelve different digital morphological features which were further orthogonalized into five principal variables. They used a set of 1800 leaves for classification of 32 kinds of plants and to identify each kind of plant at least 10 pieces of leaves from the test samples are considered for checking the accuracy of the algorithm that came out to be 90.312 %.

$$\text{Gray} = 0.2989 * \text{Red} + 0.5870 * \text{Green} + 0.1140 * \text{Blue} \quad (8)$$

After this boundary enhancement can be done with the laplacian filter of following 3 x 3 spatial masks:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

After the application of spatial masks we obtain a set of geometric features such as Diameter, Physiological length, Physiological width, leaf area and leaf perimeter and after that 12 morphological features as aspect ratio, form factor, rectangularity, narrow factor, perimeter ratio of Physiological length, Physiological width and 5 vein features were extracted for leaf classification and recognition. PCA is further used for transforming the data into a new coordinate system.

Du and Zhang [13] proposed a new methodology for plant classification and leaf recognition. The method is named as move median centers (MMC) hypersphere classifier. They explained the algorithm of MMC, classification stage and the Data preprocessing for the application of the MMC. The advantage of MMC classifier is that it not only saves the storage space but also reduces the classification time. In comparison to 1-NN and k-NN classifiers the method is more robust as it is not dependent on the contour features.

Hossain and Amin [14] proposed a computationally effective technique for plant species recognition using a leaf image. The proposed methodology is only applied to the plants with wide flat leaves are generally two dimensional in nature. There are five major phases. Firstly, leaf images are captured and then the base point of the leaf and a small number of reference points on the leaf blades are decided by the user. A binary image is generated according to the above reference points on the leaf blades. After this stage the leaf is aligned horizontally by keeping its base point to the left of the image. Then area, major axis, minor axis, eccentricity, perimeter, equivalent diameter, convex area and extent are obtained as done before. A distinguished collection of features is obtained from the leaves by segmenting it across the major axis and parallel to the minor axis. In the next stage the feature points are normalized by considering the ratio of the slice lengths and leaf lengths (major axis). Then these features are provided as inputs to the PNN.

Prasad et al. [16] proposed a new technique for feature extraction from a plant leaf image. An automated living plant species identification is proposed which would be helpful for botanical students to carry out their research for plant species identification. A novel multi-resolution and multidirectional Curvelet transform is executed on sub segmented leaf images to obtain leaf information. The accuracy rate can be enhanced if the orientation of the object in the image is not taken into account. These coefficients are given as the input to a trained SVM classifier for categorization of the result.

Abdul kadir [17] proposed foliage plant identification systems. The proposed methodology can handle two or more plants that have the same or may be similar shape, but the color patterns on the leaves must be different. In case of same color patterns Zernike moments were used in combination with features like color moments, geometric features and gray-level co-occurrence matrix (GCLM). In order to implement the system two different approaches were designed. In the first approach a distance measure is used and in the second approach PNN is used. The experimental results have proven that Zernike Moments have a better prospect as features in leaf identification and recognition systems when they are combined with other features.

An automated leaf recognition system has been proposed by J. Chaki and R. Parekh [18] based on leaf images. It specifically uses two modeling technique, namely Moments Invariant (M-I) and Centroid Radii (C-R) model. In case of MI, the first four normalized central moments have been considered and studied in various combinations. In addition, there are 7 moment features that can be used to describe shapes and these are invariant to rotation, translation and scaling. In case of CR model, the identification of the boundary of the leaf shape is done by an edge detector. Each pixel is identified by either its color or by the x and y coordinates on a two dimensional plane. The classifier that was used for discrimination is a neural network.

An automated system based on the leaf vein shape for plant identification was proposed by Lee and Hong [19]. The proposed methodology uses major vein and frequency domain data. FFT methods are used with distances between contour and centroid on detecting leaf. In total 21 features were extracted for the leaf recognition, which are the distance feature between centroid and all points of leaf contour. Frequency domain data by FFT that was performed using the distances and 10 features of all 21 leaf features were extracted using morphological features using 4 basic geometric features and 5 vein features and the last 1 feature was extracted using the convex hull.

Leaf recognition of plant species based on ANN was proposed by Sajeevan G [20]. The leaf shape features such as aspect ratio, apex ratio, base angle, centroid deviation ratio, moment ratio and circularity were extracted with the help of image processing techniques. These features are used as input parameters to the neural network for classifying the plants. The experiments were performed on a set of 534 leaves of 20 kinds. Out of these, 400 leaves were trained. The 134 testing samples were recognized with an accuracy rate of 92 %.

Rahim et al. [21] proposed normalization of the tip and base of leaf as both of them influence data extraction process. Center Contour Gradient CCG is used as the extraction method which calculates the gradient between pairs of boundary points corresponding to interval angle. The experimental results have shown that CCG has outperformed CCD (Centroid Contour Distance) as it captures the curvature of the tip and base of leaf. CCD is used to find out the distance between centroid point and the pixel's on the leaf's contour point, whereas CCG is used in calculating the gradient between two pixels on the leaf's contour corresponding to the interval angle. The accuracy in classifying the tip of the leaf using CCG is 99.47%

A simple method for leaf recognition based on bisection of leaves was proposed by Caner Uluturk and Aybars Ugur [22]. After preprocessing 7 low cost morphological features are extracted. In the next stage leaf orientation is done about its major axis and the two parts are acquired by slicing leaf on its centroid vertically. Area, extent and eccentricity features are extracted for each part and their proportions to each other are taken as new features. These all 10 features are used as an input to probabilistic neural network (PNN). PNN is trained with 1120 leaf images from 32 different plant species which are taken from FLAVIA dataset. Nearly 160 leaf images of the plant species are used for testing. The experiments and comparisons show that method based on half leaf features has reached one of the best results with 92.5% recognition accuracy. An Overview of the various plant classification techniques:

Table 1: Comparison of Various Plant Classification Techniques

S. No	Researchers	Leaf Features	Algorithm Used	Accuracy
1	Z. Wang, Z. Chi, and D. Feng	Centroid Contour Distance, eccentricity and angle code histogram.	Curvature scale space (CSS) and modified Fourier descriptor (MFD) method.	An average recall rate of 75.6% was achieved
2	H. Fu, Z. Chi, D. Feng, and J. Song	Vein Texture is extracted	Leaf classification and centroid-contour distance curve method	94.26%
3	J. Du, D. Huang, X. Wang, and X. Gu	Shape of plant leaf is used as identification tool.	Computer-aided plant species identification (CAPSI) approach with Douglas-Peucker approximation and modified dynamic programming (MDP) algorithm	
4	X. Gu, J. Du, and X. Wang,	The result of segmentation of leaf's skeleton based on the combination of Wavelet transform and Gaussian interpolation.	Combination of wavelet transform (WT) and Gaussian interpolation, k-nearest neighbour classifier (k-NN) and a radial basis probabilistic neural network (RBPNN) are used with run-length features (RLF)	Close to 95%

5	X. Wang, J. Du, and G. Zhang	Features like rectangularity, circularity, eccentricity and seven moment invariants are used	hyper sphere classifier	92.20%
6	J. Du, D. Huang, X. Wang, and X. Gu	It performs plant recognition through modified Fourier descriptors of leaf shape.	Radial basis probabilistic neural network which is trained by orthogonal least square algorithm (OLSA) and optimized by recursive OLSA.	96.20%
7	S. Wu, F. Bao, E. Xu, Y. Wang, Y. Chang, and Q. Xiang	12 digital morphological features are used. The base point of leaf and a few reference points on the leaf blades are also used.	PCA technique and training, testing with Probabilistic Neural Network	90.31%
8	J.-X. Du, X.-F. Wang and G.-J. Zhang	Digital Morphological features are used as parameter to the algorithms	Move median centers (MMC) hypersphere classifier	The recognition rate is greater than 75%
9	Hossain and M.A. Amin	Features like eccentricity, area, perimeter, major axis, minor axis, equivalent diameter and convex area	Features are selected after 5 major phases and used as inputs to the probabilistic neural network	91.41%

10	S. Prasad, P. Kumar and R.C. Tripathi	Feature used for extraction is shape and texture	New multi-resolution and multidirectional curvelet transform is applied	95.60%
11	Abdul Kadir, L.E. Nugroho, A.susanto, P.Insap Santosa	Zernike moments were combined with other features like geometric features, color moments and gray level cooccurrence matrix.	A distance measure and Probabilistic Neural Networks are used.	92.20%
12	Jyotismita Chaki, Ranjan Parekh	Moments Invariant and Centroid Radii are used.	Neural networks are used as classifiers for discrimination	Ranging from 90-100 %
13	Kue-Bum Lee and Kwang-Seok Hong	Leaf Vein Shape	frequency domain data by using FFT methods	97.19%
14	Sajeevan G	Aspect Ratio, apex ratio, base angle, centroid deviation ratio, moment ratio and circularity.	ANN is used to identify plant by inputting leaf image	92.00%
15	Mei Fern Bong, Ghazali bin Sulong, Mohd Shafry Mohd Rahim	The tip and base of leaf are used as parameters to the normalization method.	CCG is used which calculates the gradient between pairs of boundary points corresponding to internal angle.	99.47%

16	Caner Uluturk and Aybars Ugur	Leaf area, extent and eccentricity features are extracted for the bisected leaf.	It proposes a simple method based on bisection of leaves for recognition.	92.50%
----	-------------------------------	--	---	--------

3. CONCLUSION

In this paper we have discussed several plants identification systems and leaf recognition system. It must be clear from the above comparative study that there is no single method that provides a solution for all problems but this analysis helps in identifying the best approach for leaf features extraction and classification. The accuracy measure helps in analyzing the performance of the system. Hence, other features will be researched and extracted in future for better performance.

4. REFERENCES

- [1] R. C. Gonzalez, R. E. Woods, "Digital Image Processing". Prentice Hall, 2004.
- [2] 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.
- [3] H. Fu, Z. Chi, D. Feng, and J. Song, "Machine learning techniques for ontology-based leaf classification," IEEE 2004 8th International Conference on Control, Automation, Robotics and Vision, Kunming, China, 2004.
- [4] J. Du, D. Huang, X. Wang, and X. Gu, "Computer-aided plant species identification (CAPSI) based on leaf shape matching technique," Transactions of the Institute of Measurement and Control. 28, 3 (2006) pp. 275-284.
- [5] X. Gu, J. Du, and X. Wang, "Leaf recognition based on the combination of wavelet transform and gaussian interpolation," in Proceedings of International Conference on Intelligent Computing 2005, ser. LNCS 3644. Springer, 2005.
- [6] X. Wang, J. Du, and G. Zhang, "Recognition of leaf images based on shape features using a hypersphere classifier," in Proceedings of International Conference on Intelligent Computing 2005, ser. LNCS 3644. Springer, 2005.
- [7] J. Du, D. Huang, X. Wang, and X. Gu, "Shape recognition based on radial basis probabilistic neural network and application to plant species identification," in Proceedings of 2005 International Symposium of Neural Networks, ser. LNCS 3497. Springer, 2005.
- [8] M K HU, "Visual pattern recognition by moment invariants," IRE Trans. Info Theory, vol IT-8, pp. 179-187, Feb. 1962.
- [9] T.H. Reiss, "The Revised Fundamental Theorem of Moment Invariants", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-13, No. 8, August 1991, pp 830-834.
- [10] Sidhartha Maître, "Moment invariant ",IEEE Preceding vol. 67, no. 4, april 1979.
- [11] Jan Flusser, "Rotation Moment Invariants for Recognition of Symmetric Objects," IEEE Trans. On image processing, vol. 15, no. 12, pp. 3784-3790, December 2006.
- [12] S. 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, December 2007.
- [13] J.-X. Du, X.-F. Wang and G.-J. Zhang, "Leaf shape based plant species recognition," *Applied Mathematics and Computation*, vol. 185, 2007.
- [14] J. Hossain and M.A. Amin, "Leaf shape identification based plant biometrics", 13th International Conference on Computer and Information Technology (ICIT), Pp. 458-463, 2010.
- [15] Sandeep kumar E. "Leaf, colour, Area And Edge Features Based Approach For Identification Of Indian Plants" *Indian Journal of Computer Science and Engineering (IJCSSE)*, ISSN : 0976-5166, Vol. 3 No.3 Jun-Jul 2012.
- [16] S. Prasad, P. Kumar and R.C. Tripathi, "Plant leaf species identification using Curvelet transform", 2nd International Conference on Computer and Communication Technology (ICCT), Pp. 646–652, 2011.
- [17] Abdul Kadir, L.E. Nugroho, A.susanto, P.Insap Santosa, "Experiments of Zernike Moments for Leaf Identification", *Journal of Theoretical and Applied Information Technology*, 15 July 2012. Vol. 41 No.1.
- [18] 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, No. 10, 2011
- [19] Kue-Bum Lee and Kwang-Seok Hong, "An Implementation of Leaf Recognition System using Leaf Vein and Shape", *International Journal of Bio-Science and Bio-Technology*, Vol. 5 No. 2, April 2013.
- [20] Sanjeevan G, "Plant Recognition from Leaf Image through Artificial Neural Network" *International Journal of Computer Applications*, Jan 2013. Vol. 62 No.17.
- [21] Mei Fern Bong, Ghazali bin Sulong, Mohd Shafry Mohd Rahim, "Recognition of Leaf Based on Its Tip and Base using Centroid Contour Gradient" *International Journal of Computer Science Issues*, March 2013, Vol. 10, Issue 2, No 2.
- [22] Caner Uluturk and Aybars Ugur, "Recognition of Leaves Based on Morphological Features Derived From Two Half-Regions"